2016

Metric Representations Of Networks

Santiago Segarra
University of Pennsylvania, santiagosegarra@gmail.com

Metric Representations Of Networks

Abstract
The goal of this thesis is to analyze networks by first projecting them onto structured metric-like spaces -- governed by a generalized triangle inequality -- and then leveraging this structure to facilitate the analysis. Networks encode relationships between pairs of nodes, however, the relationship between two nodes can be independent of the other ones and need not be defined for every pair. This is not true for metric spaces, where the triangle inequality imposes conditions that must be satisfied by triads of distances and these must be defined for every pair of nodes. In general terms, this additional structure facilitates the analysis and algorithm design in metric spaces. In deriving metric projections for networks, an axiomatic approach is pursued where we encode as axioms intuitively desirable properties and then seek for admissible projections satisfying these axioms. Although small variations are introduced throughout the thesis, the axioms of projection -- a network that already has the desired metric structure must remain unchanged -- and transformation -- when reducing dissimilarities in a network the projected distances cannot increase -- shape all of the axiomatic constructions considered. Notwithstanding their apparent weakness, the aforementioned axioms serve as a solid foundation for the theory of metric representations of networks.

We begin by focusing on hierarchical clustering of asymmetric networks, which can be framed as a network projection problem onto ultrametric spaces. We show that the set of admissible methods is infinite but bounded in a well-defined sense and state additional desirable properties to further winnow the admissibility landscape. Algorithms for the clustering methods developed are also derived and implemented. We then shift focus to projections onto generalized q-metric spaces, a parametric family containing among others the (regular) metric and ultrametric spaces. A uniqueness result is shown for the projection of symmetric networks whereas for asymmetric networks we prove that all admissible projections are contained between two extreme methods. Furthermore, projections are illustrated via their implementation for efficient search and data visualization. Lastly, our analysis is extended to encompass projections of dioid spaces, natural algebraic generalizations of weighted networks.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Electrical & Systems Engineering

First Advisor
Alejandro Ribeiro

Keywords
Axiomatic construction, Hierarchical clustering, Metrics, Networks

Subject Categories
Computer Sciences | Electrical and Electronics | Mathematics

METRIC REPRESENTATIONS OF NETWORKS

Santiago Segarra

A DISSERTATION

in

Electrical and Systems Engineering

Presented to the Faculties of the University of Pennsylvania
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
2016

Supervisor of Dissertation

Alejandro Ribeiro, Rosenbluth Associate Professor of Electrical and Systems Engineering

Graduate Group Chairperson

Alejandro Ribeiro, Rosenbluth Associate Professor of Electrical and Systems Engineering

Dissertation Committee
Daniel E. Koditschek, Alfred Fitler Moore Professor, Electrical and Systems Engineering
Robert Ghrist, Andrea Mitchell Professor of Mathematics and Electrical and Systems Engineering
Facundo Mémoli, Associate Professor of Mathematics and Computer Science and Engineering, Ohio State University

METRIC REPRESENTATIONS OF NETWORKS

To my parents: Silvia y José.

Acknowledgments

It is a fair assessment to claim that no aspect of myself remained unchanged during the past five years. Even though time was a necessary condition for these changes, it was far from sufficient, with the main determinant being the people with whom I interacted: both new people I encountered along the way and known people that traversed the way along with me, albeit not always geographically close.

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Alejandro Ribeiro. Without his useful comments, thoughtful remarks, motivating discussions, and constant support this thesis would not have been possible. More fundamentally, I want to thank him for giving me the privilege of being his student, for teaching me what research is, and for always encouraging me to improve myself.

Due thanks go to Prof. Dan Koditschek, Prof. Rob Ghrist, and Prof. Facundo Mémoli for agreeing to serve on my committee and for their interest in my work. I sincerely look up to them and this has been an additional motivation in the generation of this thesis. In particular, I want to thank Facundo for our close collaboration during the first part of my Ph.D. from which I learned a lot and without which this thesis would not have been possible either. I have great memories of long afternoons in front of a board discussing some of the theorems contained in these pages.

Throughout my graduate studies, I had the pleasure to closely collaborate with Prof. Antonio G. Marques, Prof. Geert Leus, and Prof. Gonzalo Mateos and I greatly benefited from these interactions both at a professional and a personal level. Although none of the technical content here presented emerged from these collaborations, they partially shaped my mindset as a researcher and, thus, their influence is indirectly present in this work. Furthermore, I want to thank all my friends and colleagues for making Philly feel like home. From playing soccer to discussing a research problem, everything is twice as enjoyable when surrounded by the amazing group of people I had the luck to be a part of. Special thanks to all the current and former inhabitants of Moore 306 for the long hours shared together and, in particular, to Mark Eisen, Weiyu Huang, and Fernando Gama with whom I had the chance to collaborate.

Finally, my deepest gratitude goes to my family and friends who have little to do with me as a researcher but a lot to do with who I am. To my friends back home, thank you for making me feel like I never left. To my parents, Silvia and José, thank you for your constant love and disinterested support without which I would not be writing these words. For this and so much more I will be always grateful and this thesis, dedicated to both of you, is intended as a token of my gratitude. To my brothers, Alejandro, Pablo, and Mariano who are extraordinary role models, to their lovely wives and adorable daughters, thank you for being so close despite being so far away. Last but not least, to you Belén, thank you for wanting to embark on this journey together, I cannot wait to see where it takes us, and thank you for so many things that, if listed all, would render this thesis twice as long.

Santiago Segarra, Philadelphia, August 2016

ABSTRACT

METRIC REPRESENTATIONS OF NETWORKS

Santiago Segarra
Alejandro Ribeiro

The goal of this thesis is to analyze networks by first projecting them onto structured metric-like spaces – governed by a generalized triangle inequality – and then leveraging this structure to facilitate the analysis. Networks encode relationships between pairs of nodes, however, the relationship between two nodes can be independent of the other ones and need not be defined for every pair. This is not true for metric spaces, where the triangle inequality imposes conditions that must be satisfied by triads of distances and these must be defined for every pair of nodes. In general terms, this additional structure facilitates the analysis and algorithm design in metric spaces. In deriving metric projections for networks, an axiomatic approach is pursued where we encode as axioms intuitively desirable properties and then seek for *admissible* projections satisfying these axioms. Although small variations are introduced throughout the thesis, the axioms of projection – a network that already has the desired metric structure must remain unchanged – and transformation – when reducing dissimilarities in a network the projected distances cannot increase – shape all of the axiomatic constructions considered. Notwithstanding their apparent weakness, the aforementioned axioms serve as a solid foundation for the theory of metric representations of networks.

We begin by focusing on hierarchical clustering of asymmetric networks, which can be framed as a network projection problem onto ultrametric spaces. We show that the set of admissible methods is infinite but bounded in a well-defined sense and state additional desirable properties to further winnow the admissibility landscape. Algorithms for the clustering methods developed are also derived and implemented. We then shift focus to projections onto generalized q-metric spaces, a parametric family containing among others the (regular) metric and ultrametric spaces. A uniqueness result is shown for the projection of symmetric networks whereas for asymmetric networks we prove that all admissible projections are contained between two extreme methods. Furthermore, projections are illustrated via their implementation for efficient search and data visualization. Lastly, our analysis is extended to encompass projections of dioid spaces, natural algebraic generalizations of weighted networks.

Contents

List of Tables

List of Figures

Chapter 1

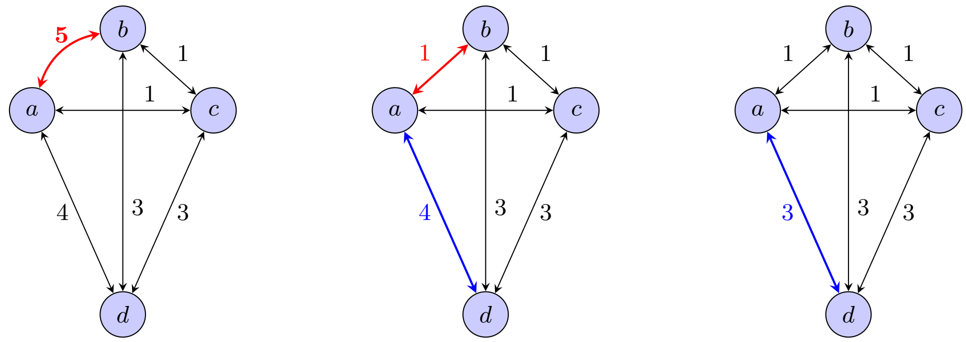
Introduction

Data is getting big, but more than big it is getting pervasive. As our lives integrate with the digital world, larger traces of our actions get recorded. Such pervasive collection leads to the emergence of information structures for which analytical tools are not yet well-developed. Networks, i.e., structures encoding relationships between pairs of elements, belong in this category and, at the same time, play a main role in our current scientific understanding of a wide range of disciplines including biology [7, 53], sociology [46, 61], and medicine [88].

An evident obstacle in understanding large scale complex networks is their massive size, with popular information networks – such as the World Wide Web – and online social networks – such as Facebook – containing more than a billion nodes. Our contention, however, is that a substantial part of the difficulty in analyzing and efficiently managing complex networks comes from the lack in structure that networks present. This loose nature contrasts with the rigidity of a closely related construction: the metric space. In a nutshell, the main goal of this thesis is to analyze networks from an alternative perspective where we first project them onto structured metric-like spaces and then leverage this structure to facilitate the analysis.

1.1 Motivation and context

A point cloud in Euclidean space is a simpler object than a weighted network of equal dimension because the triangle inequality endows the cloud with a structure the network lacks. Determining, say, whether a is more similar to b than it is to d in the metric network in Fig. 1.1(b) is straightforward – node a is closer to b than to d. However, answering the same question in the (non-metric) dissimilarity network in Fig. 1.1(a) is not as immediate. The direct dissimilarity between a and b is registered as 5 which we could use to conclude that a is closer to d than to b since the direct dissimilarity between a and d is reported as 4. Yet, we still have that a is close to c, which is in turn close to b, and this intuitively



(a) (b) (c)

Figure 1.1: Examples of a general weighted, a metric, and an ultrametric network. (a) Network of (non-metric) dissimilarities. Even simple questions are difficult to answer. E.g., is a closer to d than to b as the direct dissimilarities indicate or is the mutual proximity of a and b to c sufficient to claim the opposite? (b) Metric network. Similarity comparisons between nodes are straightforward. E.g. it is immediate to see that a is closer to b than it is to d. Nevertheless, node groupings are not obvious. Is node d at a distance of 3 or 4 from the cluster $\{a, b, c\}$? (c) Ultrametric network. The structure imposed by the strong triangle inequality induces a hierarchy of clusters in the network.

seems to imply that a and b are not that different after all. We can interpret nodes as social agents, edge dissimilarities as representing the frequency with which two nodes exchange opinions, and our goal as the study of the propagation of opinions in the network. In that case we know that a and b do not interact frequently but they will be highly influenced by each other's opinions through their mutual frequent interaction with node c. Arguably, this indirect frequent interaction with b has a larger effect on the opinion of a than the direct but somewhat infrequent interaction with d.

This simple example illustrates the very fundamental fact that questions that are difficult to answer for arbitrary weighted networks become simple, or at least simpler, when the network has a metric structure. E.g., consider a problem of proximity search in which we are given a network and an element whose dissimilarity to different nodes of the network can be determined and are asked to find the element that is least dissimilar to the given one. Finding the least dissimilar node in an arbitrary network requires comparison against all nodes and incurs a complexity that is linear in the number of nodes. In a metric space, however, the triangle inequality encodes a transitive notion of proximity. If two points are close to each other in a metric space and one of them is close to a third point, then the other one is also close to this third point. This characteristic can be exploited to design efficient search methods using metric trees whose complexity is logarithmic in the number of nodes [83, 84, 91]. Likewise, many hard combinatorial problems on graphs are known to be polynomial-time approximable in metric spaces but not approximable in generic networks. The traveling salesman problem, for instance, is not approximable in generic graphs but is approximable in polynomial time to within a factor of 3/2 in metric spaces [20]. In either case, the advantage of the metric space is that the triangle inequality endows it with a structure that an arbitrary network lacks. It is this structure that makes network analysis and algorithm design tractable.

If some problems are challenging in generic networks but not so challenging in metric spaces, a route to network analysis is to project networks onto metric spaces. We are then searching for a projection operator that takes a dataset like the network in Fig. 1.1(a) as input and generates a dataset with a metric structure like that of the network in Fig. 1.1(b). The question that arises, then, is the design of methods – i.e., shall we replace the dissimilarity between a and b of Fig. 1.1(a) by 1 as shown in Fig. 1.1(b), or by 2, 1.5, or $\sqrt{2}$? – and corresponding algorithms to implement these projections. One of our goals is then to develop a mathematical theory for the metric representation of network data in order to fundament the design of methods and algorithms to project networks onto metric spaces.

The concept of an abstract metric space, introduced in the early 20th century [28], encompasses a wide variety of scientific and engineering constructions where the notion of distance is present. During the first half of the past century, metric spaces were regarded as mere presentations of underlying topological spaces and a lot of effort was put on the study of embedding general metric spaces into more familiar ones [5]. However, in the late sixties, there was a partial shift in the focus of analysis, and the first formal studies of metric spaces *as such* – not seen as representations of some underlying topological space – appeared [42], specifically in the field of category theory [2]. We leverage the fact that the fundamental understanding of metric spaces is more developed than that of networks in order to gain insight on the latter by projecting them onto the former and using analytical tools designed for the study of metric spaces.

The traditional way of mapping a generic dissimilarity function between pairs of points to a metric space is through multidimensional scaling (MDS) [23]. Different problem formulations give rise to the definition of different types of MDS with a basic distinction between metric MDS, where the input consists of quantitative similarities [58, 82], and non-metric MDS where dissimilarities can be ordinal [49, 73]. However, all these techniques have in common that one of the end goals is to facilitate visualization of the data [50]. Thus, unlike the type of projections considered in this thesis, MDS embeds the input dissimilarities into familiar and low-dimensional metric spaces such as $\mathbb{R}^2$ or $\mathbb{R}^3$.

Some problems are still hard to elucidate even in metric spaces, thus requiring projections onto even more structured spaces. For example, if one is interested in grouping or clustering the nodes in the metric network in Fig. 1.1(b), it is evident that nodes a, b, and c are closer together than they are to d. However, the distance between this latter singleton to the cluster $\{a, b, c\}$ is unclear. Should it be the maximum distance 4, the minimum distance 3, or the average distance 10/3? Notice that this ambiguity does not arise in the

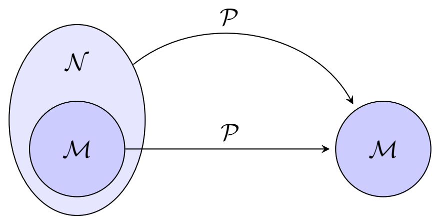


Figure 1.2: Axiom of Projection. The projection method $\mathcal{P}$ is admissible if the set of metric spaces $\mathcal{M}$ is a fixed set of $\mathcal{P}$.

network in Fig. 1.1(c) where it is clear that nodes a, b, and c form a cluster at a distance – or resolution, as will be formally introduced in Chapter 2 – of 1, which is at a distance of 3 from the remaining node d. The fact that clusters can be readily extracted from the network in Fig. 1.1(c) is not coincidental. This network can be shown to have an ultrametric structure, i.e., it is a metric network that satisfies a stronger version of the triangle inequality. This type of networks can be equivalently represented as dendrograms [44], which are the outputs of hierarchical clustering methods. Putting it differently, projecting arbitrary weighted networks onto ultrametric spaces is an alternative way of framing the problem of hierarchical clustering in networks.

Clustering, i.e. partitioning a dataset into groups such that objects in one group are more similar to each other than they are to objects outside the group, is a fundamental tool for the advancement of knowledge in a wide range of disciplines such as genetics [19], computer vision [30], sociology [35], and marketing [66]. Motivated by its relevance, literally hundreds of methods that can be applied to the determination of hierarchical [43, 51] and non-hierarchical clusters in finite metric (thus symmetric) spaces exist [1,21,62–64,68,74,87]. Of particular relevance to our work is the case of hierarchical clustering where, instead of a single partition, we look for a family of nested partitions indexed by a resolution parameter and graphically represented by a tree-like structure called dendrogram. Even in the case of asymmetric networks in which the dissimilarity from node x to node x' may differ from the one from x' to x [70], multiple methods have been developed to extend the notion of clustering into this less intuitive domain [6,40,55,60,65,76,79,93]. Although not as developed as its practice [34], the theoretical framework for clustering has been developed over the last decade for non-hierarchical [11, 47, 56, 57, 85, 92] and hierarchical clustering [10, 12, 13, 15].

Metric and ultrametric spaces are two examples of weighted networks having an added structure induced by the triangle and strong triangle inequalities, respectively. Throughout this thesis, a wide gamut of metric-like structures for networks will be presented and our main goal will be the design of methods and associated algorithms to induce a desired metric structure in a given network. In other words, our objective will be to design projections from arbitrary weighted networks onto networks possessing a metric-like structure.

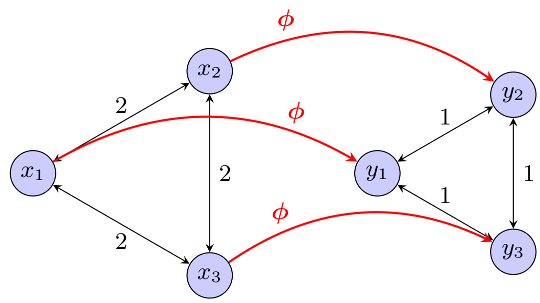


Figure 1.3: Axiom of Injective Transformation. If nodes can be mapped injectively to a network with smaller dissimilarities, the distances in the projection of the latter cannot be larger than the corresponding distances in the projection of the original one.

Devising methods to create metric structure is not difficult. Indeed, if one is interested in inducing regular metric spaces, it can be shown that it suffices to replace each arc by the minimum norm among all paths that link the given nodes. Using the 1-norm this is equivalent to the shortest path distance between the adjacent nodes, but an infinite number of methods are possible since the choice of norm for the path is arbitrary. It therefore seems that the important question is rather the opposite of devising projection methods: Out of the many ways of inducing metric structure, which method is most desirable? Inspired by the success of axiomatic approaches in the study of clustering [10, 11, 47], we adopt an axiomatic strategy in answering this question. Desirable properties are stated as axioms and we proceed to search for methods that are admissible with respect to them. In particular, the core of the theory is built on the following two axioms:

(AA1) Axiom of Projection. If the projection method is applied to a network that already possesses the desired metric structure, the outcome is identical to the original network; see Figure 1.2.

(AA2) Axiom of Injective Transformation. Consider a network and reduce some pairwise dissimilarities but increase none. The respective outcomes of the projection method are such that distances in the projection of the transformed network are not larger than distances in the projection of the original network; see Figure 1.3.

Axioms (AA1) and (AA2) state very reasonable conditions for admissibility of a projection method $\mathcal{P}$. The Axiom of Injective Transformation (AA2) simply requires that smaller networks have smaller projections. The Axiom of Projection (AA1) is a minimal condition for a method to be interpreted as a projection. If a network already belongs to the set of metric spaces, a projection onto this space cannot alter the given network. Given their apparent weakness, one should question the wisdom of attempting to build a theory of metric representations supported on axioms (AA1) and (AA2). However, as we show throughout the thesis, the joint consideration of the aforementioned axioms induces more structure that what can be grasped at first sight.

1.2 Thesis outline and contributions

Under the general formulation of projecting networks onto metric structures, this thesis contributes to the current understanding of different problems. Among these, the most popular is hierarchical clustering of networks which, although not always stated in these terms, corresponds to the projection of networks onto ultrametric spaces. This is the focus of Part I. Due to the existence of previous axiomatic approaches for the hierarchical clustering of *symmetric* networks [10, 11], we emphasize the clustering of possibly *asymmetric* networks, thus, extending and generalizing existing works. The principal contributions of this first part include: i) laying the main and alternative axiomatic frameworks for the study of hierarchical clustering in asymmetric networks; ii) describing the landscape of admissible methods satisfying such axioms; iii) stating additional desirable clustering properties and finding a complete characterization of methods satisfying these properties; and iv) deriving and implementing algorithms for the clustering methods developed.

Part I of this thesis reveals that the admissible projections onto ultrametric spaces are few, does the same hold true for more general metric representations? The affirmative answer to this question is developed in Part II, where we present a theory of metric representations built on minimal assumptions and rooted on the Axioms of Projection (AA1) and Injective Transformation (AA2). In this case, the landscape of existing works is more barren even for the case of symmetric networks, thus, we begin by studying the projection of symmetric networks onto generalized q-metric spaces to then move into the richer domain of asymmetric networks. In this second part we also depart from the classical concept of a weighted network to work with more abstract constructions founded on the algebraic concept of dioids. The main contributions of this second part include: i) stating the first axiomatic framework for the metric representation of networks; ii) characterizing a unique canonical projection method for symmetric networks and describing the bounded set of admissible methods for asymmetric networks; and iii) extending the analysis to encompass more general constructions via the incorporation of dioid spaces. A detailed explanation of the contributions of each chapter is presented next.

Chapter 2 opens Part I of the thesis by presenting the mathematical concepts needed for the study of hierarchical clustering in asymmetric networks. In particular, dendrograms are introduced and their equivalence with ultrametric spaces – fundamental for the theory developed – is formally stated. As already mentioned, a dendrogram is a graphical tree-like representation of a nested collection of partitions indexed by a resolution parameter; see e.g. Fig. 2.2. Even though the Axioms of Projection (AA1) and Injective Transformation (AA2) constitute the backbone of the axiomatic framework considered throughout the thesis, in this first part we utilize two minor variations on these that best suit the existing clustering literature. These are the Axioms of Value (A1) and Transformation (A2), formally stated in Section 2.2, that correspond to the following intuitions:

(A1) Axiom of Value. For an asymmetric network with two nodes, the nodes are clustered together at a resolution equal to the maximum of the two dissimilarities between them.

(A2) Axiom of Transformation. If we consider a network and map it to another such that no pairwise dissimilarity is increased by the mapping, the resolution at which two nodes become part of the same cluster is not larger than the resolution at which they were clustered together in the dendrogram of the original network.

The intuition supporting the Axiom of Transformation is that if some nodes become closer to each other, it may be that new clusters arise, but no cluster can disappear. In contrast to axiom (AA2), the maps considered in axiom (A2) need not be injective. The intuition supporting the Axiom of Value is that the two nodes in a two-node network form a single cluster at resolutions that allow them to influence each other. A hierarchical clustering method satisfying axioms (A1) and (A2) is said to be *admissible.*

The Axiom of Value equates clustering in asymmetric two-node networks with direct mutual influence. Our first theoretical study is the relationship between clustering and mutual influence in networks of arbitrary size (Section 2.3). In particular, we show that the outcome of any admissible hierarchical clustering method is such that a necessary condition for two nodes to cluster together is the existence of paths that allow for direct or indirect influence between them. We can interpret this result as showing that the requirement of direct influence in the two-node network in the Axiom of Value (A1) induces a requirement for, possibly indirect, influence in general networks. This result is termed the Property of Influence and plays an instrumental role in the theoretical developments presented in this first part. The material discussed in this chapter appeared in [13, 18].

Chapter 3 introduces reciprocal and nonreciprocal clustering, two hierarchical clustering methods that abide by axioms (A1) and (A2). Reciprocal clustering requires clusters to form through edges that are similar in both directions whereas nonreciprocal clustering allows clusters to form through cycles of small dissimilarity. More specifically, reciprocal clustering defines the cost of an edge as the maximum of the two directed dissimilarities. Nodes are clustered together at a given resolution if there exists a path that links them such that all links in the path have a cost smaller than said resolution. In nonreciprocal clustering we consider directed paths and define the cost of a path as the maximum dissimilarity encountered when traversing it from beginning to end. Nodes are clustered together at a given resolution if it is possible to find directed paths in both directions whose edge costs do not exceed the given resolution. Observe that both of these methods rely on the determination of paths of minimax cost linking any pair of nodes. This fact is instrumental in the derivation of algorithms for the computation of output dendrograms.

A fundamental result regarding admissible clustering methods is the proof that any method that satisfies axioms (A1) and (A2) lies between reciprocal and nonreciprocal clustering in a well-defined sense (Section 3.2). Specifically, any clustering method that satisfies axioms (A1) and (A2) forms clusters at resolutions larger than the resolutions at which they are formed by nonreciprocal clustering, and smaller than the resolutions at which they are formed by reciprocal clustering. When restricted to symmetric networks, reciprocal and nonreciprocal clustering yield equivalent outputs, which coincide with the output of single linkage. This observation is consistent with the existence and uniqueness result in [10] since axioms (A1) and (A2) are reduced to two of the axioms considered in [10] when we restrict attention to metric data. The derivations in this thesis show that the existence and uniqueness result in [10] is true for all symmetric, not necessarily metric, datasets and that a third axiom considered there is redundant because it is implied by the other two.

We then unveil some of the clustering methods that lie between reciprocal and nonreciprocal clustering and study their properties (Section 3.3). Three families of intermediate clustering methods are introduced. The grafting methods consist in attaching the clustering output structures of the reciprocal and nonreciprocal methods in such a way that admissibility is guaranteed. We further present an operation between methods that can be regarded as a convex combination in the space of clustering methods. This operation is shown to preserve admissibility therefore giving rise to a second family of admissible methods. A third family is defined in the form of semi-reciprocal methods that allow the formation of cyclic influences in a more restrictive sense than nonreciprocal clustering but more permissive than reciprocal clustering.

In some applications the requirement for bidirectional influence in the Axiom of Value is not justified as unidirectional influence suffices to establish proximity. This alternative value statement leads to the study of alternative axiomatic constructions and their corresponding admissible hierarchical clustering methods (Section 3.4). We first propose an Alternative Axiom of Value in which clusters in two-node networks are formed at the minimum of the two dissimilarities:

(A1") Alternative Axiom of Value. For a network with two nodes, the nodes are clustered together at a resolution equal to the *minimum* of the two dissimilarities between them.

Under this axiomatic framework we define unilateral clustering as a method in which influence propagates through paths of nodes that are close in at least one direction (Section 3.4.1). Contrary to the case of admissibility with respect to (A1)-(A2) in which a range of methods exists, unilateral clustering is the unique method that is admissible with respect to (A1") and (A2). A second alternative is to take an agnostic position and allow nodes in two-node networks to cluster at any resolution between the minimum and the maximum dissimilarity between them. All methods considered in this chapter satisfy this agnostic axiom and, not surprisingly, outcomes of methods that satisfy this agnostic axiom are uniformly bounded between unilateral and reciprocal clustering.

Besides the characterization of methods that are admissible with respect to different sets of axioms, we also develop algorithms to compute their output dendrograms. The determination of algorithms for all of the methods introduced is given by the computation of matrix powers in a min-max dioid algebra [32]. In this dioid algebra we operate in the field of positive reals and define the addition operation between two scalars to be their minimum and the product operation to be their maximum (Section 3.5). From this definition it follows that the (i, j)-th entry of the n-th dioid power of a matrix of network dissimilarities represents the minimax cost of a path linking node i to node j with at most n edges. As we have already mentioned, reciprocal and nonreciprocal clustering require the determination of paths of minimax cost. Similarly, other clustering methods introduced can be interpreted as minimax path costs of a previously modified matrix of dissimilarities which can therefore be framed in terms of dioid matrix powers as well. The relation between dioid algebras and the generalized problem of projecting networks onto metric spaces exceeds the design of algorithms. However, a detailed analysis of this is postponed until Chapter 11. The results in this chapter appeared in [12–14, 16, 18].

Chapter 4 lays the foundations of hierarchical quasi-clustering of asymmetric networks. Dendrograms, which represent the outputs of hierarchical clustering methods, are symmetric structures in that node x being clustered together with node x' at a given resolution is equivalent to x' being clustered together with x at that resolution. Having a symmetric output is, perhaps, a mismatched requirement for the processing of asymmetric data. This mismatch motivates the development of asymmetric structures that generalize the concept of a dendrogram.

Recall that a dendrogram is a collection of nested partitions indexed by a resolution parameter and each partition is induced by an equivalence relation, i.e., a relation satisfying the reflexivity, *symmetry*, and transitivity properties. Hence, the symmetry in hierarchical clustering derives from the symmetry property of equivalence relations which we lift to construct the asymmetric equivalent of hierarchical clustering.

To do so we define a quasi-equivalence relation as one that is reflexive and transitive but not necessarily symmetric and define a quasi-partition as the structure induced by a quasi-equivalence relation – these structures are also known as partial orders [37]. Quasi-partitions contain disjoint blocks of nodes like regular partitions but also include an influence structure between the blocks derived from the asymmetry in the original network. A quasi-dendrogram is further defined as a nested collection of quasi-partitions, and a hierarchical quasi-clustering method as a map from the space of networks to the space of quasi-dendrograms.

As in the case of (regular) hierarchical clustering we proceed to study admissibility with respect to asymmetric versions of the Axioms of Value and Transformation (Section 4.2). We show that there is a unique quasi-clustering method admissible with respect to these axioms and that this method is an asymmetric version of the single linkage clustering method. The analysis hinges upon an equivalence between quasi-dendrograms and quasi-ultrametrics that generalizes the known equivalence between dendrograms and ultrametrics. Our results in quasi-clustering were published in [15].

Chapter 5 further winnows the space of admissible clustering methods – bounded between reciprocal and nonreciprocal clustering – by imposing additional desirable features. These features are framed in the form of the properties of scale preservation, representability, excisiveness, and stability.

Scale preservation is defined as the requirement that the formation of clusters does not depend on the scale used to measure dissimilarities (Section 5.1). Formally, we say that a monotone non-decreasing function applied to all dissimilarities in the network represents a scale transformation. A method is scale preserving if the output dendrogram of the rescaled network is such that the resolution at which nodes merge in a common cluster is obtained by applying the rescaling function to the corresponding merging resolutions in the original network. This condition ensures that scale preserving methods maintain the clustering structure since nodes might cluster at different resolutions but the order in which they cluster together remains unchanged.

Representability is an attempt to restrict attention to methods that can be described through the specification of their effect on particular exemplar networks (Section 5.2). These exemplar networks are called representers and are interpreted as minimal clustering units. When given an arbitrary network we cluster nodes together by mapping linearly scaled representers with overlapping images such that a path of scaled representers links the two nodes under consideration. The merging resolution for these nodes in the output dendrogram is given by the smallest linear scaling needed to implement this mapping so that there are no increases in pairwise dissimilarities.

Excisiveness encodes the property that clustering a previously clustered network does not generate new clusters (Section 5.3). In particular, if we apply an excisive hierarchical clustering method to any cluster of a given network, we obtain the same hierarchical structure within the cluster that was obtained when clustering the whole network. This condition guarantees that the hierarchical structure of each cluster is determined by relations between the nodes within that cluster and is not affected by nodes outside of it. Our definitions of representability and excisiveness are generalizations of similar concepts defined for non-hierarchical clustering in [11].

Requiring the simultaneous fulfillment of (A1), (A2), scale preservation, and representability reduces the space of admissible methods to a unique family generated by a particular class of representers. This class is composed of representers in which dissimilarities between pairs of nodes are either undefined or equal to 1. We can think of this particular family of representers – called structure representers – as encoding influence structures. The combination of admissibility, excisiveness, and linear scale preservation – i.e., scale preservation for linear rescalings – is shown to imply representability (Section 5.3). Further considering that linear scale preservation is a particular case of scale preservation it follows that admissible methods that are also excisive and scale preserving are representable by structure representers. This is a strong result since it provides a full characterization of the set of methods that satisfy (A1), (A2), excisiveness, and scale preservation.

In order to study the stability of clustering methods with respect to perturbations of a network, we adapt the Gromov-Hausdorff distance between finite metric spaces [9, Chapter 7.3] to a distance between asymmetric networks (Section 5.4). This distance allows us to compare any two networks, even when they have different node sets. Since dendrograms are equivalent to finite ultrametric spaces which in turn are particular cases of asymmetric networks, we can use the Gromov-Hausdorff distance to quantify the difference between two dendrograms obtained when clustering two different networks. We then say that a clustering method is stable if the clustering outputs of similar networks are close to each other. More precisely, we say that a clustering method is stable if, for any pair of networks, the distance between the output dendrograms can be uniformly bounded by the distance between the original networks. In particular, stability of a method guarantees robustness to the presence of noise in the dissimilarity values. Although not every method considered is stable, we show stability for most of the methods including the reciprocal, nonreciprocal, semi-reciprocal, and unilateral clustering methods.

The results on excisiveness and scale preservation appeared in [17].

Chapter 6 illustrates the clustering methods introduced thus far via their application to the network of internal migration between states of the United States (U.S.) and the network of interactions between economic sectors of the U.S. economy for the year 2011. The purpose of these examples is to understand which information can be extracted by performing hierarchical clustering and quasi-clustering analyses based on the different methods proposed. Analyzing migration clusters provides information on population mixing whereas exploring interactions between economic sectors unveils their relative importances and their differing levels of coupling.

The clusters that appear in the reciprocal dendrogram of the migration network reveal that population movements are dominated by geographical proximity. In particular, the reciprocal dendrogram shows that the strongest bidirectional migration flows correspond to pairs of states sharing urban areas. For this dataset the outputs of the reciprocal and nonreciprocal dendrogram are very similar and is indicative of the rarity of migrational cycles. Combining this observation with the fact that reciprocal and nonreciprocal clustering are uniform lower and upper bounds on all methods that satisfy (A1) and (A2), it further follows that all the methods that satisfy these axioms yield similar clustering outputs. The application of unilateral clustering – which does *not* satisfy axioms (A1) and (A2) – reveals regional separations more marked than the ones that appear in the reciprocal and nonreciprocal dendrograms. For coarse resolutions we observe a clear East-West separation whereas for finer resolutions we observe clustering around the most populous states. This latter pattern is indicative of the ability of unilateral clustering to capture the unidirectional influence of the populous states on the smaller ones, as opposed to the methods satisfying (A1)-(A2) which capture bidirectional influence. To study the influence between clusters we apply the directed single linkage quasi-clustering method. Analysis of the output quasi-dendrograms show, e.g., the dominant roles of California and Massachusetts in the population influxes into the West Coast and New England, respectively.

The network of interactions between sectors of the U.S. economy records how much of a sector's output is used as input to another sector. In contrast to the migration matrix the reciprocal and nonreciprocal dendrograms uncover different clustering structures. The reciprocal dendrogram generates distinctive clusters of sectors that have significant interactions. These include a cluster of service sectors such as financial, professional, insurance, and support services; and a cluster of extractive industries such as mining, primary metals, and oil and gas extraction. The nonreciprocal dendrogram does not have distinctive separate clusters but rather a single cluster around which sectors coalesce as the resolution coarsens. This pattern indicates that considering cycles of influence yields a different understanding of interactions between sectors of the U.S. economy than what can be weaned from the direct mutual influence required by reciprocal clustering. An intermediate picture that considers cycles of restricted length is obtained by using cyclic methods, a particular class of representable methods. Unilateral clustering yields clusters that group around large sectors of the economy, as happened with large states in the migration network. We finally consider the use of the directed single linkage quasi-clustering method to understand influences between clusters. These numerical experiments were partially discussed in [12, 16].

Chapter 7 closes the first part of the thesis by presenting a summary of the results in the form of a taxonomic analysis. More concretely, in the preceding chapters we have contributed to the theory of hierarchical clustering in asymmetric networks in two intertwined directions. First, we have derived novel clustering methods and associated algorithms. These methods include, among others, reciprocal, nonreciprocal, semi-reciprocal, and unilateral clustering. Secondly, we have introduced a series of desirable properties to be satisfied by clustering methods. The more fundamental properties were termed as axioms, constituting the basal backbone of our construction. Additional properties – scale invariance, excisiveness, representability, and stability – were used to further winnow the space of admissible methods. In this chapter, even though no new results are introduced, we facilitate the understanding of the previous material by reviewing, comparing, and contrasting our contributions in Part I.

Chapter 8 launches Part II of the thesis where we depart from the exclusive study of hierarchical clustering – i.e., projection of networks onto ultrametric spaces – to study the more general problem of projecting networks onto spaces with metric structure. We begin by presenting mathematical concepts, complementary to those already introduced in Chapter 2, that serve as basis for the results in this second part. In particular, we introduce q-metric spaces, an extension of metric spaces that enables the generalization of our framework to a variety of structured spaces. For instance, both the (regular) metric spaces and the ultrametric spaces can be recovered as particular cases of q-metric spaces. Formally, the problem addressed in this chapter is how to project *symmetric* networks onto q-metric spaces for all possible q. As customary throughout the thesis, we follow an axiomatic approach and, in Section 8.2 we formally define the Axioms of Projection (AA1) and Injective Transformation (AA2) intuitively introduced in Section 1.1. Mimicking the lexicon in Part I, we deem as *admissible* a projection method that satisfies these two axioms. The apparent weakness of these axioms contrasts with their stringent theoretical consequences. More specifically, in Section 8.3 we show that there is a unique admissible way of projecting networks onto q-metric spaces, that we denominate as the *canonical* q-metric projection. Putting it differently, any conceivable way of inducing a q-metric in a network different from the canonical way must violate either axiom (AA1) or axiom (AA2) or both. For the particular case of (regular) metric spaces, this implies that the shortest path between two nodes is the only admissible distance between them. Moreover, when focusing on projections onto ultrametric spaces, single linkage clustering is identified as the only admissible hierarchical clustering method – in accordance with our results in Part I.

In Section 8.4 we show that the Axioms of Projection and Transformation confer three practical properties to canonical projections: optimality, stability, and nestedness. The former implies that the canonical projection of a network can be used to approximate the solution of combinatorial optimization problems on the network. It was already argued in Section 1.1 that a number of combinatorial problems in arbitrary networks are approximable in the presence of metric structure. Moreover, we show that the canonical metric projection outputs the uniformly largest metric among those dominated by the dissimilarities in the input network. Combining these two facts, the canonical projection can be used to provide tight lower bounds for the solution of combinatorial graph problems such as the traveling salesman and graph bisection. Stability ensures that two networks that are similar are projected onto similar metric spaces, bounding the effect of input noise on the projected metric space. The effect of noise is shown to decrease for larger q and, for $q = \infty$, we recover a stability result that resembles – although not equivalent – the one in Chapter 5. Nestedness implies that the q-metric space obtained when projecting an arbitrary network is invariant to intermediate projections onto other q'-metric spaces with laxer structure. A direct consequence of this is that if one is interested in computing the single linkage hierarchical clustering output of a given network then there is no gain (or loss) in first projecting the network onto a metric space and then computing the clustering output of the resulting metric space. The uniqueness result for metric projections was the main subject of [72].

Chapter 9 returns the focus to the study of *asymmetric* networks. Formally, in this chapter we analyze projections from asymmetric networks onto (regular) metric spaces. The uniqueness result in Chapter 8 is a generalization of the fact that single linkage is the only admissible hierarchical clustering method as derived in Part I. However, our study of hierarchical clustering – or projections onto ultrametric spaces – suggests that projections of symmetric networks and asymmetric networks are fundamentally different. The results in this chapter indicate that this is the case for projections onto metric spaces as well.

In order to handle the asymmetry in the input networks we begin by introducing a symmetrizing function, which plays a role akin to that of the maximum function in the Axiom of Value (A1) introduced in Part I. The results in this chapter are not restrained to a specific symmetrizing function but hold for any function that satisfies a series of requirements. Using the symmetrizing function, we update the Axiom of Projection (AA1) to accommodate asymmetric networks as inputs giving rise to the Axiom of Symmetrization (AS1), formally stated in Section 9.1 but corresponding to the following intuition:

(AS1) Axiom of Symmetrization. If the projection method is applied to a quasi-metric space then its action is the same as the symmetrization function.

A quasi-metric space satisfies an asymmetric version of the triangle inequality, hence, a mere symmetrization is enough to induce a metric structure. Axiom (AS1) requires that in this case, an admissible projection should not introduce any additional changes. Axiom (AS1) implies the Axiom of Projection (AA1) since when the input is metric, the symmetrization function – and hence the projection – leaves the network unaltered.

Defining admissibility in terms of axioms (AS1) and (AA2), in Section 9.2 we introduce two admissible methods: symmetrize-then-project (SymPro) and project-then-symmetrize (ProSym). As their names suggest, in SymPro we first induce the symmetry property in the network and then the triangle inequality whereas the opposite is true for ProSym. These methods, the metric analogues of reciprocal and nonreciprocal clustering, bound the set of all admissible (regular) metric projections in a well-defined sense (Section 9.3). More precisely, the distances between pairs of nodes output by any admissible metric projection are upper bounded by the SymPro distances and lower bounded by the ProSym distances. Finally, we show that there exist projection methods contained strictly between ProSym and SymPro by effectively constructing an intermediate method. This method, project-then-symmetrize-then-project (PSP), can be intuitively understood as first forcing a partial version of the triangle inequality, then imposing symmetry, and then completing the enforcement of the triangle inequality.

Chapter 10 contains extensions and applications of the results developed in the preceding two chapters. In terms of extensions, in Section 10.1 we study the projection of asymmetric networks onto quasi-metric spaces while in Section 10.2 we extend the analysis to quasi-q-metric spaces. After introducing the necessary modifications in the axiomatic framework to accommodate (asymmetric) quasi-metric outputs, we show that there is a unique admissible way of performing the desired projection task. As was seen in Chapter 9, when we force the output of the projections to be symmetric, there are multiple admissible ways of inducing a metric structure. By contrast, when we solve the symmetry mismatch between the input and the output by allowing asymmetric outputs, we obtain a uniqueness result. This can be thought of as a generalization of the quasi-clustering uniqueness result in Chapter 4. In Section 10.2 we also study the projections of asymmetric networks onto q-metric spaces for arbitrary q. I.e., we extend the analysis for $q = 1$ in Chapter 9 and show that under mild conditions the definitions of SymPro and ProSym can be generalized for all q as well as the fact that all admissible methods are bounded by them.

In terms of applications, in Section 10.3 we propose an efficient strategy for approximating search in networks by first projecting a general network onto a (quasi-)q-metric space and then leveraging this structure for search via the construction of generalized (quasi-)q-metric trees. Search problems in networks involve nodes to represent objects, edge weights to represent their dissimilarity, and a search specification that describes the object of interest. The goal is to find the node closest to the specification. While arbitrary networks have no structure that can be exploited for efficient search, the q-triangle inequality of q-metric spaces can be used to reduce the linear cost of a brute force search to an expected cost that is logarithmic in the number of objects. We demonstrate that the parameter q can be used to control the tradeoff between computational gain and search performance. Lastly, in Section 10.4 we utilize metric projections as intermediate steps for the visualization of asymmetric data. Visualization methods facilitate the understanding of high-dimensional data by projecting it into familiar low-dimensional domains. Although MDS is a common way of performing such a projection, symmetric data is needed as input. Hence, we first project the asymmetric data onto a (symmetric) metric space and then apply MDS to the resulting dataset. We illustrate this procedure via the visualization of the economic network introduced in Chapter 6. The visualizations obtained by implementing SymPro and ProSym before the MDS embedding reveal different aspects of the asymmetric network. While the former produces an almost perfect separation between service-providing and good-producing industries, the latter yields a grouping of the sectors by industry type, such as food, metals, or chemicals. This two-step procedure is further justified by drawing connections with Isomap, an established non-linear data visualization method.

The numerical experiments on efficient search were partially discussed in [72].

Chapter 11 generalizes part of the results introduced throughout the thesis to a higher level of algebraic abstraction by relying on dioids. A dioid is a set endowed with two operations that we can interpret as addition and multiplication. However, the most important difference with the more common triadic algebraic structure of rings is that the addition induces a group structure in rings whereas in dioids it induces a canonical order. This order can be used to write inequalities in dioids and, ultimately, to generalize the concept of a triangle inequality to this abstract domain. After introducing in Section 11.1 the algebraic concepts needed for the development of the results in the chapter, in Section 11.2 we formally define dioid spaces and dioid metric spaces. A dioid space can be conceived as the natural generalization of weighted networks, where the weights take values in the underlying dioid that need not coincide with the positive reals. In this way, fairly dissimilar constructions such as weighted networks, unweighted networks, and networks whose edges take as values subsets of a finite vocabulary, can all be viewed as dioid spaces for different dioid algebras. A dioid *metric* space, on the other hand, is a dioid space with the additional structure imposed by a generalized triangle inequality. Hence, our goal in this chapter is to study projections from the set of dioid spaces onto dioid metric spaces for arbitrary dioids. After restating the Axioms of Projection (AA1) and Injective Transformation (AA2) in dioid terms, in Section 11.3 we define a canonical projection to achieve our goal and show that this is the only admissible projection onto dioid metric spaces. The projected value between two nodes is obtained by summing over all paths linking these nodes the product of the dioid values encountered when traversing these paths in order, where the sum and multiplication are defined by the underlying dioid. This uniqueness result is an abstraction of the result in Chapter 8 in the sense that the latter can be recovered when specializing the underlying dioid to $(\bar{\mathbb{R}}_+, \min, +)$. Moreover, the uniqueness result of single linkage as the only admissible hierarchical clustering method for symmetric networks can be recovered as well when considering the underlying dioid $(\bar{\mathbb{R}}_+, \min, \max)$. Other uniqueness results obtained by specializing the underlying dioid are further discussed in Section 11.4.

The thesis closes with conclusions and directions for future research in **Chapter 12**.

Part I

Hierarchical Clustering of Asymmetric Networks

Chapter 2

Axiomatic construction of hierarchical clustering

We begin by introducing notions related to network theory and clustering needed for the development of the results presented in this first part of the thesis (Section 2.1). In particular, we revisit the known equivalence between dendrograms and ultrametrics (Section 2.1.1), which is instrumental to our proofs. The Axioms of Value and Transformation, introduced at an intuitive level in Section 1.2, are formally stated in Section 2.2. We close the chapter with Section 2.3 containing our first theoretical study of the relationship between clustering and mutual influence in networks of arbitrary size.

2.1 Networks, hierarchical clustering, and ultrametrics

We say that a network N is a pair (X, A_X) where X is a finite set of points or nodes and $A_X : X \times X \to \mathbb{R}_+$ is a dissimilarity function. The dissimilarity $A_X(x, x')$ between nodes $x \in X$ and $x' \in X$ is assumed to be non-negative for all pairs (x, x') and 0 if and only if $x = x'$. We do not, however, require A_X to be a metric on the finite set X. Dissimilarity values $A_X(x, x')$ need not satisfy the triangle inequality and, more consequential for the problem considered here, they may be asymmetric in that it is possible to have $A_X(x, x') \neq A_X(x', x)$ for some $x \neq x'$. In some discussions it is convenient to reinterpret the dissimilarity function A_X as the possibly asymmetric matrix $A_X \in \mathbb{R}_+^{n \times n}$ with $(A_X)_{i,j} = A_X(x_i, x_j)$ for all $i, j \in \{1, ..., n\}$. The diagonal elements $(A_X)_{i,i} = A_X(x_i, x_i)$ are zero. As it does not lead to confusion we use A_X to denote both, the dissimilarity function and its matrix representation. We further denote by $\tilde{\mathcal{N}}$ the set of all possibly asymmetric networks N. Networks $N \in \tilde{\mathcal{N}}$ can have different node sets X as well as different dissimilarities A_X.

The smallest nontrivial networks contain two nodes p and q and two dissimilarities α and β as depicted in Fig. 2.3. Since they appear often throughout, consider the dissimilarity function $A_{p,q}$ with $A_{p,q}(p,q) = \alpha$ and $A_{p,q}(q,p) = \beta$ for some $\alpha, \beta > 0$ and define the *two-node network* $\vec{\Delta}_2(\alpha, \beta)$ with parameters α and β as

$$\vec{\Delta}_2(\alpha, \beta) := (\{p, q\}, A_{p,q}). \tag{2.1}$$

We define a clustering of the set X as a partition P_X; i.e., a collection of sets $P_X = \{B_1, \ldots, B_J\}$ which are pairwise disjoint, $B_i \cap B_j = \emptyset$ for $i \neq j$, and are required to cover X, $\cup_{i=1}^{J} B_i = X$. The sets $B_1, B_2, \ldots B_J$ are called the *blocks* or *clusters* of P_X. We define the *power set* $\mathcal{P}(X)$ of X as the set containing every subset of X, thus $B_i \in \mathcal{P}(X)$ for all i. An equivalence relation $\sim$ on X is a binary operation such that for all $x, x', x'' \in X$ we have that $x \sim x$, $x \sim x'$ if and only if $x' \sim x$, and $x \sim x'$ combined with $x' \sim x''$ implies $x \sim x''$. A partition $P = \{B_1, \ldots, B_J\}$ of X induces and is induced by an equivalence relation $\sim_P$ on X where for all $x, x' \in X$ we have that $x \sim_P x'$ if and only if x and x' belong to the same block B_i for some i. In this first part of the thesis, we focus on hierarchical clustering methods. The output of hierarchical clustering methods is not a single partition P_X but a nested collection D_X of partitions $D_X(\delta)$ of X indexed by a resolution parameter $\delta \geq 0$. In consistency with our previous notation, for a given D_X, we say that two nodes x and x' are equivalent at resolution $\delta \geq 0$ and write $x \sim_{D_X(\delta)} x'$ if and only if nodes x and x' are in the same cluster of $D_X(\delta)$. The nested collection D_X is termed a *dendrogram* and is required to satisfy the following properties:

(D1) Boundary conditions. For $\delta = 0$ the partition $D_X(0)$ clusters each $x \in X$ into a separate singleton and for some δ_0 sufficiently large $D_X(\delta_0)$ clusters all elements of X into a single set,

$$D_X(0) = \Big\{\{x\},\, x \in X\Big\}, \qquad D_X(\delta_0) = \Big\{X\Big\} \quad \text{for some } \delta_0 > 0. \tag{2.2}$$

(D2) Hierarchy. As δ increases clusters can be combined but not separated. I.e., for any $\delta_1 < \delta_2$ any pair of points x, x' for which $x \sim_{D_X(\delta_1)} x'$ must be $x \sim_{D_X(\delta_2)} x'$.

(D3) Right continuity. For all $\delta \geq 0$, there exists $\epsilon > 0$ such that $D_X(\delta) = D_X(\delta')$ for all $\delta' \in [\delta, \delta + \epsilon]$.

The second boundary condition in (2.2) together with (D2) imply that we must have $D_X(\delta) = \big\{X\big\}$ for all $\delta \geq \delta_0$. We denote by $[x]_\delta$ the equivalence class to which the node $x \in X$ belongs at resolution δ, i.e. $[x]_\delta := \{x' \in X \,\big|\, x \sim_{D_X(\delta)} x'\}$. From requirement (D1) we must have that $[x]_0 = x$ and $[x]_{\delta_0} = \{X\}$ for all $x \in X$.

The interpretation of a dendrogram is that of a structure which yields different clusterings at different resolutions. At resolution $\delta = 0$ each point is in a cluster of its own. As the resolution parameter δ increases, nodes start forming clusters. According to condition

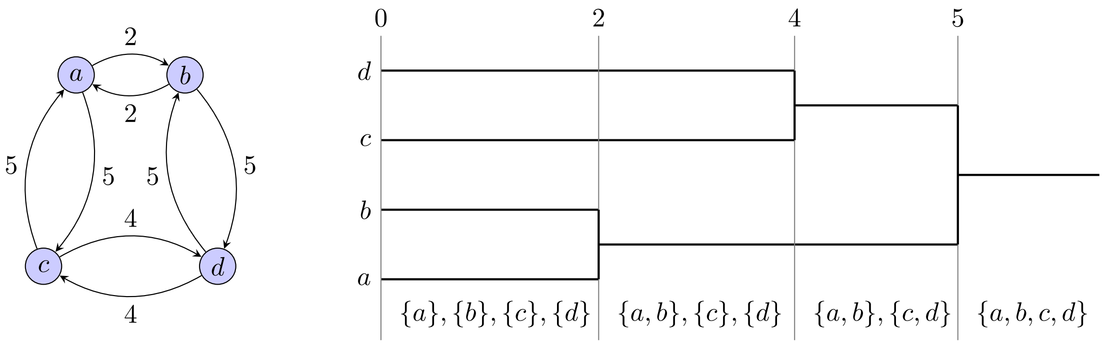


Figure 2.1: Single linkage dendrogram for a symmetric network. Dendrograms are trees representing the outcome of hierarchical clustering algorithms. The single linkage dendrogram as defined by (2.8) for the network on the left is shown on the right. For resolutions $\delta < 2$ each node is in a separate partition, for $2 \leq \delta < 4$ nodes a and b form the cluster $\{a, b\}$, for $4 \leq \delta < 5$ we add the cluster $\{c, d\}$, and for $5 \leq \delta$ all nodes are part of a single cluster.

(D2), nodes become ever more clustered since once they join together in a cluster, they stay together in the same cluster for all larger resolutions. Eventually, the resolutions become coarse enough so that all nodes become members of the same cluster and stay that way as δ keeps increasing. A dendrogram can be represented as a tree; see e.g. Fig. 2.2. Its root represents $D_X(\delta_0)$ with all nodes clustered together and the leaves represent $D_X(0)$ with each node separately clustered. Forks in the tree happen at resolutions δ at which the partitions become finer – or coarser if we move from leaves to root.

Denoting by $\mathcal{D}$ the space of all dendrograms we define a hierarchical clustering method as a function

$$\mathcal{H} : \tilde{\mathcal{N}} \to \mathcal{D}, \tag{2.3}$$

from the space of networks $\tilde{\mathcal{N}}$ to the space of dendrograms $\mathcal{D}$ such that the underlying space X is preserved. For the network $N_X = (X, A_X)$ we denote by $D_X = \mathcal{H}(N_X)$ the output of clustering method $\mathcal{H}$.

In the description of hierarchical clustering methods $\mathcal{H}$ in general, and in those derived on this thesis in particular, the concepts of *path*, *path cost*, and *minimum path cost* are important. Given a network (X, A_X) and $x, x' \in X$, a path $P_{xx'}$ is an *ordered* sequence of nodes in X,

$$P_{xx'} = [x = x_0, x_1, \ldots, x_{l-1}, x_l = x'], \tag{2.4}$$

which starts at x and finishes at x'. We say that $P_{xx'}$ links or connects x to x'. Given two paths $P_{xx'} = [x = x_0, x_1, ..., x_l = x']$ and $P_{x'x''} = [x' = x'_0, x'_1, ..., x'_{l'} = x'']$ such that the end point x' of the first one coincides with the starting point of the second one we define the *concatenated path* $P_{xx'} \uplus P_{x'x''}$ as

$$P_{xx'} \uplus P_{x'x''} := [x = x_0, \ldots, x_l = x' = x'_0, \ldots, x'_{l'} = x'']. \tag{2.5}$$

It follows from (2.5) that the concatenation operation $\uplus$ is associative in that $\big[P_{xx'} \uplus P_{x'x''}\big] \uplus P_{x''x'''} = P_{xx'} \uplus \big[P_{x'x''} \uplus P_{x''x'''}\big]$. Observe that the path $P_{xx'} = [x = x_0, x_1, \ldots, x_{l-1}, x_l = x']$ and its reverse $P_{x'x} = [x' = x_l, x_{l-1}, \ldots, x_1, x_0 = x]$ are different entities even if the intermediate hops are the same.

The *links* of a path are the edges connecting its consecutive nodes in the direction imposed by the path. We define the *cost* of a given path $P_{xx'} := [x = x_0, \ldots, x_l = x']$ as

$$\max_{i | x_i \in P_{xx'}} A_X(x_i, x_{i+1}), \tag{2.6}$$

i.e., the maximum dissimilarity encountered when traversing its links in order. The directed minimum path cost $\tilde{u}^*_X(x, x')$ between x and x' is then defined as the minimum cost among all the paths connecting x to x',

$$\tilde{u}^*_X(x, x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} A_X(x_i, x_{i+1}). \tag{2.7}$$

In asymmetric networks the minimum path costs $\tilde{u}^*_X(x, x')$ and $\tilde{u}^*_X(x', x)$ are different in general but they are equal on symmetric networks. In this latter case, the costs $\tilde{u}^*_X(x, x') = \tilde{u}^*_X(x', x)$ are instrumental in the definition of the single linkage dendrogram [10]. Indeed, for resolution δ, single linkage makes x and x' part of the same cluster if and only if they can be linked through a path of cost not exceeding δ. Formally, the equivalence classes at resolution δ in the single linkage dendrogram SL_X over a symmetric network (X, A_X) are defined by

$$x \sim_{\mathrm{SL}_X(\delta)} x' \iff \tilde{u}^*_X(x, x') = \tilde{u}^*_X(x', x) \leq \delta. \tag{2.8}$$

Recall that in (2.8) the costs $\tilde{u}^*_X(x, x') = \tilde{u}^*_X(x', x)$ are equal because we are assuming metric dissimilarities, which in particular implies $A_X(x_i, x_{i+1}) = A_X(x_{i+1}, x_i)$ in the paths $P_{xx'}$ in (2.7). Fig. 2.1 shows a symmetric network along with the corresponding single linkage dendrogram. For resolutions $\delta < 2$ the dendrogram partitions are $D_X(\delta) = \big\{\{a\}, \{b\}, \{c\}, \{d\}\big\}$. For resolutions $2 \leq \delta < 4$ nodes a and b get clustered together to yield $D_X(\delta) = \big\{\{a, b\}, \{c\}, \{d\}\big\}$. As we keep coarsening the resolution, c and d also get clustered together yielding $D_X(\delta) = \big\{\{a, b\}, \{c, d\}\big\}$ for resolutions $4 \leq \delta < 5$. For $5 \leq \delta$ all nodes are part of a single cluster, $D_X(\delta) = \big\{\{a, b, c, d\}\big\}$ because we can build paths between any pair of nodes incurring maximum cost smaller than or equal to δ.

We further define a *loop* as a path of the form P_{xx} for some $x \in X$ such that P_{xx} contains at least one node other than x. Since a loop is a particular case of a path, the cost of a loop is given by (2.6). Furthermore, consistently with (2.7), we define the *minimum loop cost* $\mathrm{mlc}(X, A_X)$ of a network (X, A_X) as the minimum across all possible loops of each individual loop cost,

$$\mathrm{mlc}(X, A_X) = \min_{x} \min_{P_{xx}} \max_{i | x_i \in P_{xx}} A_X(x_i, x_{i+1}), \tag{2.9}$$

where, we recall, P_{xx} contains at least one node different from x. Another relevant property of a network (X, A_X) is the *separation* of the network $\mathrm{sep}(X, A_X)$ which we define as its minimum positive dissimilarity,

$$\mathrm{sep}(X, A_X) := \min_{x \neq x'} A_X(x, x'). \tag{2.10}$$

Notice that from (2.9) and (2.10) we must have

$$\mathrm{sep}(X, A_X) \leq \mathrm{mlc}(X, A_X). \tag{2.11}$$

Further observe that in the particular case of networks with symmetric dissimilarities the two quantities coincide, i.e., $\mathrm{sep}(X, A_X) = \mathrm{mlc}(X, A_X)$.

2.1.1 Dendrograms as ultrametrics

Dendrograms are convenient graphical representations but otherwise cumbersome to handle. A mathematically more suitable representation is obtained when one identifies dendrograms with finite *ultrametric* spaces. An ultrametric defined on the set X is a metric function $u_X : X \times X \to \mathbb{R}_+$ that satisfies a stronger triangle inequality as we formally define next.

Definition 1 *Given a set X, an ultrametric $u_X : X \times X \to \mathbb{R}_+$ is a function from pairs of elements to the non-negative reals satisfying the following properties for all $x, x', x'' \in X$:*

(i) Identity: $u_X(x, x') = 0$ if and only if $x = x'$.

(ii) Symmetry: $u_X(x, x') = u_X(x', x)$.

(iii) Strong triangle inequality:

$$u_X(x, x'') \leq \max\Big(u_X(x, x'), u_X(x', x'')\Big). \tag{2.12}$$

Since (2.12) implies the usual triangle inequality $u_X(x, x'') \leq u_X(x, x') + u_X(x', x'')$ for all $x, x', x'' \in X$, ultrametric spaces are particular cases of metric spaces.

Our interest in ultrametrics stems from the fact that it is possible to establish a structure preserving bijective mapping between dendrograms and ultrametrics as proved by the following construction and theorem; see also Fig. 2.2.

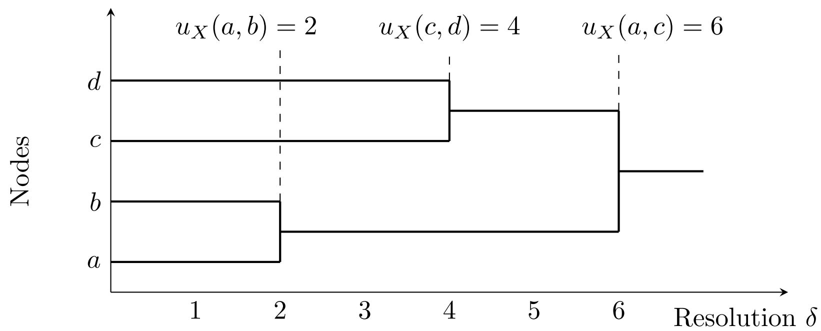


Figure 2.2: Equivalence of dendrograms and ultrametrics. Given a dendrogram D_X define the function $u_X(x,x') := \min\big\{\delta \geq 0 \,|\, x \sim_{D_X(\delta)} x'\big\}$. This function is an ultrametric because it satisfies the identity property, the strong triangle inequality (2.12) and is symmetric.

Consider the map $\Psi : \mathcal{D} \to \mathcal{U}$ from the space of dendrograms to the space of networks endowed with ultrametrics, defined as follows: for a given dendrogram D_X over the finite set X write $\Psi(D_X) = (X, u_X)$, where we define $u_X(x,x')$ for all $x, x' \in X$ as the smallest resolution at which x and x' are clustered together $u_X(x,x') := \min\{\delta \geq 0 \,|\, x \sim_{D_X(\delta)} x'\}$. We also consider the map $\Upsilon : \mathcal{U} \to \mathcal{D}$ constructed as follows: for a given ultrametric u_X on the finite set X and each $\delta \geq 0$ define the relation $\sim_{u_X(\delta)}$ on X as $x \sim_{u_X(\delta)} x' \iff u_X(x,x') \leq \delta$. Further define $D_X(\delta) := \big\{X \mod \sim_{u_X(\delta)}\big\}$ and $\Upsilon(X, u_X) := D_X$.

Theorem 1 ([10]) *The maps Ψ and Υ are both well defined. Furthermore, $\Psi \circ \Upsilon$ is the identity on $\mathcal{U}$ and $\Upsilon \circ \Psi$ is the identity on $\mathcal{D}$.*

Given the equivalence between dendrograms and ultrametrics established by Theorem 1 we can regard hierarchical clustering methods $\mathcal{H}$ as inducing ultrametrics in node sets X based on dissimilarity functions A_X. However, ultrametrics are particular cases of dissimilarity functions. Thus, we can reinterpret the method $\mathcal{H}$ as a map [cf. (2.3)]

$$\mathcal{H} : \tilde{\mathcal{N}} \to \mathcal{U} \tag{2.13}$$

mapping the space of networks $\tilde{\mathcal{N}}$ to the space $\mathcal{U} \subset \tilde{\mathcal{N}}$ of networks endowed with ultrametrics. For all $x, x' \in X$, the ultrametric value $u_X(x,x')$ induced by $\mathcal{H}$ is the minimum resolution at which x and x' are co-clustered by $\mathcal{H}$. Observe that the outcome of a hierarchical clustering method defines an ultrametric in the set X even when the original data does not correspond to a metric, as is the case of asymmetric networks. We say that two methods $\mathcal{H}_1$ and $\mathcal{H}_2$ are *equivalent*, and we write $\mathcal{H}_1 \equiv \mathcal{H}_2$, if and only if $\mathcal{H}_1(N) = \mathcal{H}_2(N)$ for all $N \in \tilde{\mathcal{N}}$.

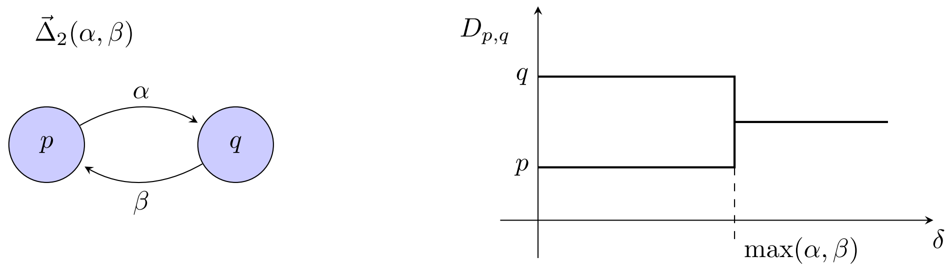


Figure 2.3: Axiom of Value. Nodes in a two-node network cluster at the minimum resolution at which both can influence each other.

2.2 Axioms of Value and Transformation

To study hierarchical clustering methods on asymmetric networks we start from intuitive notions that we translate into the Axioms of Value and Transformation discussed in this section.

The Axiom of Value is obtained from considering the two-node network $\vec{\Delta}_2(\alpha,\beta)$ defined in Section 2.1 and depicted in Fig. 2.3. We say that node x is able to influence node x' at resolution δ if the dissimilarity from x to x' is not greater than δ. In two-node networks, our intuition dictates that a cluster is formed if nodes p and q are able to influence each other. This implies that the output dendrogram should be such that p and q are part of the same cluster at resolutions $\delta \geq \max(\alpha,\beta)$ that allow direct mutual influence. Conversely, we expect nodes p and q to be in separate clusters at resolutions $0 \leq \delta < \max(\alpha,\beta)$ that do *not* allow for mutual influence. At resolutions $\delta < \min(\alpha,\beta)$ there is no influence between the nodes and at resolutions $\min(\alpha,\beta) \leq \delta < \max(\alpha,\beta)$ there is unilateral influence from one node over the other. In either of the latter two cases the nodes are different in nature. If we think of dissimilarities as, e.g., trust, it means one node is trustworthy whereas the other is not. If we think of the network as a Markov path, at resolutions $0 \leq \delta < \max(\alpha,\beta)$ the states are different singleton equivalence classes – one of the states would be transient and the other one absorbent. Given that, according to (2.13), a hierarchical clustering method is a map $\mathcal{H}$ from networks to ultrametrics, we formalize this intuition as the following requirement on the set of admissible maps:

(A1) Axiom of Value. The ultrametric $(\{p,q\}, u_{p,q}) = \mathcal{H}(\vec{\Delta}_2(\alpha,\beta))$ produced by $\mathcal{H}$ applied to the two-node network $\vec{\Delta}_2(\alpha,\beta)$ satisfies $u_{p,q}(p,q) = \max(\alpha,\beta)$.

Clustering nodes p and q together at resolution $\delta = \max(\alpha,\beta)$ is somewhat arbitrary, as any monotone increasing function of $\max(\alpha,\beta)$ would be admissible. As a value claim, however, it means that the clustering resolution parameter δ is expressed in the same units as the elements of the dissimilarity function.

The second restriction on the space of allowable methods $\mathcal{H}$ formalizes our expectations for the behavior of $\mathcal{H}$ when confronted with a transformation of the underlying set X

and the dissimilarity function A_X; see Fig. 2.4. Consider networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ and denote by $D_X = \mathcal{H}(X, A_X)$ and $D_Y = \mathcal{H}(Y, A_Y)$ the corresponding dendrogram outputs. If we map all the nodes of the network $N_X = (X, A_X)$ into nodes of the network $N_Y = (Y, A_Y)$ in such a way that no pairwise dissimilarity is increased we expect the latter network to be more clustered than the former at any given resolution. Intuitively, nodes in N_Y are more capable of influencing each other, thus, clusters should be formed more easily. In terms of the respective dendrograms we expect that nodes co-clustered at resolution δ in D_X are mapped to nodes that are also co-clustered at this resolution in D_Y. In order to formalize this notion, we introduce the concept of a *dissimilarity reducing map*. Given two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$, map $\phi : X \to Y$ is dissimilarity reducing if it holds that $A_X(x, x') \geq A_Y(\phi(x), \phi(x'))$ for all $x, x' \in X$.

The Axiom of Transformation that we introduce next is a formal statement of the intuition described above:

> *(A2) Axiom of Transformation.* Consider two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ and a dissimilarity reducing map $\phi : X \to Y$, i.e. a map ϕ such that for all $x, x' \in X$ it holds that $A_X(x, x') \geq A_Y(\phi(x), \phi(x'))$. Then, for all $x, x' \in X$, the output ultrametrics $(X, u_X) = \mathcal{H}(X, A_X)$ and $(Y, u_Y) = \mathcal{H}(Y, A_Y)$ satisfy
>
> $$u_X(x, x') \geq u_Y(\phi(x), \phi(x')). \tag{2.14}$$

We say that a hierarchical clustering method $\mathcal{H}$ is admissible with respect to (A1) and (A2), or *admissible* for short, if it satisfies axioms (A1) and (A2). Axiom (A1) states that units of the resolution parameter δ are the same units of the elements of the dissimilarity function. Axiom (A2) states that if we reduce dissimilarities, clusters may be combined but cannot be separated.

For the particular case of symmetric networks (X, A_X) we defined the single linkage dendrogram SL_X through the equivalence relations in (2.8). According to Theorem 1 this dendrogram is equivalent to an ultrametric space that we denote by (X, u_X^{SL}). More specifically, as is well known [10], the single linkage ultrametric u_X^{SL} in symmetric networks is given by

$$u_X^{\mathrm{SL}}(x, x') \;=\; \tilde{u}_X^*(x, x') = \tilde{u}_X^*(x', x) \;=\; \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} A_X(x_i, x_{i+1}), \tag{2.15}$$

where we also used (2.7) to write the last equality.

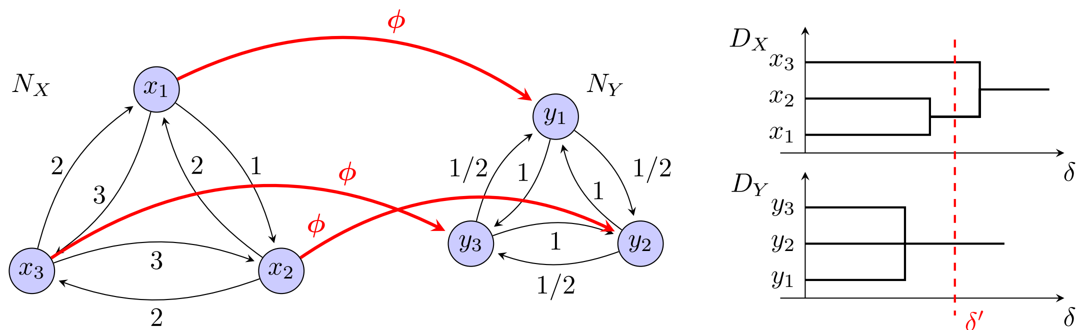


Figure 2.4: Axiom of Transformation. If the network N_X can be mapped to the network N_Y using a dissimilarity reducing map ϕ, then for every resolution δ nodes clustered together in $D_X(\delta)$ must also be clustered in $D_Y(\delta)$. E.g., since points x_1 and x_2 are clustered together at resolution δ', their image through ϕ, i.e. $y_1 = \phi(x_1)$ and $y_2 = \phi(x_2)$, must also be clustered together at this resolution.

2.3 Influence modalities

The Axiom of Value states that for two nodes to belong to the same cluster they have to be able to exercise mutual influence on each other. When we consider a network with more than two nodes the concept of mutual influence is more difficult because it is possible to have direct influence as well as indirect paths of influence through other nodes. In this section we introduce two intuitive notions of mutual influence in networks of arbitrary size and show that they can be derived from the Axioms of Value and Transformation. Besides their intrinsic value, these influence modalities are important for later developments in this thesis; see, e.g. the proof of Theorem 4.

Consider first the intuitive notion that for two nodes to be part of a cluster there has to be a way for each of them to exercise influence on the other, either directly or indirectly. To formalize this idea, recall the concept of minimum loop cost (2.9) which we exemplify in Fig. 2.5. For this network, the loops $[a, b, a]$ and $[b, a, b]$ have maximum cost 2 corresponding to the link (b, a) in both cases. All other two-node loops have cost 3. All of the counterclockwise loops, e.g., $[a, c, b, a]$, have cost 3 and any of the clockwise loops have cost 1. Thus, the minimum loop cost of this network is $\text{mlc}(X, A_X) = 1$.

For resolutions $0 \leq \delta < \text{mlc}(X, A_X)$ it is impossible to find paths of mutual influence with maximum cost smaller than δ between any pair of points. Indeed, suppose we can link x to x' with a path of maximum cost smaller than δ, and also link x' to x with a path having the same property. Then, we can form a loop with cost smaller than δ by concatenating these two paths. Thus, the intuitive notion that clusters cannot form at resolutions for which it is impossible to observe mutual influence can be translated into the requirement that no clusters can be formed at resolutions $0 \leq \delta < \text{mlc}(X, A_X)$. In terms of ultrametrics, this implies that it must be $u_X(x, x') \geq \text{mlc}(X, A_X)$ for any pair of different nodes $x, x' \in X$ as we formally state next:

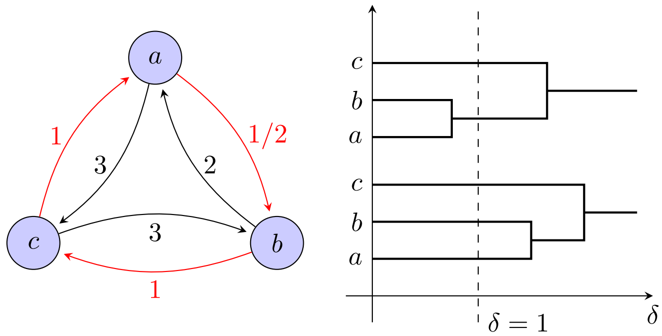


Figure 2.5: Property of Influence. No clusters can be formed at resolutions for which it is impossible to form influence loops. Here, the loop of minimum cost is formed by circling the network clockwise where the maximum cost encountered is $A_X(b, c) = A_X(c, a) = 1$. The top dendrogram is an invalid outcome because it has a and b clustering together at resolution $\delta < 1$. The bottom dendrogram satisfies the Property of Influence (P1), [cf (2.16)].

(P1) Property of Influence. For any network $N_X = (X, A_X)$ the output ultrametric $(X, u_X) = \mathcal{H}(X, A_X)$ corresponding to the application of hierarchical clustering method $\mathcal{H}$ is such that the ultrametric $u_X(x, x')$ between any two distinct points x and x' cannot be smaller than the minimum loop cost $\text{mlc}(X, A_X)$ [cf. (2.9)] of the network

$$u_X(x, x') \geq \text{mlc}(X, A_X) \qquad \text{for all} \quad x \neq x'. \tag{2.16}$$

Since for the network in Fig. 2.5 the minimum loop cost is $\text{mlc}(X, A_X) = 1$, then the Property of Influence implies that $u_X(x, x') \geq \text{mlc}(X, A_X) = 1$ for any pair of nodes $x \neq x'$. Equivalently, the output dendrogram is such that for resolutions $\delta < \text{mlc}(X, A_X) = 1$ each node is in its own block. Observe that (P1) does not imply that a cluster with more than one node *is* formed at resolution $\delta = \text{mlc}(X, A_X)$ but states that achieving this minimum resolution is a necessary condition for the formation of clusters.

A second intuitive statement about influence in networks of arbitrary size comes in the form of the *Extended Axiom of Value*. To introduce this concept define a family of canonical asymmetric networks

$$\vec{\Delta}_n(\alpha, \beta) = (\{1, \ldots, n\}, A_{n,\alpha,\beta}), \tag{2.17}$$

with $\alpha, \beta > 0$ where the underlying set $\{1, \ldots, n\}$ is the set of the first n natural numbers and the dissimilarity function $A_{n,\alpha,\beta}$ between points i and j depends on whether $i > j$; see Fig. 2.6. For points $i < j$ we make the dissimilarity $A_{n,\alpha,\beta}(i, j) = \alpha$ whereas for points $i > j$ we have $A_{n,\alpha,\beta}(i, j) = \beta$. Or, in matrix form,

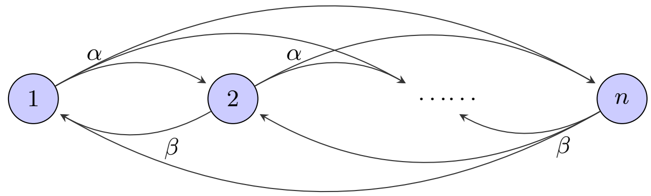


Figure 2.6: Canonical network $\vec{\Delta}_n(\alpha,\beta)$ for the Extended Axiom of Value. Edges from a node to another node identified with a higher number (i.e. arrows pointing to the right) have weight α, whereas edges going to nodes identified with lower numbers (i.e. arrows pointing to the left) have weight β. All admissible methods $\mathcal{H}$ cluster the n nodes together at resolution $\max(\alpha,\beta)$.

$$A_{n,\alpha,\beta} := \begin{pmatrix} 0 & \alpha & \alpha & \alpha & \cdots & \alpha \\ \beta & 0 & \alpha & \alpha & \cdots & \alpha \\ \beta & \beta & 0 & \alpha & \cdots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta & \beta & \beta & \cdots & 0 & \alpha \\ \beta & \beta & \beta & \cdots & \beta & 0 \end{pmatrix}. \tag{2.18}$$

In the network $\vec{\Delta}_n(\alpha,\beta)$ all pairs of nodes have dissimilarities α in one direction and β in the other direction. This symmetry entails that all nodes should cluster together at the same resolution, and the requirement of mutual influence along with consistency with the Axiom of Value entails that this resolution should be $\max(\alpha,\beta)$. Before formalizing this definition notice that having clustering outcomes that depend on the ordering of the nodes in the space $\{1,\ldots,n\}$ is not desirable. Thus, we consider a permutation $\Pi = (\pi_1,\pi_2,\ldots,\pi_n)$ of $\{1,2,\ldots,n\}$ and the action $\Pi(A)$ of Π on a $n\times n$ matrix A, which we define by $(\Pi(A))_{i,j} = A_{\pi_i,\pi_j}$ for all i and j. Define now the network $\vec{\Delta}_n(\alpha,\beta,\Pi) := (\{1,\ldots,n\},\Pi(A_{n,\alpha,\beta}))$ with underlying set $\{1,\ldots,n\}$ and dissimilarity matrix given $\Pi(A_{n,\alpha,\beta})$. With this definition we can now formally introduce the Extended Axiom of Value as follows:

(A1') Extended Axiom of Value. Consider the network $\vec{\Delta}_n(\alpha,\beta,\Pi) = (\{1,\ldots,n\},\Pi(A_{n,\alpha,\beta}))$. Then, for all indices $n\in\mathbb{N}$, constants $\alpha,\beta>0$, and permutations Π, the outcome $(\{1,\ldots,n\},u) = \mathcal{H}\big(\vec{\Delta}_n(\alpha,\beta,\Pi)\big)$ of hierarchical clustering method $\mathcal{H}$ applied to the network $\vec{\Delta}_n(\alpha,\beta,\Pi)$ satisfies, for all pairs of nodes $i\neq j$,

$$u(i,j) = \max(\alpha,\beta). \tag{2.19}$$

Observe that the Axiom of Value (A1) is subsumed into the Extended Axiom of Value for $n=2$. Further note that the minimum loop cost of the canonical network $\vec{\Delta}_n(\alpha,\beta)$ is $\mathrm{mlc}\big(\vec{\Delta}_n(\alpha,\beta)\big) = \max(\alpha,\beta)$ because forming a loop requires traversing a link while moving right and a link while moving left at least once in Fig. 2.6. Since a permutation of indices does not alter the minimum loop cost of the network we have that

$$\mathrm{mlc}\big(\vec{\Delta}_n(\alpha,\beta,\Pi)\big) = \mathrm{mlc}\big(\vec{\Delta}_n(\alpha,\beta)\big) = \max(\alpha,\beta). \tag{2.20}$$

By the Property of Influence (P1) it follows from (2.20) and (2.16) that for the network $\vec{\Delta}_n(\alpha,\beta,\Pi)$ we must have $u(i,j) \geq \mathrm{mlc}(\vec{\Delta}_n(\alpha,\beta)) = \max(\alpha,\beta)$ for $i \neq j$. By the Extended Axiom of Value (A1') we have $u(i,j) = \max(\alpha,\beta)$ for $i \neq j$, which means that (A1') and (P1) are compatible requirements. We can then think of two alternative axiomatic formulations where admissible methods are required to abide by the Axiom of Transformation (A2), the Property of Influence (P1), and either the (regular) Axiom of Value (A1) or the Extended Axiom of Value (A1') – axiom (A1) and (P1) are compatible because (A1) is a particular case of (A1') which we already argued is compatible with (P1). We will see in the following section that these two alternative axiomatic formulations are equivalent to each other in the sense that a clustering method satisfies one set of axioms if and only if it satisfies the other. We further show that (P1) and (A1') are implied by (A1) and (A2). As a consequence, it follows that both alternative axiomatic formulations are equivalent to simply requiring validity of axioms (A1) and (A2).

2.3.1 Equivalent axiomatic formulations

We begin by proving the equivalence between admissibility with respect to (A1)-(A2) and (A1')-(A2). A proof that methods admissible with respect to (A1') and (A2) satisfy the Property of Influence (P1) is presented next to conclude that (A1)-(A2) imply (P1) as a corollary. Paths $P_{xx'}$ for nodes $x, x' \in X$, as defined in (2.4), and the concept of directed minimum path cost $\tilde{u}^*_X$ in (2.7) are instrumental in the proof of the first result pertaining to the equivalence between admissibility with respect to (A1)-(A2) and (A1')-(A2).

Theorem 2 *Assume the hierarchical clustering method $\mathcal{H}$ satisfies the Axiom of Transformation (A2). Then, $\mathcal{H}$ satisfies the Axiom of Value (A1) if and only if it satisfies the Extended Axiom of Value (A1').*

In proving Theorem 2, we make use of the following lemma.

Lemma 1 *A network $N = (X, A_X)$ and a positive constant δ are given. Then, for any pair of nodes $x, x' \in X$ whose minimum path cost [cf. (2.7)] satisfies*

$$\tilde{u}^*_X(x,x') \geq \delta, \tag{2.21}$$

there exists a partition $P_\delta(x,x') = \{B_\delta(x), B_\delta(x')\}$ of the node set X into blocks $B_\delta(x)$ and

$B_\delta(x')$ with $x \in B_\delta(x)$ and $x' \in B_\delta(x')$ such that for all points $b \in B_\delta(x)$ and $b' \in B_\delta(x')$

$$A_X(b, b') \geq \delta. \tag{2.22}$$

Proof: We prove this result by contradiction. If a partition $P_\delta(x, x') = \{B_\delta(x), B_\delta(x')\}$ with $x \in B_\delta(x)$ and $x' \in B_\delta(x)$ and satisfying (2.22) does not exist for all pairs of points $x, x' \in X$ satisfying (2.21), then there is at least one pair of nodes $x, x' \in X$ satisfying (2.21) such that for *all* partitions of X into two blocks $P = \{B, B'\}$ with $x \in B$ and $x' \in B'$ we can find at least a pair of elements $b_P \in B$ and $b'_P \in B'$ for which

$$A_X(b_P, b'_P) < \delta. \tag{2.23}$$

Begin by considering the partition $P_1 = \{B_1, B'_1\}$ where $B_1 = \{x\}$ and $B'_1 = X \backslash \{x\}$. Since (2.23) is true for all partitions having $x \in B$ and $x' \in B'$ and x is the unique element of B_1, there must exist a node $b'_{P_1} \in B'_1$ such that

$$A_X(x, b'_{P_1}) < \delta. \tag{2.24}$$

Hence, the path $P_{xb'_{P_1}} = [x, b'_{P_1}]$ composed of these two nodes has cost smaller than δ. Moreover, since $\tilde{u}^*_X(x, b'_{P_1})$ represents the minimum cost among all paths $P_{xb'_{P_1}}$ linking x to b'_{P_1}, we can assert that

$$\tilde{u}^*_X(x, b'_{P_1}) \leq A_X(x, b'_{P_1}) < \delta. \tag{2.25}$$

Consider now the partition $P_2 = \{B_2, B'_2\}$ where $B_2 = \{x, b'_{P_1}\}$ and $B'_2 = X \backslash B_2$. From (2.23), there must exist a node $b'_{P_2} \in B'_2$ that satisfies at least one of the two following conditions

$$A_X(x, b'_{P_2}) < \delta, \tag{2.26}$$

$$A_X(b'_{P_1}, b'_{P_2}) < \delta. \tag{2.27}$$

If (2.26) is true, the path $P_{xb'_{P_2}} = [x, b'_{P_2}]$ has cost smaller than δ. If (2.27) is true, we combine the dissimilarity bound with the one in (2.24) to conclude that the path $P_{xb'_{P_2}} = [x, b'_{P_1}, b'_{P_2}]$ has cost smaller than δ. In either case we conclude that there exists a path $P_{xb'_{P_2}}$ linking x to b'_{P_2} whose cost is smaller than δ. Therefore, the minimum path cost must satisfy

$$\tilde{u}^*_X(x, b'_{P_2}) < \delta. \tag{2.28}$$

Repeat the process by considering the partition P_3 with $B_3 = \{x, b'_{P_1}, b'_{P_2}\}$ and $B'_3 = X \backslash B_3$. As we did in arguing (2.26)-(2.27) it must follow from (2.23) that there exists a point b'_{P_3} such that at least one of the dissimilarities $A_X(x, b'_{P_3})$, $A_X(b'_{P_1}, b'_{P_3})$, or $A_X(b'_{P_2}, b'_{P_3})$ is smaller than δ. This observation implies that at least one of the paths $[x, b'_{P_3}]$, $[x, b'_{P_1}, b'_{P_3}]$, $[x, b'_{P_2}, b'_{P_3}]$, or $[x, b'_{P_1}, b'_{P_2}, b'_{P_3}]$ has cost smaller than δ from where it follows that

$$\tilde{u}^*_X(x, b'_{P_3}) < \delta. \tag{2.29}$$

This recursive construction can be repeated $n-1$ times to obtain partitions $P_1, P_2, ..., P_{n-1}$ and corresponding nodes $b'_{P_1}, b'_{P_2}, ...b'_{P_{n-1}}$ such that the minimum path cost satisfies

$$\tilde{u}^*_X(x, b'_{P_i}) < \delta, \qquad \text{for all } i. \tag{2.30}$$

Observe now that the nodes b'_{P_i} are distinct by construction and distinct from x. Since there are n nodes in the network it must be that $x' = b'_{P_k}$ for some $k \in \{1, \ldots, n-1\}$. It then follows from (2.30) that

$$\tilde{u}^*_X(x, x') < \delta. \tag{2.31}$$

This is a contradiction because $x, x' \in X$ were assumed to satisfy (2.21). Thus, the assumption that (2.23) is true for *all* partitions is incorrect. Hence, the claim that there is a partition $P_\delta(x, x') = \{B_\delta(x), B_\delta(x')\}$ satisfying (2.22) must be true. ■

Proof of Theorem 2: To prove that (A1)-(A2) imply (A1')-(A2) let $\mathcal{H}$ be a method that satisfies (A1) and (A2) and denote by $(\{1, 2, \ldots, n\}, u_{n,\alpha,\beta}) = \mathcal{H}(\vec{\Delta}_n(\alpha, \beta, \Pi))$ the output ultrametric resulting of applying $\mathcal{H}$ to the network $\vec{\Delta}_n(\alpha, \beta, \Pi)$ considered in the Extended Axiom of Value (A1'). We want to prove that (A1') is satisfied which means that we have to show that for all indices $n \in \mathbb{N}$, constants $\alpha, \beta > 0$, permutations Π, and points $i \neq j$, we have $u_{n,\alpha,\beta}(i, j) = \max(\alpha, \beta)$. We will do so by showing both

$$u_{n,\alpha,\beta}(i, j) \;\leq\; \max(\alpha, \beta), \tag{2.32}$$

$$u_{n,\alpha,\beta}(i, j) \;\geq\; \max(\alpha, \beta), \tag{2.33}$$

for all $n \in \mathbb{N}$, $\alpha, \beta > 0$, Π, and $i \neq j$.

To prove (2.32) define a symmetric two node network $N_{p,q} = (\{p, q\}, A_{p,q})$ where $A_{p,q}(p, q) = A_{p,q}(q, p) = \max(\alpha, \beta)$ and denote by $\big(\{p, q\}, u_{p,q}\big) = \mathcal{H}(N_{p,q})$ the outcome of method $\mathcal{H}$ when applied to $N_{p,q}$. Since the method $\mathcal{H}$ abides by (A1) we must have

$$u_{p,q}(p, q) = \max\big(\max(\alpha, \beta), \max(\alpha, \beta)\big) = \max(\alpha, \beta). \tag{2.34}$$

Consider now the map $\phi_{i,j} : \{p, q\} \to \{1, \ldots, n\}$ from the two node network $N_{p,q}$ to the permuted canonical network $\vec{\Delta}_n(\alpha, \beta, \Pi)$ where $\phi_{i,j}(p) = i$ and $\phi_{i,j}(q) = j$. Since dissimilarities in $\vec{\Delta}_n(\alpha, \beta, \Pi)$ are either α or β and the dissimilarities in $N_{p,q}$ are $\max(\alpha, \beta)$ it follows that the map $\phi_{i,j}$ is dissimilarity reducing regardless of the particular values of i and j. Since the method $\mathcal{H}$ was assumed to satisfy (A2) as well, we must have

$$u_{p,q}(p,q) \geq u_{n,\alpha,\beta}\big(\phi_{i,j}(p), \phi_{i,j}(q)\big) = u_{n,\alpha,\beta}(i,j). \tag{2.35}$$

The inequality in (2.32) follows form substituting (2.34) into (2.35).

In order to show inequality (2.33), pick two arbitrary distinct nodes $i, j \in \{1, \ldots, n\}$ in the node set of $\vec{\Delta}_n(\alpha, \beta, \Pi)$. Denote by P_{ij} and P_{ji} two minimizing paths in the definition (2.7) of the directed minimum path costs $\tilde{u}^*_{n,\alpha,\beta}(i,j)$ and $\tilde{u}^*_{n,\alpha,\beta}(j,i)$ respectively. Observe that at least one of the following two inequalities must be true

$$\tilde{u}^*_{n,\alpha,\beta}(i,j) \geq \max(\alpha, \beta), \tag{2.36}$$

$$\tilde{u}^*_{n,\alpha,\beta}(j,i) \geq \max(\alpha, \beta). \tag{2.37}$$

Indeed, if both (2.36) and (2.37) were false, the concatenation of P_{ij} and P_{ji} would form a loop $P_{ii} = P_{ij} \uplus P_{ji}$ of cost strictly less than $\max(\alpha, \beta)$. This cannot be true because $\max(\alpha, \beta)$ is the minimum loop cost of the network $\vec{\Delta}_n(\alpha, \beta, \Pi)$ as we already showed in (2.20).

Without loss of generality assume (2.36) is true and consider $\delta = \max(\alpha, \beta)$. By Lemma 1 we are therefore guaranteed to find a partition of the node set $\{1, \ldots, n\}$ into two blocks $B_\delta(i)$ and $B_\delta(j)$ with $i \in B_\delta(i)$ and $j \in B_\delta(j)$ such that for all $b \in B_\delta(i)$ and $b' \in B_\delta(j)$ it holds that

$$\Pi(A_{n,\alpha,\beta})(b, b') \geq \delta = \max(\alpha, \beta). \tag{2.38}$$

Define a two node network $N_{r,s} = (\{r, s\}, A_{r,s})$ where $A_{r,s}(r,s) = \max(\alpha, \beta)$ and $A_{r,s}(s,r) = \min(\alpha, \beta)$ and denote by $(\{r, s\}, u_{r,s}) = \mathcal{H}(N_{r,s})$ the outcome of applying the method $\mathcal{H}$ to the network $N_{r,s}$. Since the method $\mathcal{H}$ satisfies (A1) we must have

$$u_{r,s}(r,s) = \max\big(\max(\alpha, \beta), \min(\alpha, \beta)\big) = \max(\alpha, \beta). \tag{2.39}$$

Consider the map $\phi'_{i,j} : \{1, \ldots, n\} \to \{r, s\}$ such that $\phi'_{i,j}(b) = r$ for all $b \in B_\delta(i)$ and $\phi'_{i,j}(b') = s$ for all $b' \in B_\delta(j)$. The map $\phi'_{i,j}$ is dissimilarity reducing because

$$\Pi(A_{n,\alpha,\beta})(k,l) \geq A_{r,s}(\phi'_{i,j}(k), \phi'_{i,j}(l)), \tag{2.40}$$

for all $k, l \in \{1, \ldots, n\}$. To see the validity of (2.40) consider three different possible cases. If k and l belong both to the same block, i.e., either $k, l \in B_\delta(i)$ or $k, l \in B_\delta(j)$, then $\phi'_{i,j}(k) = \phi'_{i,j}(l)$ and the dissimilarity $A_{r,s}(\phi'_{i,j}(k), \phi'_{i,j}(l)) = 0$ which cannot exceed the nonnegative $\Pi(A_{n,\alpha,\beta})(k,l)$. If $k \in B_\delta(j)$ and $l \in B_\delta(i)$ it holds that $A_{r,s}(\phi'_{i,j}(k), \phi'_{i,j}(l)) = A_{r,s}(s,r) = \min(\alpha, \beta)$ which cannot exceed $\Pi(A_{n,\alpha,\beta})(k,l)$ which is either equal to α or β.

If $k \in B_\delta(i)$ and $l \in B_\delta(j)$, then we have $A_{r,s}(\phi'_{i,j}(k), \phi'_{i,j}(l)) = A_{r,s}(r,s) = \max(\alpha, \beta)$ but we also have $\Pi(A_{n,\alpha,\beta})(k,l) = \max(\alpha,\beta)$ as it follows by taking $b = k$ and $b' = l$ in (2.38).

Since $\mathcal{H}$ satisfies the Axiom of Transformation (A2) and the map $\phi'_{i,j}$ is dissimilarity reducing we must have

$$u_{n,\alpha,\beta}(i,j) \geq u_{r,s}\big(\phi'_{i,j}(i), \phi'_{i,j}(j)\big) = u_{r,s}(r,s). \tag{2.41}$$

Substituting (2.39) in (2.41) we obtain the inequality in (2.33). Combining this result with the validity of (2.32) that we already established it follows that $u_{n,\alpha,\beta}(i,j) = \max(\alpha,\beta)$ for all $n \in \mathbb{N}$, $\alpha, \beta > 0$, Π, and $i \neq j$. Thus, admissibility with respect to (A1)-(A2) implies admissibility with respect to (A1')-(A2). That admissibility with respect to (A1')-(A2) implies admissibility with respect to (A1)-(A2) is immediate because (A1) is a particular case of (A1'). ■

The Extended Axiom of Value (A1') is stronger than the (regular) Axiom of Value (A1). However, Theorem 2 shows that when considered together with the Axiom of Transformation (A2), both Axioms of Value are equivalent in the restrictions they impose in the set of admissible clustering methods. In the following theorem we show that the Property of Influence (P1) can be derived from axioms (A1') and (A2).

Theorem 3 *If a clustering method $\mathcal{H}$ satisfies the Axioms of Extended Value (A1') and Transformation (A2) then it satisfies the Property of Influence (P1).*

The following lemma is instrumental towards the proof of Theorem 3.

Lemma 2 *Let $N = (X, A_X)$ be an arbitrary network with n nodes and $\vec{\Delta}_n(\alpha,\beta)$ be the canonical network in* (2.18) *with $0 < \alpha \leq \mathrm{sep}(X, A_X)$ [cf.* (2.10)*] and $\beta = \mathrm{mlc}(X, A_X)$ [cf.* (2.9)*]. Then, there exists a bijective map $\phi : X \to \{1, \ldots, n\}$ such that*

$$A_X(x,x') \geq A_{n,\alpha,\beta}(\phi(x), \phi(x')), \tag{2.42}$$

for all $x, x' \in X$.

Proof: To construct the map ϕ consider the function $F : X \to \mathcal{P}(X)$ from the node set X to its power set $\mathcal{P}(X)$ such that

$$F(x) := \{x' \in X \,|\, x' \neq x \;,\; A_X(x',x) < \beta\}, \tag{2.43}$$

for all $x \in X$. Having $r \in F(s)$ for some $r, s \in X$ implies that $A_X(r,s) < \beta = \mathrm{mlc}(X, A_X)$. An important observation is that we must have a node $x \in X$ whose F-image is empty. Otherwise, pick a node $x_n \in X$ and construct the path $[x_0, x_1, \ldots, x_n]$ where the ith element of the path x_{i-1} is in the F-image of x_i. From the definition of the map F it follows that all dissimilarities along this path satisfy $A_X(x_{i-1}, x_i) < \beta = \mathrm{mlc}(X, A_X)$. But since the path $[x_0, x_1, \dots, x_n]$ contains $n+1$ elements, at least one node must be repeated. Hence, we have found a loop for which all dissimilarities are bounded above by $\beta = \mathrm{mlc}(X, A_X)$, which is impossible because it contradicts the definition of the minimum loop cost in (2.9). We can then find a node x_{i_1} for which $F(x_{i_1}) = \emptyset$. Fix $\phi(x_{i_1}) = 1$.

Select now a node $x_{i_2} \neq x_{i_1}$ whose F-image is either $\{x_{i_1}\}$ or $\emptyset$, which we write jointly as $F(x_{i_2}) \subseteq \{x_{i_1}\}$. Such a node must exist, otherwise, pick a node $x_{n-1} \in X \backslash \{x_{i_1}\}$ and construct the path $[x_0, x_1, \dots, x_{n-1}]$ where $x_{i-1} \in F(x_i) \backslash \{x_{i_1}\}$, i.e. x_{i-1} is in the F-image of x_i and $x_{i-1} \neq \{x_{i_1}\}$. Since the path $[x_0, x_1, ..., x_{n-1}]$ contains n elements from the set $X \backslash \{x_{i_1}\}$ of cardinality $n-1$, at least one node must be repeated. Hence, we have found a loop where all dissimilarities between consecutive nodes satisfy $A_X(x_{i-1}, x_i) < \beta = \mathrm{mlc}(X, A_X)$, contradicting the definition of minimum loop cost. We can then find a node $x_{i_2} \neq x_{i_1}$ for which $F(x_{i_2}) \subseteq \{x_{i_1}\}$. Fix $\phi(x_{i_2}) = 2$.

Repeat this process k times so that at step k we have $\phi(x_{i_k}) = k$ for a node $x_{i_k} \notin \{x_{i_1}, x_{i_2}, \dots x_{i_{k-1}}\}$ whose F-image is a subset of the nodes already picked, that is

$$F(x_{i_k}) \subseteq \{x_{i_1}, x_{i_2}, \dots x_{i_{k-1}}\}. \tag{2.44}$$

This node must exist, otherwise, we could start with a node $x_{n-k+1} \in X \backslash \{x_{i_1}, x_{i_2}, \dots x_{i_{k-1}}\}$ and construct a path $[x_0, x_1, \dots, x_{n-k+1}]$ where $x_{i-1} \in F(x_i) \backslash \{x_{i_1}, x_{i_2}, \dots x_{i_{k-1}}\}$ and arrive to the same contradiction as for the case $k = 2$.

Since all the nodes x_{i_k} are different, the map ϕ with $\phi(x_{i_k}) = k$ is bijective. By construction, ϕ is such that for all $l > k$, $x_{i_l} \notin F(x_{i_k})$. From (2.43), this implies that the dissimilarity from x_{i_l} to x_{i_k} must satisfy

$$A_X(x_{i_l}, x_{i_k}) \geq \beta, \qquad \text{for all } l > k. \tag{2.45}$$

Moreover, from the definition of the canonical matrix $A_{n,\alpha,\beta}$ in (2.18) we have that for $l > k$

$$A_{n,\alpha,\beta}(\phi(x_{i_l}), \phi(x_{i_k})) = A_{n,\alpha,\beta}(l, k) = \beta. \tag{2.46}$$

By comparing (2.46) with (2.45) we conclude that (2.42) is true for all points with $\phi(x) > \phi(x')$. When $\phi(x) < \phi(x')$, we have $A_{n,\alpha,\beta}(\phi(x), \phi(x')) = \alpha$ which was assumed to be bounded above by the separation of the network (X, A_X), thus, $A_{n,\alpha,\beta}(\phi(x), \phi(x'))$ is not greater than any positive dissimilarity in the range of A_X. ■

Proof of Theorem 3: Consider a given arbitrary network $N = (X, A_X)$ with $X = \{x_1, x_2, ..., x_n\}$ and denote by $(X, u_X) = \mathcal{H}(X, A_X)$ the output of applying the clustering method $\mathcal{H}$ to the network N. The method $\mathcal{H}$ is known to satisfy (A1') and (A2) and we want to show that it satisfies (P1) for which we need to show that $u_X(x,x') \geq \text{mlc}(X, A_X)$ for all x, x' [cf. (2.16)].

Consider the canonical network $\vec{\Delta}_n(\alpha, \beta) = (\{1, \ldots, n\}, A_{n,\alpha,\beta})$ in (2.18) with $\beta = \text{mlc}(X, A_X)$ being the minimum loop cost of the network N [cf. (2.9)] and $\alpha > 0$ a constant not exceeding the separation of the network as defined in (2.10). Thus, we must have $\alpha \leq \text{sep}(X, A_X) \leq \text{mlc}(X, A_X) = \beta$. Notice that the networks N and $\vec{\Delta}_n(\alpha, \beta)$ have the same number of nodes.

Denote by $(\{1, \ldots, n\}, u_{\alpha,\beta}) = \mathcal{H}(\vec{\Delta}_n(\alpha, \beta))$ the ultrametric space obtained when we apply the clustering method $\mathcal{H}$ to the network $\vec{\Delta}_n(\alpha, \beta)$. Since $\mathcal{H}$ satisfies the Extended Axiom of Value (A1'), then for all indices $i, j \in \{1, \ldots, n\}$ with $i \neq j$ we have that

$$u_{\alpha,\beta}(i,j) = \max(\alpha, \beta) = \beta = \text{mlc}(X, A_X). \tag{2.47}$$

Further, focus on the bijective dissimilarity reducing map considered in Lemma 2 and notice that since the method $\mathcal{H}$ satisfies the Axiom of Transformation (A2) it follows that for all $x, x' \in X$

$$u_X(x,x') \geq u_{\alpha,\beta}(\phi(x), \phi(x')). \tag{2.48}$$

Since the equality in (2.47) is true for all $i \neq j$ and since all points $x \neq x'$ are mapped to points $\phi(x) \neq \phi(x')$ because ϕ is bijective, (2.48) implies

$$u_X(x,x') \geq \beta = \text{mlc}(X, A_X), \tag{2.49}$$

for all distinct $x, x' \in X$. This is the definition of the Property of Influence (P1). ■

The fact that (P1) is implied by (A1') and (A2) as claimed by Theorem 3 implies that adding (P1) as a third axiom on top of these two is moot. Since we have already established in Theorem 2 that (A1) and (A2) yield the same space of admissible methods as (A1') and (A2) we can conclude as a corollary of theorems 2 and 3 that (P1) is also satisfied by all methods $\mathcal{H}$ that satisfy (A1) and (A2).

Corollary 1 *If a given clustering method $\mathcal{H}$ satisfies the Axioms of Value (A1) and Transformation (A2), then it also satisfies the Property of Influence (P1).*

Proof: If a clustering method $\mathcal{H}$ satisfies (A1) and (A2) it follows from Theorem 2 that it must satisfy (A1') and (A2). But if the latter is true it follows from Theorem 3 that it must satisfy property (P1). ■

In the discussion leading to the definition of the Axiom of Value (A1) in Section 2.2 we argued that the intuitive notion of a cluster dictates that it must be possible for co-clustered nodes to influence each other. In the discussion leading to the definition of the Property of Influence (P1) at the beginning of this section we argued that in networks with more than two nodes the natural extension is that co-clustered nodes must be able to influence each other either directly or through their indirect influence on other intermediate nodes. The Property of Influence is a codification of this intuition because it states the impossibility of cluster formation at resolutions where influence loops cannot be formed. While (P1) and (A1) seem quite different and seemingly independent, we have shown in this section that if a method satisfies axioms (A1) and (A2) it must satisfy (P1). Therefore, requiring direct influence on a two-node network as in (A1) restricts the mechanisms for indirect influence propagation so that clusters cannot be formed at resolutions that do not allow for mutual, possibly indirect, influence as stated in (P1). In that sense the restriction of indirect influence propagation in (P1) is not just *intuitively* reasonable but *formally* implied by the more straightforward restrictions on direct influence in (A1) and dissimilarity reducing maps in (A2).

Chapter 3

Admissible clustering methods and algorithms

Having stated the Axioms of Value and Transformation as required properties for admissibility, we now question the existence of admissible methods. Indeed, in Section 3.1 we derive two hierarchical clustering methods that abide by axioms (A1) and (A2), namely reciprocal and nonreciprocal clustering. The former requires clusters to form through edges exhibiting low dissimilarities in both directions whereas the latter allows clusters to form through cycles of small dissimilarities. A fundamental result regarding admissibility is that any clustering method that satisfies axioms (A1) and (A2) lies between reciprocal and nonreciprocal clustering in a well-defined sense (Section 3.2). After establishing the extreme nature of reciprocal and nonreciprocal clustering, in Section 3.3 we seek for admissible methods contained between them.

In some applications the requirement for bidirectional influence in the Axiom of Value is not justified as unidirectional influence might suffice to establish proximity. This alternative value statement leads to the study of alternative axiomatic constructions in Section 3.4. Finally, in Section 3.5 we develop algorithms to compute the ultrametrics associated with the methods introduced throughout the chapter.

3.1 Reciprocal and nonreciprocal clustering

A clustering method satisfying axioms (A1)-(A2) can be constructed by considering the symmetric dissimilarity

$$\bar{A}_X(x, x') := \max(A_X(x, x'), A_X(x', x)), \tag{3.1}$$

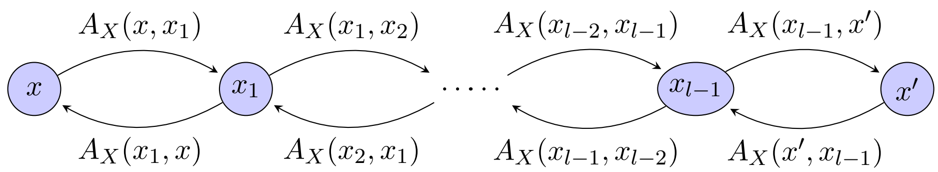


Figure 3.1: Reciprocal clustering. Nodes x and x' are clustered together at resolution δ if they can be joined with a (reciprocal) path whose maximum dissimilarity is smaller than or equal to δ in both directions [cf. (3.2)]. Of all methods that satisfy the Axioms of Value and Transformation, reciprocal clustering yields the largest ultrametric between any pair of nodes.

for all $x, x' \in X$. This effectively reduces the problem to clustering of symmetric data, a scenario in which the single linkage method in (2.8) is known to satisfy axioms similar to (A1)-(A2), [10]. Drawing upon this connection we define the *reciprocal* clustering method $\mathcal{H}^{\mathrm{R}}$ with output $(X, u_X^{\mathrm{R}}) = \mathcal{H}^{\mathrm{R}}(X, A_X)$ as the one for which the ultrametric $u_X^{\mathrm{R}}(x, x')$ between points x and x' is given by

$$u_X^{\mathrm{R}}(x, x') := \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \bar{A}_X(x_i, x_{i+1}). \tag{3.2}$$

An illustration of the definition in (3.2) is shown in Fig. 3.1. We search for paths $P_{xx'}$ linking nodes x and x'. For a given path we walk from x to x' and for every link, connecting say x_i with x_{i+1}, we determine the maximum dissimilarity in both directions, i.e. the value of $\bar{A}_X(x_i, x_{i+1})$. We then determine the maximum across all the links in the path. The reciprocal ultrametric $u_X^{\mathrm{R}}(x, x')$ between points x and x' is the minimum of this value across all possible paths. Recalling the equivalence of dendrograms and ultrametrics provided by Theorem 1 we know that R_X, the dendrogram produced by reciprocal clustering, clusters x and x' together for resolutions $\delta \geq u_X^{\mathrm{R}}(x, x')$. Combining the latter observation with (3.2), we can write the reciprocal clustering equivalence classes as

$$x \sim_{\mathrm{R}_X(\delta)} x' \iff \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \bar{A}_X(x_i, x_{i+1}) \leq \delta. \tag{3.3}$$

Comparing (3.3) with the definition of single linkage in (2.8), where $\tilde{u}_X^*(x, x')$ is defined in (2.7), we see that reciprocal clustering is equivalent to single linkage for the network $N = (X, \bar{A}_X)$ where dissimilarities between nodes are symmetrized to the maximum value of each directed dissimilarity.

For the method $\mathcal{H}^{\mathrm{R}}$ specified in (3.2) to be a properly defined hierarchical clustering method, we need to establish that u_X^{R} is a valid ultrametric. We know that this is the case because u_X^{R} is equivalent to the definition of single linkage for the symmetric dissimilarity $\bar{A}_X$. Otherwise, we can also verify that u_X^{R} as defined by (3.2) is indeed an ultrametric in the set X. It is clear that $u_X^{\mathrm{R}}(x, x') = 0$ only if $x = x'$ and that $u_X^{\mathrm{R}}(x, x') = u_X^{\mathrm{R}}(x', x)$ because the definition is symmetric on x and x'. To verify that the strong triangle inequality in (2.12) holds, let $P^*_{xx'}$ and $P^*_{x'x''}$ be paths that achieve the minimum in (3.2) for $u^{\mathrm{R}}_X(x,x')$ and $u^{\mathrm{R}}_X(x',x'')$, respectively. The maximum cost in the concatenated path $P_{xx''} = P^*_{xx'} \uplus P^*_{x'x''}$ does not exceed the maximum cost in each individual path. Thus, while the maximum cost may be smaller on a different path, the path $P_{xx''}$ suffices to bound $u^{\mathrm{R}}_X(x,x'') \leq \max\big(u^{\mathrm{R}}_X(x,x'), u^{\mathrm{R}}_X(x',x'')\big)$ as in (2.12). It is also possible to prove that $\mathcal{H}^{\mathrm{R}}$ satisfies axioms (A1)-(A2) as the following proposition shows.

Proposition 1 *The reciprocal clustering method $\mathcal{H}^{\mathrm{R}}$ is valid and admissible. I.e., u^{R}_X defined by* (3.2) *is an ultrametric for all networks $N = (X, A_X)$ and $\mathcal{H}^{\mathrm{R}}$ satisfies axioms (A1)-(A2).*

Proof: That u^{R}_X conforms to the definition of an ultrametric is already proved in the paragraph preceding this proposition. To see that the Axiom of Value (A1) is satisfied pick an arbitrary two-node network $\vec{\Delta}_2(\alpha,\beta)$ as defined in Section 2.1 and denote by $(\{p,q\}, u^{\mathrm{R}}_{p,q}) = \mathcal{H}^{\mathrm{R}}(\vec{\Delta}_2(\alpha,\beta))$ the output of applying the reciprocal clustering method to $\vec{\Delta}_2(\alpha,\beta)$. Since there is only one possible path between p and q consisting of a single link, applying the definition in (3.2) yields

$$u^{\mathrm{R}}_{p,q}(p,q) = \max\Big(A_{p,q}(p,q), A_{p,q}(q,p)\Big) = \max(\alpha,\beta). \tag{3.4}$$

Axiom (A1) is thereby satisfied.

To show fulfillment of axiom (A2), consider two networks (X, A_X) and (Y, A_Y) and a dissimilarity reducing map $\phi : X \to Y$. Let $(X, u^{\mathrm{R}}_X) = \mathcal{H}^{\mathrm{R}}(X, A_X)$ and $(Y, u^{\mathrm{R}}_Y) = \mathcal{H}^{\mathrm{R}}(Y, A_Y)$ be the outputs of applying the reciprocal clustering method to networks (X, A_X) and (Y, A_Y). For an arbitrary pair of nodes $x, x' \in X$, denote by $P^{X*}_{xx'} = [x = x_0, \dots, x_l = x']$ a path that achieves the minimum reciprocal cost in (3.2) so as to write

$$u^{\mathrm{R}}_X(x,x') = \max_{i | x_i \in P^{X*}_{xx'}} \bar{A}_X(x_i, x_{i+1}). \tag{3.5}$$

Consider the transformed path $P^Y_{\phi(x)\phi(x')} = [\phi(x) = \phi(x_0), \dots, \phi(x_l) = \phi(x')]$ in the space Y. Since the transformation ϕ does not increase dissimilarities we have that for all links in this path $A_Y(\phi(x_i), \phi(x_{i+1})) \leq A_X(x_i, x_{i+1})$ and $A_Y(\phi(x_{i+1}), \phi(x_i)) \leq A_X(x_{i+1}, x_i)$. Combining this observation with (3.5) we obtain,

$$\max_{\phi(x_i) \in P^Y_{\phi(x)\phi(x')}} \bar{A}_Y(\phi(x_i), \phi(x_{i+1})) \leq u^{\mathrm{R}}_X(x,x'). \tag{3.6}$$

Further note that $P^Y_{\phi(x)\phi(x')}$ is a particular path joining $\phi(x)$ and $\phi(x')$ whereas the reciprocal

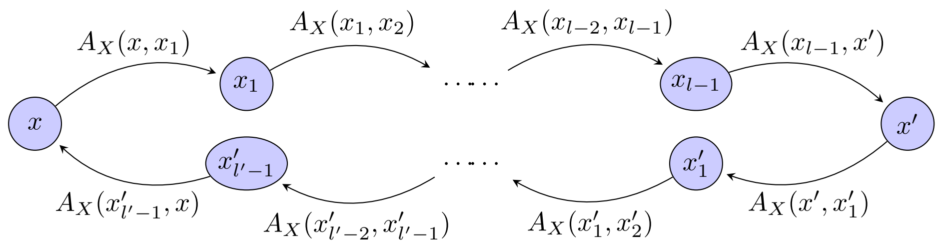


Figure 3.2: Nonreciprocal clustering. Nodes x and x' are co-clustered at resolution δ if they can be joined in both directions with possibly different (nonreciprocal) paths of maximum dissimilarity not greater than δ [cf. (3.8)]. Of all methods abiding by the Axioms of Value and Transformation, nonreciprocal clustering yields the smallest ultrametric between any pair of nodes.

ultrametric is the minimum across paths. Therefore,

$$u_Y^{\mathrm{R}}(\phi(x), \phi(x')) \leq \max_{\phi(x_i) \in P^Y_{\phi(x)\phi(x')}} \bar{A}_Y(\phi(x_i), \phi(x_{i+1})). \tag{3.7}$$

Substituting (3.6) in (3.7), it follows that $u_Y^{\mathrm{R}}(\phi(x), \phi(x')) \leq u_X^{\mathrm{R}}(x, x')$, as wanted. ■

In reciprocal clustering, nodes x and x' belong to the same cluster at a resolution δ whenever we can go back and forth from x to x' at a maximum cost δ through the same path. In nonreciprocal clustering we relax the restriction about the path being the same in both directions and cluster nodes x and x' together if there are paths, possibly different, linking x to x' and x' to x. To state this definition in terms of ultrametrics consider a given network $N = (X, A_X)$ and recall the definition of the unidirectional minimum path cost $\tilde{u}^*_X$ in (2.7). We define the *nonreciprocal* clustering method $\mathcal{H}^{\mathrm{NR}}$ with output $(X, u_X^{\mathrm{NR}}) = \mathcal{H}^{\mathrm{NR}}(X, A_X)$ as the one for which the ultrametric $u_X^{\mathrm{NR}}(x, x')$ between points x and x' is given by the maximum of the unidirectional minimum path costs $\tilde{u}^*_X(x, x')$ and $\tilde{u}^*_X(x', x)$ in each direction,

$$u_X^{\mathrm{NR}}(x, x') = \max\Big(\tilde{u}^*_X(x, x'),\ \tilde{u}^*_X(x', x)\Big). \tag{3.8}$$

An illustration of the definition in (3.8) is shown in Fig. 3.2. We consider forward paths $P_{xx'}$ going from x to x' and backward paths $P_{x'x}$ going from x' to x. For each of these paths we determine the maximum dissimilarity across all the links in the path. We then search independently for the best forward path $P_{xx'}$ and the best backward path $P_{x'x}$ that minimize the respective maximum dissimilarities across all possible paths. The nonreciprocal ultrametric $u_X^{\mathrm{NR}}(x, x')$ between points x and x' is the maximum of these two minimum values.

As is the case with reciprocal clustering we can verify that u_X^{NR} is a properly defined ultrametric and that, as a consequence, the nonreciprocal clustering method $\mathcal{H}^{\mathrm{NR}}$

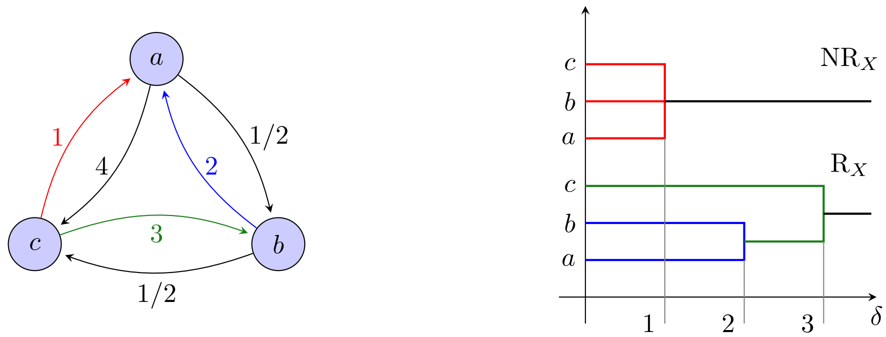


Figure 3.3: Reciprocal and nonreciprocal dendrograms. An example network with its corresponding reciprocal (bottom) and nonreciprocal (top) dendrograms is shown. The optimal reciprocal path linking a and b is $[a, b]$ the optimal path linking b and c is $[b, c]$ and the optimal path linking a and c is $[a, b, c]$. The optimal nonreciprocal paths linking a and b are $[a, b]$ and $[b, c, a]$. Of these two the cost of $[b, c, a]$ is larger.

is properly defined. Identity and symmetry are immediate. For the strong triangle inequality consider paths $P^*_{xx'}$ and $P^*_{x'x''}$ that achieve the minimum costs in $\tilde{u}^*_X(x, x')$ and $\tilde{u}^*_X(x', x'')$ as well as the paths $P^*_{x''x'}$ and $P^*_{x'x}$ that achieve the minimum costs in $\tilde{u}^*_X(x'', x')$ and $\tilde{u}^*_X(x', x)$. The concatenation of these paths permits concluding that $u^{\mathrm{NR}}_X(x, x'') \leq \max\left(u^{\mathrm{NR}}_X(x, x'), u^{\mathrm{NR}}_X(x', x'')\right)$, which is the strong triangle inequality in (2.12). The method $\mathcal{H}^{\mathrm{NR}}$ also satisfies axioms (A1)-(A2) as the following proposition shows.

Proposition 2 *The nonreciprocal clustering method $\mathcal{H}^{\mathrm{NR}}$ is valid and admissible. I.e., u^{NR}_X defined by (3.8) is an ultrametric for all networks $N = (X, A_X)$ and $\mathcal{H}^{\mathrm{NR}}$ satisfies axioms (A1)-(A2).*

Proof: That $\mathcal{H}^{\mathrm{NR}}$ outputs valid ultrametrics was already argued prior to the statement of Proposition 2. The proof of admissibility is analogous to the proof of Proposition 1 and presented for completeness. For axiom (A1) notice that for the two-node network $\vec{\Delta}_2(\alpha, \beta)$ we have $\tilde{u}^*_X(p, q) = \alpha$ and $\tilde{u}^*_X(q, p) = \beta$ because there is only one possible path selection. According to (3.8) we then have

$$u^{\mathrm{NR}}_X(p, q) = \max\left(\tilde{u}^*_X(p, q), \tilde{u}^*_X(q, p)\right) = \max(\alpha, \beta). \tag{3.9}$$

To prove that axiom (A2) is satisfied consider arbitrary points $x, x' \in X$ and denote by $P^*_{xx'}$ one path achieving the minimum path cost in (2.7),

$$\tilde{u}^*_X(x, x') = \max_{i | x_i \in P^*_{xx'}} A(x_i, x_{i+1}). \tag{3.10}$$

Consider the transformed path $P^Y_{\phi(x)\phi(x')} = [\phi(x) = \phi(x_0), \ldots, \phi(x_l) = \phi(x')]$ in the space Y. Since the map $\phi : X \to Y$ reduces dissimilarities we have that for all links in this path $A_Y(\phi(x_i), \phi(x_{i+1})) \leq A_X(x_i, x_{i+1})$. Consequently,

$$\max_{i | x_i \in P^Y_{\phi(x)\phi(x')}} A_Y(\phi(x_i), \phi(x_{i+1})) \leq \max_{i | x_i \in P^*_{xx'}} A_X(x_i, x_{i+1}). \tag{3.11}$$

Further note that the minimum path cost $\tilde{u}^*_Y(\phi(x), \phi(x'))$ among all paths linking $\phi(x)$ to $\phi(x')$ cannot exceed the cost in the given path $P^Y_{\phi(x)\phi(x')}$. Combining this observation with the inequality in (3.11) it follows that

$$\tilde{u}^*_Y(\phi(x), \phi(x')) \leq \max_{i | x_i \in P^*_{xx'}} A_X(x_i, x_{i+1}) = \tilde{u}^*_X(x, x'), \tag{3.12}$$

where we also used (3.10) to write the equality.

The bound in (3.12) is true for arbitrary ordered pair x, x'. In particular, it is true if we reverse the order to consider the pair x', x. Consequently, we can write

$$\max\Big(\tilde{u}^*_Y(\phi(x),\phi(x')),\ \tilde{u}^*_Y(\phi(x'), \phi(x))\Big) \leq \max\Big(\tilde{u}^*_X(x, x'),\ \tilde{u}^*_X(x', x)\Big), \tag{3.13}$$

because both maximands in the left are smaller than their corresponding maximand in the right. To complete the proof just notice that the expressions in (3.13) correspond to the nonreciprocal ultrametric distances $u^{\text{NR}}_Y(\phi(x), \phi(x'))$ and $u^{\text{NR}}_X(x, x')$ [cf. (3.8)]. ■

We denote by NR_X the dendrogram output by the nonreciprocal method $\mathcal{H}^{\text{NR}}$, equivalent to u^{NR}_X by Theorem 1. The reciprocal and nonreciprocal dendrograms for an example network are shown in Fig. 3.3. In the reciprocal dendrogram nodes a and b cluster together at resolution $\delta = 2$ due to their direct connections $A_X(a, b) = 1/2$ and $A_X(b, a) = 2$. Node c joins this cluster at resolution $\delta = 3$ because it links bidirectionally with b through the direct path $[b, c]$ whose maximum cost is $A_X(c, b) = 3$. The optimal reciprocal path linking a and c is $[a, b, c]$ whose maximum cost is also $A_X(c, b) = 3$. In the nonreciprocal dendrogram we can link nodes with different paths in each direction. As a consequence, a and b cluster together at resolution $\delta = 1$ because the directed cost of the path $[a, b]$ is $A_X(a, b) = 1/2$ and the directed cost of the path $[b, c, a]$ is $A_X(c, a) = 1$. Similar paths demonstrate that a and c as well as b and c also cluster together at resolution $\delta = 1$.

3.2 Extreme ultrametrics

Given that we have constructed two admissible methods satisfying axioms (A1)-(A2), the question whether these two constructions are the only possible ones arises and, if not, whether they are special in some sense. We will see in Section 3.3 that there are constructions other than reciprocal and nonreciprocal clustering that satisfy axioms (A1)-(A2). However, we prove in this section that reciprocal and nonreciprocal clustering are a peculiar pair in that all possible admissible clustering methods are contained between them in a well-defined sense. To explain this sense properly, observe that since reciprocal paths [cf. Fig. 3.1] are particular cases of nonreciprocal paths [cf. Fig. 3.2] we must have that for all pairs of nodes x, x'

$$u_X^{\mathrm{NR}}(x, x') \leq u_X^{\mathrm{R}}(x, x'). \tag{3.14}$$

I.e., nonreciprocal ultrametrics do not exceed reciprocal ultrametrics. An important characterization is that any method $\mathcal{H}$ satisfying axioms (A1)-(A2) yields ultrametrics that lie between u_X^{NR} and u_X^{R} as we formally state in the following theorem.

Theorem 4 *Consider an admissible clustering method $\mathcal{H}$ satisfying axioms (A1)-(A2). For an arbitrary given network $N = (X, A_X)$ denote by $(X, u_X) = \mathcal{H}(X, A_X)$ the output of $\mathcal{H}$ applied to N. Then, for all pairs of nodes x, x'*

$$u_X^{\mathrm{NR}}(x, x') \leq u_X(x, x') \leq u_X^{\mathrm{R}}(x, x'), \tag{3.15}$$

where $u_X^{\mathrm{NR}}(x, x')$ and $u_X^{\mathrm{R}}(x, x')$ denote the nonreciprocal and reciprocal ultrametrics as defined by (3.8) *and* (3.2)*, respectively.*

Proof of $\mathbf{u_X^{\mathrm{NR}}(x, x') \leq u_X(x, x')}$: Recall that validity of (A1)-(A2) implies validity of (P1) by Corollary 1. Consider the nonreciprocal clustering equivalence relation $\sim_{\mathrm{NR}_X(\delta)}$ at resolution δ according to which $x \sim_{\mathrm{NR}_X(\delta)} x'$ if and only if x and x' belong to the same nonreciprocal cluster at resolution δ. Notice that this is true if and only if $u_X^{\mathrm{NR}}(x, x') \leq \delta$. Further consider the space $Z := X \mod \sim_{\mathrm{NR}_X(\delta)}$ of corresponding equivalence classes and the map $\phi_\delta : X \to Z$ that maps each point of X to its equivalence class. Notice that x and x' are mapped to the same point z if they belong to the same cluster at resolution δ, which allows us to write

$$\phi_\delta(x) = \phi_\delta(x') \iff u_X^{\mathrm{NR}}(x, x') \leq \delta. \tag{3.16}$$

We define the network $N_Z = (Z, A_Z)$ by endowing Z with the dissimilarity A_Z derived from the dissimilarity A_X as

$$A_Z(z, z') = \min_{x \in \phi_\delta^{-1}(z), x' \in \phi_\delta^{-1}(z')} A_X(x, x'). \tag{3.17}$$

The dissimilarity $A_Z(z, z')$ compares all the dissimilarities $A_X(x, x')$ between a member of the equivalence class z and a member of the equivalence class z' and sets $A_Z(z, z')$ to the value corresponding to the least dissimilar pair; see Fig. 3.4. Notice that according to construction, the map ϕ_δ is dissimilarity reducing

$$A_X(x, x') \geq A_Z(\phi_\delta(x), \phi_\delta(x')), \tag{3.18}$$

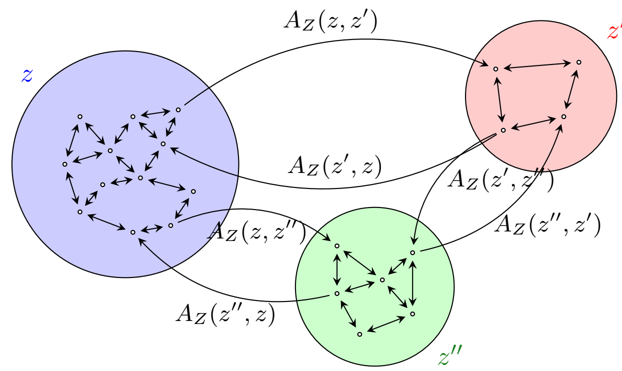


Figure 3.4: Network of equivalence classes for a given resolution. Each shaded subset of nodes represents an equivalence class. The Axiom of Transformation permits relating the clustering of nodes in the original network and the clustering of nodes in the network of equivalence classes.

because we either have $A_Z(\phi_\delta(x), \phi_\delta(x')) = 0$ if x and x' are co-clustered at resolution δ, or $A_X(x, x') \geq \min_{x \in \phi_\delta^{-1}(z), x' \in \phi_\delta^{-1}(z')} A_X(x, x') = A_Z(\phi_\delta(x), \phi_\delta(x'))$ if they are mapped to different equivalent classes.

Consider now an arbitrary method $\mathcal{H}$ satisfying axioms (A1)-(A2) and denote by $(Z, u_Z) = \mathcal{H}(Z, A_Z)$ the outcome of $\mathcal{H}$ when applied to the network N_Z. To apply property (P1) to this outcome we determine the minimum loop cost of (Z, A_Z) in the following claim.

Claim 1 *The minimum loop cost of the network N_Z is $\text{mlc}(Z, A_Z) > \delta$.*

Proof: We first establish that, according to (3.17), if $z \neq z'$ it must be that either $A_Z(z, z') > \delta$ or $A_Z(z', z) > \delta$ for otherwise z and z' would be the same equivalent class. Indeed, if both $A_Z(z, z') \leq \delta$ and $A_Z(z', z) \leq \delta$ we can build paths $P_{xx'}$ and $P_{x'x}$ with maximum cost smaller than δ for any $x \in \phi_\delta^{-1}(z)$ and $x' \in \phi_\delta^{-1}(z')$. For the path $P_{xx'}$ denote by $x_o \in \phi_\delta^{-1}(z)$ and $x'_i \in \phi_\delta^{-1}(z')$ the points achieving the minimum in (3.17) so that $A_Z(z, z') = A_X(x_o, x'_i)$. Since x and x_o are in the same equivalence class there is a path P_{xx_o} of maximum cost smaller than δ. Likewise, since x'_i and x' are in the same class there is a path $P_{x'_i x'}$ that joins them at maximum cost smaller than δ. Therefore, the concatenated path

$$P_{xx'} = P_{xx_o} \uplus [x_o, x'_i] \uplus P_{x'_i x'}, \tag{3.19}$$

has maximum cost smaller than δ. The construction of the path $P_{x'x}$ is analogous. However, the existence of these two paths implies that x and x' are clustered together at resolution δ [cf (3.8)] contradicting the assumption that z and z' are different equivalent classes.

To prove that the minimum loop cost of N_Z is $\text{mlc}(Z, A_Z) > \delta$ assume that $\text{mlc}(Z, A_Z) \leq \delta$ and denote by $[z, z', \ldots, z^{(l)}, z]$ a loop of cost smaller than δ. For any $x \in \phi_\delta^{-1}(z)$ and $x' \in \phi_\delta^{-1}(z')$ we can join x to x' using the path $P_{xx'}$ in (3.19). To join x' and x denote by

$x_i^{(k)}$ and $x_o^{(k)}$ the points for which $A_Z(z^{(k)}, z^{(k+1)}) = A_X(x_o^{(k)}, x_i^{(k+1)})$ as in (3.17). We can then join x'_o and $x_o^{(l)}$ with the concatenated path

$$P_{x'_o x_o^{(l)}} = \biguplus_{k=1}^{l-1} \left[x_o^{(k)}, x_i^{(k+1)}\right] \uplus P_{x_i^{(k+1)} x_o^{(k+1)}}. \tag{3.20}$$

The maximum cost in traversing this path is smaller than δ because the maximum cost in $P_{x_i^{(k+1)} x_o^{(k+1)}}$ is smaller than δ and because $A_X(x_o^{(k)}, x_i^{(k+1)}) \leq \delta$ by assumption. We can now join x' to x with the concatenated path

$$P_{x'x} = P_{x'x'_o} \uplus P_{x'_o x_o^{(l)}} \uplus [x_o^{(l)}, x_i] \uplus P_{x_i x}, \tag{3.21}$$

whose maximum cost is smaller than δ. Using the paths (3.19) and (3.21) it follows that $u_X^{\mathrm{NR}}(x, x') \leq \delta$ contradicting the assumption that x and x' belong to different equivalent classes. Therefore, the assumption that $\mathrm{mlc}(Z, A_Z) \leq \delta$ cannot hold. The opposite must be true. ■

Continuing with the main proof recall that $(Z, u_Z) = \mathcal{H}(Z, A_Z)$ is the outcome of the arbitrary clustering method $\mathcal{H}$ applied to the network N_Z. Since the minimum loop cost of Z satisfies $\mathrm{mlc}(Z, A_Z) > \delta$, it follows from property (P1) that for all pairs z, z',

$$u_Z(z, z') > \delta. \tag{3.22}$$

Further note that according to (3.18) and axiom (A2) we must have $u_X(x, x') \geq u_Z(z, z')$. This fact, combined with (3.22) allows us to conclude that when x and x' map to different equivalence classes

$$u_X(x, x') \geq u_Z(z, z') > \delta. \tag{3.23}$$

Notice now that according to (3.16), x and x' mapping to different equivalence classes is equivalent to $u_X^{\mathrm{NR}}(x, x') > \delta$. Consequently, we can claim that $u_X^{\mathrm{NR}}(x, x') > \delta$ implies $u_X(x, x') > \delta$, or, in set notation that

$$\{(x, x') : u_X^{\mathrm{NR}}(x, x') > \delta\} \subseteq \{(x, x') : u_X(x, x') > \delta\}. \tag{3.24}$$

Because (3.24) is true for arbitrary $\delta > 0$ it implies that $u_X^{\mathrm{NR}}(x, x') \leq u_X(x, x')$ for all $x, x' \in X$ as in the first inequality in (3.15). ■

Proof of $\mathbf{u_X(x, x') \leq u_X^{R}(x, x')}$: Consider points x and x' with reciprocal ultrametric $u_X^{\mathrm{R}}(x, x') = \delta$. Let $P_{xx'} = [x = x_0, \dots, x_l = x']$ be a path achieving the minimum in (3.2) so that we can write

$$\delta = u_X^{\mathrm{R}}(x,x') = \max_i \max\Big(A_X(x_i,x_{i+1}), A_X(x_{i+1},x_i)\Big). \tag{3.25}$$

Turn attention to the symmetric two-point network $N_{p,q} = (\{p,q\}, A_{p,q})$ with $A_{p,q}(p,q) = A_{p,q}(q,p) = \delta$. Denote the output of clustering method $\mathcal{H}$ applied to network $N_{p,q}$ as $(\{p,q\}, u_{p,q}) = \mathcal{H}(\{p,q\}, A_{p,q})$. Notice that according to axiom (A1) we have $u_{p,q}(p,q) = \max(\delta,\delta) = \delta$.

Focus now on transformations $\phi_i : \{p,q\} \to X$ given by $\phi_i(p) = x_i$, $\phi_i(q) = x_{i+1}$ so as to map p and q to subsequent points in the path $P_{xx'}$ used in (3.25). Since it follows from (3.25) that $A_X(x_i,x_{i+1}) \leq \delta$ and $A_X(x_{i+1},x_i) \leq \delta$ for all i, it is just a simple matter of notation to observe that

$$A_X(\phi_i(p),\phi_i(q)) \leq A_{p,q}(p,q) = \delta, \quad A_X(\phi_i(q),\phi_i(p)) \leq A_{p,q}(q,p) = \delta. \tag{3.26}$$

Since according to (3.26) transformations ϕ_i are dissimilarity reducing, it follows from axiom (A2) that

$$u_X(\phi_i(p),\phi_i(q)) \leq u_{p,q}(p,q) = \delta. \tag{3.27}$$

Substituting the equivalences $\phi_i(p) = x_i$, $\phi_i(q) = x_{i+1}$ and recalling that (3.27) is true for all i we can equivalently write

$$u_X(x_i,x_{i+1}) \leq \delta, \quad \text{for all } i. \tag{3.28}$$

To complete the proof we use the fact that since u_X is an ultrametric and $P_{xx'} = [x = x_0, \ldots, x_l = x']$ is a path joining x and x' the strong triangle inequality dictates that

$$u_X(x,x') \leq \max_i u_X(x_i,x_{i+1}) \leq \delta, \tag{3.29}$$

where we used (3.28) in the second inequality. The proof of the second inequality in (3.15) follows by substituting $\delta = u_X^{\mathrm{R}}(x,x')$ [cf. (3.25)] into (3.29). ■

According to Theorem 4, nonreciprocal clustering applied to a given network $N = (X, A_X)$ yields a uniformly minimal ultrametric among those output by all clustering methods satisfying axioms (A1)-(A2). Reciprocal clustering yields a uniformly maximal ultrametric. Any other clustering method abiding by (A1)-(A2) yields an ultrametric such that the value $u_X(x,x')$ for any two points $x, x' \in X$ lies between the values $u_X^{\mathrm{NR}}(x,x')$ and $u_X^{\mathrm{R}}(x,x')$ assigned by nonreciprocal and reciprocal clustering. In terms of dendrograms, (3.15) implies that among all possible clustering methods, the smallest possible resolution at which nodes are clustered together is the one corresponding to nonreciprocal clustering. The highest possible resolution is the one that corresponds to reciprocal clustering.

3.2.1 Hierarchical clustering of symmetric networks

Restrict attention to the subspace $\mathcal{N} \subset \tilde{\mathcal{N}}$ of symmetric networks $N = (X, A_X)$ with $A_X(x, x') = A_X(x', x)$ for all $x, x' \in X$. When restricted to the space $\mathcal{N}$ reciprocal and nonreciprocal clustering are equivalent methods because, for any pair of points, minimizing nonreciprocal paths are always reciprocal – more precisely there may be multiple minimizing nonreciprocal paths but at least one of them is reciprocal. To see this formally observe that in symmetric networks the symmetrization in (3.1) is unnecessary because $\bar{A}_X(x_i, x_{i+1}) = A_X(x_i, x_{i+1}) = A_X(x_{i+1}, x_i)$ and the definition of reciprocal clustering in (3.2) reduces to

$$u_X^{\mathrm{R}}(x, x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} A_X(x_i, x_{i+1}) = \min_{P_{x'x}} \max_{i | x_i \in P_{x'x}} A_X(x_i, x_{i+1}). \tag{3.30}$$

Further note that the costs of any given path $P_{xx'} = [x = x_0, x_1, \ldots, x_{l-1}, x_l = x']$ and its reciprocal $P_{x'x} = [x' = x_l, x_{l-1}, \ldots, x_1, x_0 = x]$ are the same. It follows that directed minimum path costs $\tilde{u}_X^*(x, x') = \tilde{u}_X^*(x', x)$ are equal and according to (3.8) equal to the nonreciprocal ultrametric

$$u_X^{\mathrm{NR}}(x, x') = \tilde{u}_X^*(x, x') = \tilde{u}_X^*(x', x) = u_X^{\mathrm{R}}(x, x'). \tag{3.31}$$

To write the last equality in (3.31) we used the definitions of $\tilde{u}_X^*(x, x')$ and $\tilde{u}_X^*(x', x)$ in (2.7) which are correspondingly equivalent to the first and second equality in (3.30).

By further comparison of the ultrametric definition of single linkage in (2.15) with (3.31) the equivalence of reciprocal, nonreciprocal, and single linkage clustering in symmetric networks follows

$$u_X^{\mathrm{NR}}(x, x') = u_X^{\mathrm{SL}}(x, x') = u_X^{\mathrm{R}}(x, x'). \tag{3.32}$$

The equivalence in (3.31) along with Theorem 4 demonstrates that when considering the application of hierarchical clustering methods to symmetric networks $\mathcal{H} : \mathcal{N} \to \mathcal{U}$, there exist a unique method satisfying (A1)-(A2). The equivalence in (3.32) shows that this method is single linkage. Before stating this result formally let us define the symmetric version of the Axiom of Value:

> *(B1) Symmetric Axiom of Value.* The ultrametric $(\{p, q\}, u_{p,q}) = \mathcal{H}(\vec{\Delta}_2(\alpha, \alpha))$ produced by $\mathcal{H}$ applied to the symmetric two-node network $\vec{\Delta}_2(\alpha, \alpha)$ satisfies $u_{p,q}(p, q) = \alpha$.

Since there is only one dissimilarity in a symmetric network with two nodes, (B1) states that they cluster together at the resolution that connects them to each other. We can now invoke Theorem 4 and (3.32) to prove that single linkage is the unique hierarchical clustering method in symmetric networks that is admissible with respect to (B1) and (A2).

Corollary 2 *Let $\mathcal{H} : \mathcal{N} \to \mathcal{U}$ be a hierarchical clustering method for symmetric networks $N = (X, A_X) \in \mathcal{N}$, that is $A_X(x, x') = A_X(x', x)$ for all $x, x' \in X$, and $\mathcal{H}^{\mathrm{SL}}$ be the single linkage method with output ultrametrics as defined in* (2.15)*. If $\mathcal{H}$ satisfies axioms (B1) and (A2) then $\mathcal{H} \equiv \mathcal{H}^{\mathrm{SL}}$.*

Proof: When restricted to symmetric networks (B1) and (A1) are equivalent statements. Thus, $\mathcal{H}$ satisfies the hypotheses of Theorem 4 and as a consequence (3.15) is true for any pair of points x, x' of any network $N \in \mathcal{N}$. But by (3.32) nonreciprocal, single linkage, and reciprocal ultrametrics coincide. Thus, we can reduce (3.15) to

$$u_X^{\mathrm{SL}}(x, x') \le u_X(x, x') \le u_X^{\mathrm{SL}}(x, x'). \tag{3.33}$$

It then must be $u_X^{\mathrm{SL}}(x, x') = u_X(x, x')$ for any pair of points x, x' of any network $N \in \mathcal{N}$. This means that $\mathcal{H} \equiv \mathcal{H}^{\mathrm{SL}}$. ■

The uniqueness result claimed by Corollary 2 strengthens the uniqueness result in [10, Theorem 18]. To explain the differences consider the symmetric version of the Property of Influence. In a symmetric network there is always a loop of minimum cost of the form $[x, x', x]$ for some pair of points x, x'. Indeed, say that $P^*_{x^*x^*}$ is one of the loops achieving the minimum cost in (2.9) and let $A_X(x, x') = \mathrm{mlc}(X, A_X)$ be the maximum dissimilarity in this loop. Then, the cost of the loop $[x, x', x]$ is $A_X(x, x') = A_X(x', x) = \mathrm{mlc}(X, A_X)$ which means that either the loop $P^*_{x^*x^*}$ was already of the form $[x, x', x]$ or that the cost of the loop $[x, x', x]$ is the same as $P^*_{x^*x^*}$. In any event, there is a loop of minimum cost of the form $[x, x', x]$ which implies that in symmetric networks we must have

$$\mathrm{mlc}(X, A_X) = \min_{x \neq x'} A_X(x, x') = \mathrm{sep}(X, A_X), \tag{3.34}$$

where we recalled the definition of the separation of a network stated in (2.10) to write the second equality. With this observation we can now introduce the symmetric version of the Property of Influence (P1):

> *(Q1) Symmetric Property of Influence.* For any symmetric network $N_X = (X, A_X)$ the output $(X, u_X) = \mathcal{H}(X, A_X)$ corresponding to the application of method $\mathcal{H}$ is such that the ultrametric $u_X(x, x')$ between any two distinct points x and x' cannot be smaller than the separation of the network [cf. (3.34)], i.e. $u_X(x, x') \ge \mathrm{sep}(X, A_X)$.

In [10] admissibility is defined with respect to (B1), (A2), and (Q1), denominated as conditions (I), (II), and (III) in Theorem 18 of the mentioned paper. Corollary 2 shows that property (Q1) is redundant when given axioms (B1) and (A2) – respectively, Condition (III) of [10, Theorem 18] is redundant when given conditions (I) and (II). Corollary 2 also shows that single linkage is the unique admissible method for all symmetric, not necessarily metric, networks.

3.3 Intermediate clustering methods

Reciprocal and nonreciprocal clustering bound the range of methods satisfying axioms (A1)-(A2) in the sense specified by Theorem 4. Since $\mathcal{H}^{\mathrm{R}}$ and $\mathcal{H}^{\mathrm{NR}}$ are in general different, a question of interest is whether one can identify methods which are *intermediate* to $\mathcal{H}^{\mathrm{R}}$ and $\mathcal{H}^{\mathrm{NR}}$. We present three types of intermediate clustering methods: grafting, convex combinations, and semi-reciprocal clustering. The latter arises as a natural intermediate method in an algorithmic sense, as further discussed in Section 3.5.

3.3.1 Grafting

A family of admissible methods can be constructed by grafting branches of the nonreciprocal dendrogram into corresponding branches of the reciprocal dendrogram; see Fig. 3.5. To be precise, consider a given positive constant $\beta > 0$. For any given network $N = (X, A_X)$ compute the reciprocal and nonreciprocal dendrograms and cut all branches of the reciprocal dendrogram at resolution β. For each of these branches define the corresponding branch in the nonreciprocal tree as the one whose leaves are the same. Replacing the previously cut branches of the reciprocal tree by the corresponding branches of the nonreciprocal tree yields the $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ method. Grafting is equivalent to providing the following piecewise definition of the output ultrametric

$$u_X^{\mathrm{R/NR}}(x, x'; \beta) := \begin{cases} u_X^{\mathrm{NR}}(x, x'), & \text{if } u_X^{\mathrm{R}}(x, x') \leq \beta, \\ u_X^{\mathrm{R}}(x, x'), & \text{if } u_X^{\mathrm{R}}(x, x') > \beta. \end{cases} \tag{3.35}$$

For pairs $x, x' \in X$ having large reciprocal ultrametric value we keep this value, whereas for pairs with small reciprocal ultrametric value, we replace it by the nonreciprocal one.

To prove admissibility, we need to show that (3.35) defines an ultrametric and that the method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ satisfies axioms (A1) and (A2). This is asserted in the following proposition.

Proposition 3 *The hierarchical clustering method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ is valid and admissible. I.e., $u_X^{\mathrm{R/NR}}(\beta)$ defined in (3.35) is a valid ultrametric and $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ satisfies axioms (A1)-(A2).*

Proof: The function $u_X^{\mathrm{R/NR}}(\beta)$ fulfills the symmetry and identity properties of ultrametrics because u_X^{NR} and u_X^{R} fulfill them separately. Hence, to show that $u_X^{\mathrm{R/NR}}(\beta)$ is a properly defined ultrametric, we need to show that it satisfies the strong triangle inequality (2.12). To show this, we split the proof into two cases: $u_X^{\mathrm{R}}(x,x') \leq \beta$ and $u_X^{\mathrm{R}}(x,x') > \beta$. Note that, by definition (3.35),

$$u_X^{\mathrm{NR}}(x,x') \leq u_X^{\mathrm{R/NR}}(x,x';\beta) \leq u_X^{\mathrm{R}}(x,x'). \tag{3.36}$$

Starting with the case where $u_X^{\mathrm{R}}(x,x') \leq \beta$, since u_X^{NR} satisfies (2.12) we can state that,

$$u_X^{\mathrm{R/NR}}(x,x';\beta) = u_X^{\mathrm{NR}}(x,x') \leq \max\Big(u_X^{\mathrm{NR}}(x,x''),\, u_X^{\mathrm{NR}}(x'',x')\Big). \tag{3.37}$$

Using the lower bound inequality in (3.36) we can write

$$\max\Big(u_X^{\mathrm{NR}}(x,x''),\, u_X^{\mathrm{NR}}(x'',x')\Big) \leq \max\Big(u_X^{\mathrm{R/NR}}(x,x'';\beta),\, u_X^{\mathrm{R/NR}}(x'',x';\beta)\Big). \tag{3.38}$$

Combining (3.37) and (3.38), we see that $u_X^{\mathrm{R/NR}}(\beta)$ fulfills the strong triangle inequality in this case. As a second case, suppose that $u_X^{\mathrm{R}}(x,x') > \beta$, from the validity of the strong triangle inequality (2.12) for u_X^{R}, we can write

$$\beta < u_X^{\mathrm{R/NR}}(x,x';\beta) = u_X^{\mathrm{R}}(x,x') \leq \max\Big(u_X^{\mathrm{R}}(x,x''),\, u_X^{\mathrm{R}}(x'',x')\Big). \tag{3.39}$$

This implies that at least one of $u_X^{\mathrm{R}}(x,x'')$ and $u_X^{\mathrm{R}}(x'',x')$ is greater than β. When this occurs, $u_X^{\mathrm{R/NR}}(\beta) = u_X^{\mathrm{R}}$. Hence,

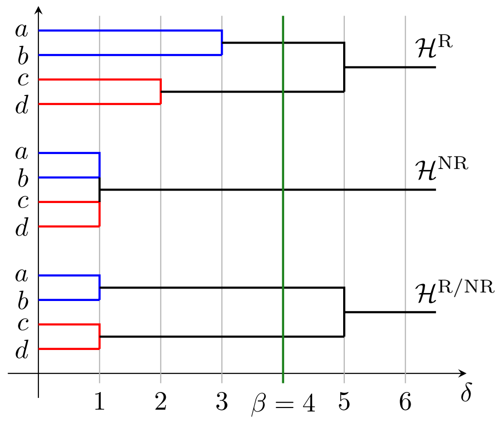
$$\max\Big(u_X^{\mathrm{R}}(x,x''),\, u_X^{\mathrm{R}}(x'',x')\Big) = \max\Big(u_X^{\mathrm{R/NR}}(x,x'';\beta),\, u_X^{\mathrm{R/NR}}(x'',x';\beta)\Big). \tag{3.40}$$

By substituting (3.40) into (3.39), we see that for this second case the strong triangle inequality is also satisfied.

To show that $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ satisfies axiom (A1) it suffices to see that in a two-node network u_X^{NR} and u_X^{R} coincide, meaning that we must have $u_X^{\mathrm{R/NR}}(\beta) = u_X^{\mathrm{NR}} = u_X^{\mathrm{R}}$. Since $\mathcal{H}^{\mathrm{R}}$ and $\mathcal{H}^{\mathrm{NR}}$ fulfill (A1), the method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ must satisfy (A1) as well.

To prove (A2) consider a dissimilarity reducing map $\phi : X \to Y$ and split consideration with regards to whether the reciprocal ultrametric is $u_X^{\mathrm{R}}(x,x') \leq \beta$ or $u_X^{\mathrm{R}}(x,x') > \beta$. When $u_X^{\mathrm{R}}(x,x') \leq \beta$ we must have $u_Y^{\mathrm{R}}(\phi(x),\phi(x')) \leq \beta$ because $\mathcal{H}^{\mathrm{R}}$ satisfies (A2) and ϕ is a dissimilarity reducing map. Hence, according to the definition in (3.35) we must have that both $u_X^{\mathrm{R/NR}}(x,x';\beta)$ and $u_Y^{\mathrm{R/NR}}(\phi(x),\phi(x');\beta)$ coincide with the nonreciprocal ultrametric and, since $\mathcal{H}^{\mathrm{NR}}$ satisfies (A2), it immediately follows that $u_X^{\mathrm{R/NR}}(x,x';\beta) \geq u_Y^{\mathrm{R/NR}}(\phi(x),\phi(x');\beta)$, showing that $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ satisfies (A2) when $u_X^{\mathrm{R}}(x,x') \leq \beta$.

In the second case, when $u_X^{\mathrm{R}}(x,x') > \beta$, the validity of (A2) for the reciprocal ultrametric



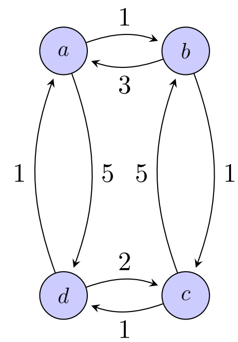


Figure 3.5: Dendrogram grafting. Reciprocal ($\mathcal{H}^{\mathrm{R}}$), nonreciprocal ($\mathcal{H}^{\mathrm{NR}}$), and grafting ($\mathcal{H}^{\mathrm{R/NR}}(\beta = 4)$) dendrograms for the given network are shown – edges not drawn have dissimilarities greater than 5. To form the latter, branches of the reciprocal dendrogram are cut at resolution $\beta = 4$ and replaced by the corresponding branches of the nonreciprocal dendrogram.

u_X^{R} allows us to write

$$u_X^{\mathrm{R/NR}}(x, x'; \beta) = u_X^{\mathrm{R}}(x, x') \geq u_Y^{\mathrm{R}}(\phi(x), \phi(x')). \tag{3.41}$$

Combining this with the fact that u_Y^{R} is an upper bound on $u_Y^{\mathrm{R/NR}}(\beta)$ [cf. (3.36)], we see that $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ satisfies (A2) also for this second case. ■

Notice that, since $u_X^{\mathrm{R/NR}}(x, x'; \beta)$ coincides with either $u_X^{\mathrm{NR}}(x, x')$ or $u_X^{\mathrm{R}}(x, x')$ for all $x, x' \in X$, it satisfies Theorem 4 as it should be the case for any admissible method.

An example implementation of $\mathcal{H}^{\mathrm{R/NR}}(\beta = 4)$ for a particular network is illustrated in Fig. 3.5. The nonreciprocal ultrametric (3.8) is $u_X^{\mathrm{NR}}(x, x') = 1$ for all $x \neq x'$ due to the outmost clockwise loop visiting all nodes at cost 1. This is represented in the nonreciprocal $\mathcal{H}^{\mathrm{NR}}$ dendrogram in Fig. 3.5. For the reciprocal ultrametric (3.2) nodes c and d merge at resolution $u_X^{\mathrm{R}}(c, d) = 2$, nodes a and b at resolution $u_X^{\mathrm{R}}(a, b) = 3$, and they all join together at resolution $\delta = 5$. This can be seen in the reciprocal $\mathcal{H}^{\mathrm{R}}$ dendrogram. To determine $u_X^{\mathrm{R/NR}}(x, x'; 4)$ use the piecewise definition in (3.35). Since the reciprocal ultrametrics $u_X^{\mathrm{R}}(c, d) = 2$ and $u_X^{\mathrm{R}}(a, b) = 3$ are smaller than $\beta = 4$ we set the grafted outcomes to the nonreciprocal ultrametrics to obtain $u_X^{\mathrm{R/NR}}(c, d) = u_X^{\mathrm{NR}}(c, d) = 1$ and $u_X^{\mathrm{R/NR}}(a, b) = u_X^{\mathrm{NR}}(a, b) = 1$. Since the remaining ultrametrics are $u_X^{\mathrm{R}}(x, x') = 5$ which exceed β we set $u_X^{\mathrm{R/NR}}(x, x'; 4) = u_X^{\mathrm{R}}(x, x') = 5$. This yields the $\mathcal{H}^{\mathrm{R/NR}}$ dendrogram in Fig. 3.5 which we interpret as cutting branches from $\mathcal{H}^{\mathrm{R}}$ that we replace by the corresponding branches of $\mathcal{H}^{\mathrm{NR}}$.

In the method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ we use the reciprocal ultrametric as a decision variable in the piecewise definition (3.35) and use nonreciprocal ultrametrics for nodes having small reciprocal ultrametrics. There are three other possible grafting combinations $\mathcal{H}^{\mathrm{R/R}}(\beta)$,

$\mathcal{H}^{\mathrm{NR/R}}(\beta)$ and $\mathcal{H}^{\mathrm{NR/NR}}(\beta)$ depending on which ultrametric is used as decision variable to swap branches and which of the two ultrametrics is used for nodes having small values of the decision ultrametric. E.g., in $\mathcal{H}^{\mathrm{R/R}}(\beta)$, we use reciprocal ultrametrics as decision variables and as the choice for small values of reciprocal ultrametrics,

$$u_X^{\mathrm{R/R}}(x,x';\beta) := \begin{cases} u_X^{\mathrm{R}}(x,x'), & \text{if } u_X^{\mathrm{R}}(x,x') \leq \beta, \\ u_X^{\mathrm{NR}}(x,x'), & \text{if } u_X^{\mathrm{R}}(x,x') > \beta. \end{cases} \tag{3.42}$$

However, the method $\mathcal{H}^{\mathrm{R/R}}(\beta)$ is *not* valid because for some networks the function $u_X^{\mathrm{R/R}}(\beta)$ is not an ultrametric as it violates the strong triangle inequality in (2.12). As a counterexample consider again the network in Fig. 3.5. Applying the definition in (3.42) we obtain that $u_X^{\mathrm{R/R}}(a,b;4) = u_X^{\mathrm{R}}(a,b) = 3$ while $u_X^{\mathrm{R/R}}(a,c;4) = u_X^{\mathrm{NR}}(a,c) = 1$ and similarly $u_X^{\mathrm{R/R}}(c,b;4) = 1$. In turn, this implies that $u_X^{\mathrm{R/R}}(a,b;4) > \max(u_X^{\mathrm{R/R}}(a,c;4), u_X^{\mathrm{R/R}}(c,b;4))$ violating the strong triangle inequality. Analogously, $\mathcal{H}^{\mathrm{NR/NR}}(\beta)$ and $\mathcal{H}^{\mathrm{NR/R}}(\beta)$ can also be shown to be invalid clustering methods.

A second valid grafting alternative can be obtained as a modification of $\mathcal{H}^{\mathrm{R/R}}(\beta)$ in which reciprocal ultrametrics are kept for pairs having small reciprocal ultrametrics, nonreciprocal ultrametrics are used for pairs having large reciprocal ultrametrics, but all nonreciprocal ultrametrics smaller than β are saturated to this value. Denoting the method by $\mathcal{H}^{\mathrm{R/R_{max}}}(\beta)$ the output ultrametrics are thereby given as

$$u_X^{\mathrm{R/R_{max}}}(x,x';\beta) := \begin{cases} u_X^{\mathrm{R}}(x,x'), & \text{if } u_X^{\mathrm{R}}(x,x') \leq \beta, \\ \max\big(\beta, u_X^{\mathrm{NR}}(x,x')\big), & \text{if } u_X^{\mathrm{R}}(x,x') > \beta. \end{cases} \tag{3.43}$$

This alternative definition outputs a valid ultrametric and $\mathcal{H}^{\mathrm{R/R_{max}}}(\beta)$ satisfies axioms (A1)-(A2) as claimed next.

Proposition 4 *The method $\mathcal{H}^{\mathrm{R/R_{max}}}(\beta)$ is valid and admissible. I.e., $u_X^{\mathrm{R/R_{max}}}(\beta)$ defined in* (3.43) *is a valid ultrametric and $\mathcal{H}^{\mathrm{R/R_{max}}}(\beta)$ satisfies axioms (A1)-(A2).*

Proof: This proof follows from a reasoning analogous to that in the proof of Proposition 3. In particular, by definition we have that [cf. (3.36)]

$$u_X^{\mathrm{NR}}(x,x') \leq u_X^{\mathrm{R/R_{max}}}(x,x';\beta) \leq u_X^{\mathrm{R}}(x,x'), \tag{3.44}$$

which immediately implies fulfillment of (A1). Also, as done for Proposition 3, the strong triangle inequality and the fulfillment of (A2) can be shown by dividing the proofs into the two cases $u_X^{\mathrm{R}}(x,x') \leq \beta$ and $u_X^{\mathrm{R}}(x,x') > \beta$. ■

Remark 1 Intuitively, the grafting combination $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ allows nonreciprocal propagation of influence for resolutions smaller than β while requiring reciprocal propagation for higher resolutions. This is of interest if we want tight clusters of small dissimilarity to be formed through loops of influence while looser clusters of higher dissimilarity are required to form through links of bidirectional influence. Conversely, the clustering method $\mathcal{H}^{\mathrm{R/R_{max}}}(\beta)$ requires reciprocal influence within tight clusters of resolution smaller than β but allows nonreciprocal influence in clusters of higher resolutions. This latter behavior is desirable in, e.g., trust propagation in social interactions, where we want tight clusters to be formed through links of mutual trust but allow looser clusters to be formed through unidirectional trust loops.

3.3.2 Convex combinations

A different family of intermediate admissible methods can be constructed by performing a convex combination of methods known to satisfy axioms (A1) and (A2). Indeed, consider two admissible clustering methods $\mathcal{H}^1$ and $\mathcal{H}^2$ and a given parameter $0 \leq \theta \leq 1$. For an arbitrary network $N = (X, A_X)$ denote by $(X, u_X^1) = \mathcal{H}^1(N)$ and $(X, u_X^2) = \mathcal{H}^2(N)$ the respective outcomes of methods $\mathcal{H}^1$ and $\mathcal{H}^2$. Construct then the dissimilarity function $A_X^{12}(\theta)$ as the convex combination of u_X^1 and u_X^2, for all $x, x' \in X$,

$$A_X^{12}(x, x'; \theta) := \theta\, u_X^1(x, x') + (1 - \theta)\, u_X^2(x, x'). \tag{3.45}$$

Although $A_X^{12}(\theta)$ is a well-defined dissimilarity function, it is not an ultrametric in general because it may violate the strong triangle inequality. Nevertheless, we can recover the ultrametric structure by applying any admissible clustering method $\mathcal{H}$ to the symmetric network $N_\theta^{12} = (X, A_X^{12}(\theta))$. Moreover, as explained after Theorem 4, single linkage is the unique admissible clustering method for symmetric networks. Thus, we define the convex combination method $\mathcal{H}_\theta^{12}$ as the application of single linkage on N_θ^{12}. Formally, we define $\mathcal{H}_\theta^{12}$ as a method whose output $(X, u_X^{12}(\theta)) = \mathcal{H}_\theta^{12}(N)$ corresponding to network $N = (X, A_X)$ is given by

$$u_X^{12}(x, x'; \theta) := \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} A_X^{12}(x_i, x_{i+1}; \theta), \tag{3.46}$$

for all $x, x' \in X$ and $A_X^{12}(\theta)$ as given in (3.45). We show that (3.46) defines a valid ultrametric and that $\mathcal{H}_\theta^{12}$ fulfills axioms (A1) and (A2) in the following proposition.

Proposition 5 *Given two admissible hierarchical clustering methods $\mathcal{H}^1$ and $\mathcal{H}^2$, the convex combination method $\mathcal{H}_\theta^{12}$ is valid and admissible. I.e., $u_X^{12}(\theta)$ defined in* (3.46) *is a valid ultrametric and $\mathcal{H}_\theta^{12}$ satisfies axioms (A1)-(A2).*

Proof: As discussed in the paragraph preceding the statement of this proposition, $u_X^{12}(\theta)$ is the output of applying single linkage to the symmetric network N_θ^{12}, immediately implying that $u_X^{12}(\theta)$ is a well-defined ultrametric.

To see that axiom (A1) is fulfilled, pick an arbitrary two-node network $(\{p, q\}, A_{p,q})$ with $A_{p,q}(p, q) = \alpha$ and $A_{p,q}(q, p) = \beta$. Since methods $\mathcal{H}^1$ and $\mathcal{H}^2$ are admissible, in particular they satisfy (A1), hence $u^1_{p,q}(p, q) = u^2_{p,q}(p, q) = \max(\alpha, \beta)$. It then follows from (3.45) that $A^{12}_{p,q}(p, q; \theta) = \max(\alpha, \beta)$ for all possible values of θ. Moreover, since in (3.46) all possible paths joining p and q must contain these two nodes as consecutive elements, we have that

$$u^{12}_{p,q}(p, q; \theta) = A^{12}_{p,q}(p, q; \theta) = \max(\alpha, \beta), \tag{3.47}$$

for all θ, satisfying axiom (A1).

Fulfillment of axiom (A2) also follows from admissibility of $\mathcal{H}^1$ and $\mathcal{H}^2$. Suppose there are two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ and a dissimilarity reducing map $\phi : X \to Y$. From the facts that $\mathcal{H}^1$ and $\mathcal{H}^2$ satisfy (A2) we have

$$u^1_X(x, x') \geq u^1_Y(\phi(x), \phi(x')), \;\; u^2_X(x, x') \geq u^2_Y(\phi(x), \phi(x')). \tag{3.48}$$

By multiplying the left inequality by θ and the right one by $(1 - \theta)$, and adding both inequalities we obtain [cf. (3.45)]

$$A^{12}_X(x, x'; \theta) \geq A^{12}_Y(\phi(x), \phi(x'); \theta), \tag{3.49}$$

for all $0 \leq \theta \leq 1$. This implies that the map ϕ is also dissimilarity reducing between the networks $(X, A^{12}_X(\theta))$ and $(Y, A^{12}_Y(\theta))$. Combining this with the fact that we apply an admissible method (single linkage) to the previous networks to obtain the ultrametric outputs, it follows that

$$u^{12}_X(x, x'; \theta) \geq u^{12}_Y(\phi(x), \phi(x'); \theta), \tag{3.50}$$

for all θ, showing that axiom (A2) is satisfied by the convex combination method. ■

The construction in (3.46) can be generalized to produce intermediate clustering methods generated by convex combinations of any number (i.e. not necessarily two) of admissible methods. These convex combinations can be seen to satisfy axioms (A1) and (A2) through recursive applications of Proposition 5.

Remark 2 Since (3.46) is equivalent to single linkage applied to the symmetric network N_θ^{12}, it follows [10, 11] that $u_X^{12}(\theta)$ is the largest ultrametric bounded above by $A_X^{12}(\theta)$, i.e., the largest ultrametric for which $u^{12}_X(x, x'; \theta) \leq A^{12}_X(x, x'; \theta)$ for all x, x'. We can then think of (3.46) as an operation ensuring a valid ultrametric definition while deviating as

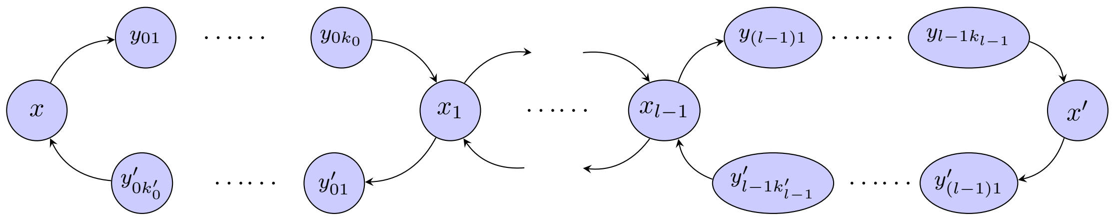


Figure 3.6: Semi-reciprocal paths. The main path joining x and x' is formed by $[x, x_1, ..., x_{l-1}, x']$. Between two consecutive nodes of the main path x_i and x_{i+1}, we have a secondary path in each direction. For $u_X^{\mathrm{SR}(t)}$, the maximum allowed node-length of secondary paths is t.

little as possible from $A_X^{12}(\theta)$, thus, retaining as much information as possible in the convex combination of u_X^1 and u_X^2.

3.3.3 Semi-reciprocal

In reciprocal clustering we require influence to propagate through bidirectional paths; see Fig. 3.1. We could reinterpret bidirectional propagation as allowing loops of node-length two in both directions. E.g., the bidirectional path between x and x_1 in Fig. 3.1 can be interpreted as a loop between x and x_1 composed by two paths $[x, x_1]$ and $[x_1, x]$ of node-length two. *Semi-reciprocal* clustering is a generalization of this concept where loops consisting of at most t nodes in each direction are allowed. Given $t \in \mathbb{N}$ such that $t \geq 2$, we use the notation $P^t_{xx'}$ to denote any path $[x = x_0, x_1, \ldots, x_l = x']$ joining x to x' where $l \leq t-1$. That is, $P^t_{xx'}$ is a path starting at x and finishing at x' with at most t nodes. We reserve the notation $P_{xx'}$ to represent a path from x to x' where no maximum is imposed on the number of nodes. Given an arbitrary network $N = (X, A_X)$, define as $A_X^{\mathrm{SR}(t)}(x, x')$ the minimum cost incurred when traveling from node x to node x' using a path of at most t nodes. I.e.,

$$A_X^{\mathrm{SR}(t)}(x, x') := \min_{P^t_{xx'}} \max_{i | x_i \in P^t_{xx'}} A_X(x_i, x_{i+1}). \tag{3.51}$$

We define the family of semi-reciprocal clustering methods $\mathcal{H}^{\mathrm{SR}(t)}$ with output $(X, u_X^{\mathrm{SR}(t)}) = \mathcal{H}^{\mathrm{SR}(t)}(X, A_X)$ as the one for which the ultrametric $u_X^{\mathrm{SR}(t)}(x, x')$ between x and x' is

$$u_X^{\mathrm{SR}(t)}(x, x') := \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \bar{A}_X^{\mathrm{SR}(t)}(x_i, x_{i+1}), \tag{3.52}$$

where the function $\bar{A}_X^{\mathrm{SR}(t)}$ is defined as

$$\bar{A}_X^{\mathrm{SR}(t)}(x_i, x_{i+1}) := \max\left(A_X^{\mathrm{SR}(t)}(x_i, x_{i+1}), A_X^{\mathrm{SR}(t)}(x_{i+1}, x_i)\right). \tag{3.53}$$

The path $P_{xx'}$ of unconstrained length in (3.52) is called the *main path*, represented by $[x = x_0, x_1, ..., x_{l-1}, x']$ in Fig. 3.6. Between consecutive nodes x_i and x_{i+1} of the main path, we build loops consisting of secondary paths in each direction, represented in Fig. 3.6 by $[x_i, y_{i1}, ..., y_{ik_i}, x_{i+1}]$ and $[x_{i+1}, y'_{i1}, ..., y'_{ik'_i}, x_i]$ for all i. For the computation of $u_X^{\mathrm{SR}(t)}(x, x')$, the maximum allowed length of secondary paths is equal to t nodes, i.e., $k_i, k'_i \leq t-2$ for all i. In particular, for $t = 2$ we recover the reciprocal path; see Fig. 3.1.

We can reinterpret (3.52) as the application of reciprocal clustering [cf. (3.2)] to a network with dissimilarities $A_X^{\mathrm{SR}(t)}$ as in (3.51), i.e., a network with dissimilarities given by the optimal choice of secondary paths. Semi-reciprocal clustering methods are valid and satisfy axioms (A1)-(A2) as shown in the following proposition.

Proposition 6 *The semi-reciprocal clustering method $\mathcal{H}^{\mathrm{SR}(t)}$ is valid and admissible for all integers $t \geq 2$. I.e., $u_X^{\mathrm{SR}(t)}$ is a valid ultrametric and $\mathcal{H}^{\mathrm{SR}(t)}$ satisfies axioms (A1)-(A2).*

Proof: We begin the proof by showing that (3.52) outputs a valid ultrametric where the only non-trivial property to be shown is the strong triangle inequality (2.12). For a fixed t, pick an arbitrary pair of nodes x and x' and an arbitrary intermediate node x''. Let us denote by $P^*_{xx''}$ and $P^*_{x''x'}$ a pair of main paths that satisfy definition (3.52) for $u_X^{\mathrm{SR}(t)}(x, x'')$ and $u_X^{\mathrm{SR}(t)}(x'', x')$ respectively. Construct $P_{xx'}$ by concatenating the aforementioned minimizing paths $P^*_{xx''}$ and $P^*_{x''x'}$. However, $P_{xx'}$ is a particular path for computing $u_X^{\mathrm{SR}(t)}(x, x')$ and need not be the minimizing one. This implies that

$$u_X^{\mathrm{SR}(t)}(x, x') \leq \max\Big(u_X^{\mathrm{SR}(t)}(x, x''), u_X^{\mathrm{SR}(t)}(x'', x')\Big), \tag{3.54}$$

proving the strong triangle inequality.

To show fulfillment of (A1), consider the network $(\{p, q\}, A_{p,q})$ with $A_{p,q}(p, q) = \alpha$ and $A_{p,q}(q, p) = \beta$. Note that in this situation, $A_{p,q}^{\mathrm{SR}(t)}(p, q) = \alpha$ and $A_{p,q}^{\mathrm{SR}(t)}(q, p) = \beta$ for all $t \geq 2$ [cf. (3.51)], since there is only one possible path between them and contains only two nodes. Hence, from (3.52),

$$u_{p,q}^{\mathrm{SR}(t)}(p, q) = \max(\alpha, \beta), \tag{3.55}$$

for all t. Consequently, axiom (A1) is satisfied.

To show fulfillment of (A2), consider two arbitrary networks (X, A_X) and (Y, A_Y) and a dissimilarity reducing map $\phi : X \to Y$ between them. Further, denote by $P_{xx'}^{X*} = [x = x_0, \ldots, x_l = x']$ a main path that achieves the minimum semi-reciprocal cost in (3.52). Then, for a fixed t, we can write

$$u_X^{\mathrm{SR}(t)}(x, x') = \max_{i | x_i \in P_{xx'}^{X*}} \bar{A}_X^{\mathrm{SR}(t)}(x_i, x_{i+1}). \tag{3.56}$$

Consider now a secondary path $P^t_{x_i x_{i+1}} = [x_i = x^{(0)}, \dots, x^{(l')} = x_{i+1}]$ between two consecutive nodes x_i and x_{i+1} of the minimizing path $P^{X*}_{xx'}$. Further, focus on the image of this secondary path under the map ϕ, that is $P^t_{\phi(x_i)\phi(x_{i+1})} := \phi\big(P^t_{x_i,x_{i+1}}\big) = [\phi(x_i) = \phi(x^{(0)}), \dots, \phi(x^{(l')}) = \phi(x_{i+1})]$ in the set Y.

Since the map $\phi : X \to Y$ is dissimilarity reducing, $A_Y(\phi(x^{(i)}), \phi(x^{(i+1)})) \leq A_X(x^{(i)}, x^{(i+1)})$ for all links in this path. Analogously, we can bound the dissimilarities in secondary paths $P^t_{x_{i+1},x_i}$ from x_{i+1} back to x_i. Thus, from (3.51) we can state that,

$$\bar{A}^{\mathrm{SR}(t)}_X(x_i, x_{i+1}) \geq \bar{A}^{\mathrm{SR}(t)}_Y(\phi(x_i), \phi(x_{i+1})). \tag{3.57}$$

Denote by $P^Y_{\phi(x)\phi(x')}$ the image of the main path $P^{X*}_{xx'}$ under the map ϕ. Notice that $P^Y_{\phi(x)\phi(x')}$ is a particular path joining $\phi(x)$ and $\phi(x')$, whereas the semi-reciprocal ultrametric computes the minimum across all main paths. Therefore,

$$u^{\mathrm{SR}(t)}_Y(\phi(x), \phi(x')) \leq \max_{i | \phi(x_i) \in P^Y_{\phi(x)\phi(x')}} \bar{A}^{\mathrm{SR}(t)}_Y(\phi(x_i), \phi(x_{i+1})). \tag{3.58}$$

By bounding the right-hand side of (3.58) using (3.57) and recalling (3.56), it follows that $u^{\mathrm{SR}(t)}_Y(\phi(x), \phi(x')) \leq u^{\mathrm{SR}(t)}_X(x, x')$. This proves that (A2) is satisfied. ■

The semi-reciprocal family is a countable family of clustering methods parameterized by integer $t \geq 2$ representing the allowed maximum node-length of secondary paths. Reciprocal and nonreciprocal ultrametrics are equivalent to semi-reciprocal ultrametrics for specific values of t. For $t = 2$ we have $u^{\mathrm{SR}(2)}_X = u^{\mathrm{R}}_X$ meaning that we recover reciprocal clustering. To see this formally, note that $A^{\mathrm{SR}(2)}_X(x, x') = A_X(x, x')$ [cf. (3.51)] since the only path of length two joining x and x' is $[x, x']$. Hence, for $t = 2$, (3.52) reduces to

$$u^{\mathrm{SR}(2)}_X(x, x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \bar{A}_X(x_i, x_{i+1}), \tag{3.59}$$

which is the definition of the reciprocal ultrametric [cf. (3.2)]. Nonreciprocal ultrametrics can be obtained as $u^{\mathrm{SR}(t)}_X = u^{\mathrm{NR}}_X$ for any parameter t exceeding the number of nodes in the network analyzed. To see this, notice that minimizing over $P_{xx'}$ is equivalent to minimizing over $P^t_{xx'}$ for all $t \geq n$, since we are looking for minimizing paths in a network with nonnegative dissimilarities. Therefore, visiting the same node twice is not an optimal choice. This implies that $P^n_{xx'}$ contains all possible minimizing paths between x and x'. Hence, by inspecting (3.51), $A^{\mathrm{SR}(t)}_X(x, x') = \tilde{u}^*_X(x, x')$ [cf. (2.7)] for all $t \geq n$. Furthermore, when $t \geq n$, the best main path that can be picked is formed only by nodes x and x' because, in this way, no additional meeting point is enforced between the paths going from x to x' and

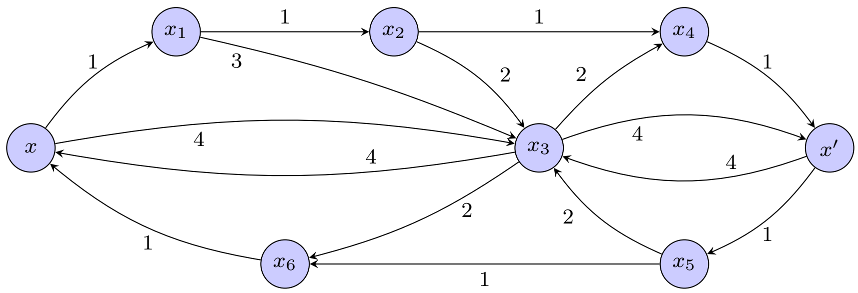


Figure 3.7: Semi-reciprocal example. Computation of semi-reciprocal ultrametrics between nodes x and x' for different values of parameter t; see text for details.

vice versa. As a consequence, definition (3.52) reduces to

$$u_X^{\mathrm{SR}(t)}(x, x') = \max\Big(\tilde{u}_X^*(x, x'), \tilde{u}_X^*(x', x)\Big), \tag{3.60}$$

for all $x, x' \in X$ and for all $t \geq n$. The right hand side of (3.60) is the definition of the nonreciprocal ultrametric [cf. (3.8)].

For the network in Fig. 3.7, we compute the semi-reciprocal ultrametrics between x and x' for different values of t. The edges which are not delineated are assigned dissimilarity values greater than 4. Since the only bidirectional path between x and x' uses x_3 as the intermediate node, we conclude that $u_X^{\mathrm{R}}(x, x') = u_X^{\mathrm{SR}(2)}(x, x') = 4$. Furthermore, by constructing a path through the outermost clockwise cycle in the network, we conclude that $u_X^{\mathrm{NR}}(x, x') = 1$. Since the longest secondary path in the minimizing path for the nonreciprocal case, $[x, x_1, x_2, x_4, x']$, has node-length 5, we may conclude that $u_X^{\mathrm{SR}(t)}(x, x') = 1$ for all $t \geq 5$. For intermediate values of t, if e.g., we fix $t = 3$, the minimizing path is given by the main path $[x, x_3, x']$ and the secondary paths $[x, x_1, x_3]$, $[x_3, x_4, x']$, $[x', x_5, x_3]$ and $[x_3, x_6, x]$ joining consecutive nodes in the main path in both directions. The maximum cost among all dissimilarities in this path is $A_X(x_1, x_3) = 3$. Hence, $u_X^{\mathrm{SR}(3)}(x, x') = 3$. The minimizing path for $t = 4$ is similar to the minimizing one for $t = 3$ but replacing the secondary path $[x, x_1, x_3]$ by $[x, x_1, x_2, x_3]$. In this way, we obtain $u_X^{\mathrm{SR}(4)}(x, x') = 2$.

Remark 3 Intuitively, when propagating influence through a network, reciprocal clustering requires bidirectional influence whereas nonreciprocal clustering allows arbitrarily large unidirectional cycles. In many applications, such as trust propagation in social networks, it is reasonable to look for an intermediate situation where influence can propagate through cycles but of limited length. Semi-reciprocal ultrametrics represent this intermediate situation where the parameter t represents the maximum length of paths through which influence can propagate in a nonreciprocal manner.

3.4 Alternative axiomatic constructions

The axiomatic framework that we adopted allows alternative constructions by modifying the underlying set of axioms. Among the axioms in Section 2.2, the Axiom of Value (A1) is perhaps the most open to interpretation. Although we required the two-node network in Fig. 2.3 to first cluster into one single block at resolution $\max(\alpha, \beta)$ corresponding to the largest dissimilarity and argued that this was reasonable in most situations, it is also reasonable to accept that in some situations the two nodes should be clustered together as long as one of them is able to influence the other. To account for this possibility we replace the Axiom of Value by the following alternative.

> *(A1") Alternative Axiom of Value.* The ultrametric $(\{p, q\}, u_{p,q}) := \mathcal{H}(\vec{\Delta}_2(\alpha, \beta))$ output by $\mathcal{H}$ from the two-node network $\vec{\Delta}_2(\alpha, \beta)$ satisfies $u_{p,q}(p, q) = \min(\alpha, \beta)$.

Axiom (A1") replaces the requirement of bidirectional influence in axiom (A1) to unidirectional influence; see Fig. 3.8. We say that a clustering method $\mathcal{H}$ is admissible with respect to the alternative axioms if it satisfies axioms (A1") and (A2).

The Property of Influence (P1), which is a keystone in the proof of Theorem 4, is not compatible with the Alternative Axiom of Value (A1"). Indeed, just observe that the minimum loop cost of the two-node network in Fig. 3.8 is $\text{mlc}(\vec{\Delta}_2(\alpha, \beta)) = \max(\alpha, \beta)$ whereas in (A1") we are requiring the output ultrametric to be $u_{p,q}(p, q) = \min(\alpha, \beta)$. We therefore have that axiom (A1") itself implies $u_{p,q}(p, q) = \min(\alpha, \beta) < \max(\alpha, \beta) = \text{mlc}(\vec{\Delta}_2(\alpha, \beta))$ for the cases when $\alpha \neq \beta$. Thus, we reformulate (P1) into the Alternative Property of Influence (P1') that we define next.

> *(P1') Alternative Property of Influence.* For any network $N_X = (X, A_X)$ the output ultrametric $(X, u_X) = \mathcal{H}(N_X)$ is such that $u_X(x, x')$ for distinct points cannot be smaller than the separation of the network, $u_X(x, x') \geq \text{sep}(N_X)$ for all $x \neq x'$.

Observe that the Alternative Property of Influence (P1') coincides with the Symmetric Property of Influence (Q1) defined in Section 3.2.1. This is not surprising because for symmetric networks the Axiom of Value (A1) and the Alternative Axiom of Value (A1") impose identical restrictions. Moreover, since the separation of a network cannot be larger than its minimum loop cost, the Alternative Property of Influence (P1') is implied by the (regular) Property of Influence (P1), but not vice versa.

The Alternative Property of Influence (P1') states that no clusters are formed at resolutions at which there are no unidirectional influences between any pair of nodes and is consistent with the Alternative Axiom of Value (A1"). Moreover, in studying methods admissible with respect to (A1") and (A2), (P1') plays a role akin to the one played by (P1) when studying methods that are admissible with respect to (A1) and (A2). In particular, as (P1) is implied by (A1) and (A2), (P1') is true if (A1") and (A2) hold as we assert in

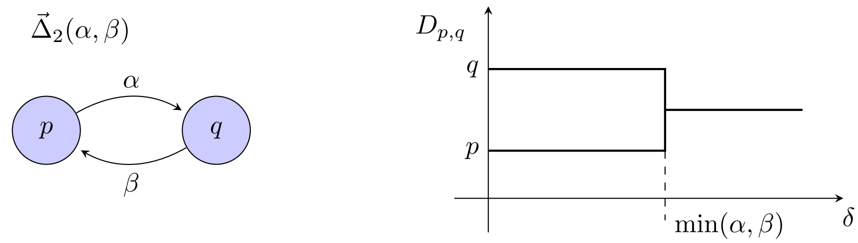


Figure 3.8: Alternative Axiom of Value. For a two-node network, nodes are clustered together at the minimum resolution at which one of them can influence the other.

the following theorem.

Theorem 5 *If a clustering method $\mathcal{H}$ satisfies the Alternative Axiom of Value (A1") and the Axiom of Transformation (A2) then it also satisfies the Alternative Property of Influence (P1').*

Proof: Suppose there exists a clustering method $\mathcal{H}$ that satisfies axioms (A1") and (A2) but does not satisfy property (P1'). This means that there exists a network $N = (X, A_X)$ with output ultrametrics $(X, u_X) = \mathcal{H}(N)$ for which $u_X(x_1, x_2) < \text{sep}(X, A_X)$ for at least one pair of nodes $x_1 \neq x_2 \in X$. Focus on a symmetric two-node network $\vec{\Delta}_2(s, s) = (\{p, q\}, A_{p,q})$ with $A_{p,q}(p, q) = A_{p,q}(q, p) = s = \text{sep}(X, A_X)$ and define $(X, u_{p,q}) = \mathcal{H}(\vec{\Delta}_2(s, s))$. From axiom (A1"), we must have that

$$u_{p,q}(p, q) = \min\Big(\text{sep}(X, A_X), \text{sep}(X, A_X)\Big) = \text{sep}(X, A_X). \tag{3.61}$$

Construct the map $\phi : X \to \{p, q\}$ from the network N to $\vec{\Delta}_2(s, s)$ that takes node x_1 to $\phi(x_1) = p$ and every other node $x \neq x_1$ to $\phi(x) = q$. No dissimilarity can be increased when applying ϕ since every dissimilarity is mapped either to zero or to $\text{sep}(X, A_X)$ which is by definition the minimum dissimilarity in the original network [cf. (2.10)]. Hence, ϕ is dissimilarity reducing and from axiom (A2) it follows that $u_X(x_1, x_2) \geq u_{p,q}(\phi(x_1), \phi(x_2)) = u_{p,q}(p, q)$. By substituting (3.61) into the previous expression, we contradict $u_X(x_1, x_2) < \text{sep}(X, A_X)$ proving that such method $\mathcal{H}$ cannot exist. ■

Theorem 5 admits the following interpretation. In (A1") we require two-node networks to cluster at the resolution where unidirectional influence occurs. When we consider (A1") in conjunction with (A2) we can translate this requirement into a statement about clustering in arbitrary networks. Such requirement is the Alternative Property of Influence (P1') which prevents nodes to cluster at resolutions at which no influence exists between *any* two nodes.

3.4.1 Unilateral clustering

Mimicking the developments in Sections 3.1 and 3.2, we move on to identify and define methods that satisfy axioms (A1")-(A2) and then bound the range of admissible methods with respect to these axioms. To do so, let $N = (X, A_X)$ be a given network and consider the dissimilarity function $\hat{A}_X(x, x') := \min(A_X(x, x'), A_X(x', x))$, for all $x, x' \in X$. Notice that, as opposed to the definition of $\bar{A}_X$ where the symmetrization is done by means of a max operation, $\hat{A}$ is defined by using a min operation. We define the *unilateral* clustering method $\mathcal{H}^{\mathrm{U}}$ with output ultrametric $(X, u_X^{\mathrm{U}}) = \mathcal{H}^{\mathrm{U}}(N)$, where u_X^{U} is defined as

$$u_X^{\mathrm{U}}(x, x') := \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \hat{A}_X(x_i, x_{i+1}), \tag{3.62}$$

for all $x, x' \in X$. To show that $\mathcal{H}^{\mathrm{U}}$ is a properly defined clustering method, we need to establish that u_X^{U} as defined in (3.62) is a valid ultrametric. However, comparing (3.62) and (2.15) we see that $\mathcal{H}^{\mathrm{U}}(X, A_X) \equiv \mathcal{H}^{\mathrm{SL}}(X, \hat{A}_X)$, i.e. applying the unilateral clustering method to an asymmetric network (X, A_X) is equivalent to applying single linkage clustering method to the symmetrized network $(X, \hat{A}_X)$. Since we know that single linkage produces a valid ultrametric when applied to any symmetric network such as $(X, \hat{A}_X)$, (3.62) is a properly defined ultrametric. Furthermore, it can be shown that $\mathcal{H}^{\mathrm{U}}$ satisfies axioms (A1") and (A2).

Proposition 7 *The unilateral clustering method $\mathcal{H}^{\mathrm{U}}$ with output ultrametrics defined in* (3.62) *satisfies axioms (A1") and (A2).*

Proof: To show fulfillment of (A1"), consider the network $\vec{\Delta}_2(\alpha, \beta)$ and define $(\{p, q\}, u_{p,q}^{\mathrm{U}}) := \mathcal{H}^{\mathrm{U}}(\vec{\Delta}_2(\alpha, \beta))$. Since every path connecting p and q must contain these two nodes as consecutive nodes, applying the definition in (3.62) yields $u_{p,q}^{\mathrm{U}}(p, q) = \min\big(A_{p,q}(p, q), A_{p,q}(q, p)\big) = \min(\alpha, \beta)$, and axiom (A1") is thereby satisfied. In order to show fulfillment of axiom (A2), the proof is analogous to the one developed in Proposition 1. The proof only differs in the appearance of minimizations instead of maximizations to account for the difference in the definitions of unilateral and reciprocal ultrametrics [cf. (3.62) and (3.2)]. ■

In the case of admissibility with respect to (A1) and (A2), nonreciprocal and reciprocal clustering are two different admissible methods which bound every other possible clustering method satisfying (A1)-(A2) (cf. Theorem 4). In contrast, in the case of admissibility with respect to (A1") and (A2), unilateral clustering is the *unique* admissible method as stated in the following theorem.

Theorem 6 *Let $\mathcal{H}$ be a hierarchical clustering method satisfying axioms (A1") and (A2). Then, $\mathcal{H} \equiv \mathcal{H}^{\mathrm{U}}$ where $\mathcal{H}^{\mathrm{U}}$ is the unilateral clustering.*

Proof: Given an arbitrary network (X, A_X), denote by $\mathcal{H}$ a clustering method that fulfills axioms (A1") and (A2) and define $(X, u_X) := \mathcal{H}(X, A_X)$. Then, we show the theorem by proving the following inequalities for all nodes $x, x' \in X$,

$$u_X^{\mathrm{U}}(x, x') \leq u_X(x, x') \leq u_X^{\mathrm{U}}(x, x'). \tag{3.63}$$

Proof of leftmost inequality in (3.63)**:** Consider the unilateral clustering equivalence relation $\sim_{\mathrm{U}_X(\delta)}$ according to which $x \sim_{\mathrm{U}_X(\delta)} x'$ if and only if x and x' belong to the same unilateral cluster at resolution δ. That is, $x \sim_{\mathrm{U}_X(\delta)} x' \iff u_X^{\mathrm{U}}(x, x') \leq \delta$. Further, as in the proof of Theorem 4, consider the set Z of equivalence classes at resolution δ. That is, $Z := X \mod \sim_{\mathrm{U}_X(\delta)}$. Also, consider the map $\phi_\delta : X \to Z$ that maps each point of X to its equivalence class. Notice that x and x' are mapped to the same point z if and only if they belong to the same block at resolution δ, consequently $\phi_\delta(x) = \phi_\delta(x') \iff u_X^{\mathrm{U}}(x, x') \leq \delta$. We define the network $N_Z = (Z, A_Z)$ by endowing Z with the dissimilarity function A_Z derived from A_X as explained in (3.17). For further details on this construction, review the corresponding proof in Theorem 4 and see Fig. 3.4. We stress the fact that the map ϕ_δ is dissimilarity reducing for all δ.

Claim 2 *The separation of the equivalence class network N_Z is* $\mathrm{sep}(N_Z) > \delta$.

Proof: First, observe that by definition of unilateral clustering (3.62), we know that,

$$u_X^{U}(x, x') \leq \min(A_X(x, x'), A_X(x', x)), \tag{3.64}$$

since a two-node path between nodes x and x' is a particular path joining the two nodes whereas the ultrametric is calculated as the minimum over all paths. Now, assume that $\mathrm{sep}(N_Z) \leq \delta$. Therefore, by (3.17) there exists a pair of nodes x and x' that belong to different equivalence classes and have $A_X(x, x') \leq \delta$. However, if x and x' belong to different equivalence classes, they cannot be clustered at resolution δ, hence, $u_X^{U}(x, x') > \delta$. Inequalities $A_X(x, x') \leq \delta$ and $u_X^{U}(x, x') > \delta$ cannot hold simultaneously since they contradict (3.64). Thus, it must be that $\mathrm{sep}(N_Z) > \delta$. ■

Define $(Z, u_Z) := \mathcal{H}(Z, A_Z)$ and, since $\mathrm{sep}(N_Z) > \delta$ (cf. Claim 2), it follows from property (P1') that for all $z \neq z'$ it holds $u_Z(z, z') > \delta$. Further, recalling that ϕ_δ is a dissimilarity reducing map, from axiom (A2) we must have $u_X(x, x') \geq u_Z(\phi_\delta(x), \phi_\delta(x')) = u_Z(z, z')$ for some $z, z' \in Z$. This fact, combined with $u_Z(z, z') > \delta$, entails that when $\phi_\delta(x)$ and $\phi_\delta(x')$ belong to different equivalence classes $u_X(x, x') \geq u_Z(\phi(x), \phi(x')) > \delta$. Notice now that $\phi_\delta(x)$ and $\phi_\delta(x')$ belonging to different equivalence classes is equivalent to $u_X^{\mathrm{U}}(x, x') > \delta$. Hence, we can state that $u_X^{\mathrm{U}}(x, x') > \delta$ implies $u_X(x, x') > \delta$ for any arbitrary $\delta > 0$. In set notation, $\{(x, x') : u_X^{\mathrm{U}}(x, x') > \delta\} \subseteq \{(x, x') : u_X(x, x') > \delta\}$. Since the previous expression is true for arbitrary $\delta > 0$, this implies that $u_X^{\mathrm{U}}(x,x') \leq u_X(x,x')$, proving the left inequality in (3.63). ■

Proof of rightmost inequality in (3.63)**:** Consider two nodes x and x' with unilateral ultrametric value $u_X^{\mathrm{U}}(x,x') = \delta$. Let $P^*_{xx'} = [x = x_0, \dots, x_l = x']$ be a minimizing path in the definition (3.62) so that we can write

$$\delta = u_X^{\mathrm{U}}(x,x') = \max_{i | x_i \in P^*_{xx'}} \min\Big(A_X(x_i, x_{i+1}), A_X(x_{i+1}, x_i)\Big). \tag{3.65}$$

Consider the two-node network $\vec{\Delta}_2(\delta, M) = (\{p,q\}, A_{p,q})$ where $M := \max_{x,x'} A_X(x,x')$ and define $(\{p,q\}, u_{p,q}) := \mathcal{H}(\{p,q\}, A_{p,q})$. Notice that according to axiom (A1") we have $u_{p,q}(p,q) = u_{p,q}(q,p) = \min(\delta, M) = \delta$, where the last equality is enforced by the definition of M.

Focus now on each link of the minimizing path in (3.65). For every successive pair of nodes x_i and x_{i+1}, we must have

$$\max\Big(A_X(x_i, x_{i+1}), A_X(x_{i+1}, x_i)\Big) \leq M, \tag{3.66}$$

$$\min\Big(A_X(x_i, x_{i+1}), A_X(x_{i+1}, x_i)\Big) \leq \delta. \tag{3.67}$$

Expression (3.66) is true since M is defined as the maximum dissimilarity in A_X. Inequality (3.67) is justified by (3.65), since δ is defined as the maximum among links of the minimum distance in both directions of the link. This observation allows the construction of dissimilarity reducing maps $\phi_i : \{p,q\} \to X$,

$$\phi_i := \begin{cases} \phi_i(p) = x_i, \phi_i(q) = x_{i+1}, & \text{if } \hat{A}_X(x_i, x_{i+1}) = A_X(x_i, x_{i+1}) \\ \phi_i(q) = x_i, \phi_i(p) = x_{i+1}, & \text{otherwise.} \end{cases} \tag{3.68}$$

In this way, we can map p and q to subsequent nodes in the path $P_{xx'}$ used in (3.65). Inequalities (3.66) and (3.67) combined with the map definition in (3.68) guarantee that ϕ_i is a dissimilarity reducing map for every i. Since clustering method $\mathcal{H}$ satisfies axiom (A2), it follows that

$$u_X(\phi_i(p), \phi_i(q)) \leq u_{p,q}(p,q) = \delta, \quad \text{for all } i. \tag{3.69}$$

Substituting $\phi_i(p)$ and $\phi_i(q)$ in (3.69) by the corresponding nodes given by the definition (3.68), we can write $u_X(x_i, x_{i+1}) = u_X(x_{i+1}, x_i) \leq \delta$, for all i, where the symmetry property of ultrametrics was used. To complete the proof we invoke the strong triangle inequality (2.12) and apply it to $P_{xx'} = [x = x_0, \dots, x_l = x']$, the minimizing path in (3.65). As a consequence, $u_X(x,x') \leq \max_i u_X(x_i, x_{i+1}) \leq \delta$. The proof of the right inequality in (3.63) is completed by substituting $\delta = u^{\mathrm{U}}_X(x,x')$ [cf. (3.65)] into the last previous expression. ■

Having proved both inequalities in (3.63), unilateral clustering is the only method that satisfies axioms (A1") and (A2), completing the global proof. ■

By Theorem 6, the space of methods that satisfy the Alternative Axiom of Value (A1") and the Axiom of Transformation (A2) is inherently simpler than the space of methods that satisfy the (regular) Axiom of value (A1) and the Axiom of Transformation (A2). Further note that in the case of symmetric networks, for all $x, x' \in X$ we have $\hat{A}_X(x,x') = A_X(x,x') = A_X(x',x)$ and, as a consequence, unilateral clustering is equivalent to single linkage as it follows from comparison of (2.15) and (3.62). Thus, the result in Theorem 6 reduces to the statement in Corollary 2, which was derived upon observing that in symmetric networks reciprocal and nonreciprocal clustering yield identical outcomes. The fact that reciprocal, nonreciprocal, and unilateral clustering all coalesce into single linkage when restricted to symmetric networks is consistent with the fact that the Axiom of Value (A1) and the Alternative Axiom of Value (A1") are both equivalent to the Symmetric Axiom of Value (B1) when restricted to symmetric dissimilarities.

3.4.2 Agnostic Axiom of Value

Axiom (A1) stipulates that every two-node network $\vec{\Delta}_2(\alpha,\beta)$ is clustered into a single block at resolution $\max(\alpha,\beta)$, whereas axiom (A1") stipulates that they should be clustered at $\min(\alpha,\beta)$. One can also be agnostic with respect to this issue and say that both of these situations are admissible. An agnostic version of axioms (A1) and (A1") is given next.

(A1"') Agnostic Axiom of Value. The ultrametric $(X, u_{p,q}) = \mathcal{H}(\vec{\Delta}_2(\alpha,\beta))$ produced by $\mathcal{H}$ applied to the two-node network $\vec{\Delta}_2(\alpha,\beta)$ satisfies $\min(\alpha,\beta) \le u_X(p,q) \le \max(\alpha,\beta)$.

Since fulfillment of (A1) or (A1") implies fulfillment of (A1"'), any admissible clustering method with respect to the original axioms (A1)-(A2) or with respect to the alternative axioms (A1")-(A2) must be admissible with respect to the agnostic axioms (A1"')-(A2). In this sense, (A1"')-(A2) is the most general combination of axioms described in this first part of the thesis. For methods that are admissible with respect to (A1"') and (A2) we can bound the range of outcome ultrametrics as stated next.

Theorem 7 *Consider a clustering method $\mathcal{H}$ satisfying axioms (A1"') and (A2). For an arbitrary given network $N = (X, A_X)$ denote by $(X, u_X) = \mathcal{H}(X, A_X)$ the outcome of $\mathcal{H}$ applied to N. Then, for all pairs of nodes $x, x' \in X$*

$$u^{\mathrm{U}}_X(x,x') \le u_X(x,x') \le u^{\mathrm{R}}_X(x,x'), \tag{3.70}$$

where $u^{\mathrm{U}}_X(x,x')$ and $u^{\mathrm{R}}_X(x,x')$ denote the unilateral and reciprocal ultrametrics as defined by (3.62) *and* (3.2)*, respectively.*

Proof: The leftmost inequality in (3.70) can be proved using the same method of proof used for the leftmost inequality in (3.63) within the proof of Theorem 6. The proof of the rightmost inequality in (3.70) is equivalent to the proof of the rightmost inequality in Theorem 4. ■

By Theorem 7, given an asymmetric network (X, A_X), any hierarchical clustering method abiding by axioms (A1''') and (A2) produces outputs contained between those corresponding to two methods. The first method, unilateral clustering, symmetrizes A_X by calculating $\hat{A}_X(x,x') = \min(A_X(x,x'), A_X(x',x))$ for all $x, x' \in X$ and computes single linkage on $(X, \hat{A}_X)$. The other method, reciprocal clustering, symmetrizes A_X by calculating $\bar{A}_X(x,x') = \max(A_X(x,x'), A_X(x',x))$ for all $x, x' \in X$ and computes single linkage on $(X, \bar{A}_X)$.

3.5 Algorithms

Recall that, for convenience, we can interpret the dissimilarity function A_X as an $n \times n$ matrix and, similarly, u_X can be regarded as a matrix of ultrametrics. By (3.2), reciprocal clustering searches for paths that minimize their maximum dissimilarity in the symmetric matrix $\bar{A}_X := \max(A_X, A_X^T)$, where the max is applied element-wise. This is equivalent to finding paths in $\bar{A}_X$ that have minimum cost in a ℓ_∞ sense. Likewise, nonreciprocal clustering searches for directed paths of minimum cost in A_X to construct the matrix $\tilde{u}^*_X$ [cf. (2.7)] and selects the maximum of the directed costs by performing the operation $u^{\mathrm{NR}}_X = \max(\tilde{u}^*_X, \tilde{u}^{*T}_X)$ [cf. (3.8)]. These operations can be performed algorithmically using matrix powers in the dioid algebra $\mathfrak{A} := (\bar{\mathbb{R}}_+ \cup, \min, \max)$ where $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$ [32].

In $\mathfrak{A}$, the regular sum is replaced by the minimization operator and the regular product by maximization. Indeed, using $\oplus$ and $\otimes$ to denote sum and product, respectively, on this dioid algebra we have $a \oplus b := \min(a,b)$ and $a \otimes b := \max(a,b)$ for all $a, b \in \bar{\mathbb{R}}_+$. In the algebra $\mathfrak{A}$, the matrix product $A \otimes B$ of two real valued matrices of compatible sizes is therefore given by the matrix with entries

$$\big[A \otimes B\big]_{ij} := \bigoplus_{k=1}^{n} \big(A_{ik} \otimes B_{kj}\big) = \min_{k \in \{1,..,n\}} \max\big(A_{ik}, B_{kj}\big). \tag{3.71}$$

For integers $k \geq 2$ dioid matrix powers $A^k_X := A_X \otimes A^{k-1}_X$ with $A^1_X := A_X$ of a dissimilarity matrix are related to ultrametric matrices u_X. We delve into this relationship in the next section.

3.5.1 Dioid powers and ultrametrics

Notice that the elements of the dioid power u_X^2 of a given ultrametric matrix u_X are given by

$$\big[u_X^2\big]_{ij} = \min_{k\in\{1,..,n\}} \max\big([u_X]_{ik}, [u_X]_{kj}\big). \tag{3.72}$$

Since u_X satisfies the strong triangle inequality we have that $[u_X]_{ij} \leq \max\big([u_X]_{ik}, [u_X]_{kj}\big)$ for all $k \in \{1, .., n\}$. And for $k = j$ in particular we further have that $\max\big([u_X]_{ij}, [u_X]_{jj}\big) = \max\big([u_X]_{ij}, 0\big) = [u_X]_{ij}$. Combining these two observations it follows that the result of the minimization in (3.72) is $\big[u_X^2\big]_{ij} = \big[u_X\big]_{ij}$ since none of its arguments is smaller that $[u_X]_{ij}$ and one of them is exactly $[u_X]_{ij}$. This being valid for all i, j implies

$$u_X^2 = u_X. \tag{3.73}$$

Furthermore, a matrix having the property in (3.73) is such that $\big[u_X\big]_{ij} = \big[u_X^2\big]_{ij} = \min_{k\in\{1,..,n\}} \max\big([u_X]_{ik}, [u_X]_{kj}\big) \leq \max\big([u_X]_{il}, [u_X]_{lj}\big)$ for all l, which is just a restatement of the strong triangle inequality. Therefore, a non-negative matrix u_X represents a finite ultrametric if and only if (3.73) is true, has null diagonal elements and positive off-diagonal elements, and is symmetric, $u_X = u_X^T$. From definition (3.71) it follows that the l-th dioid power A_X^l is such that its entry $[A_X^l]_{ij}$ represents the minimum cost of a path from node i to j containing at most l hops. We then expect dioid powers to play a key role in the construction of ultrametrics.

The *quasi-inverse* of a matrix in a dioid algebra is a useful concept that simplifies the proofs within this section. In any dioid algebra we call quasi-inverse of A, denoted by $A^\dagger$, to the limit, when it exists, of the sequence of matrices [32, Ch.4, Def. 3.1.2]

$$A^\dagger := \lim_{k\to\infty} I \oplus A \oplus A^2 \oplus ... \oplus A^k, \tag{3.74}$$

where I has zeros in the diagonal and $+\infty$ in the off-diagonal elements. The utility of the quasi-inverse resides in the fact that, given a dissimilarity matrix A_X, then [32, Ch.6, Sec 6.1]

$$[A_X^\dagger]_{ij} = \min_{P_{x_i x_j}} \max_{k | x_k \in P_{x_i x_j}} A_X(x_k, x_{k+1}). \tag{3.75}$$

I.e., the elements of the quasi-inverse $A_X^\dagger$ correspond to the directed minimum path costs $\tilde{u}_X^*$ of the associated network (X, A_X) as defined in (2.7).

3.5.2 Algorithms for admissible clustering methods

The reciprocal and nonreciprocal ultrametrics can be obtained via simple dioid matrix operations, as stated next.

Theorem 8 *For any network $N = (X, A_X)$ with n nodes the reciprocal ultrametric u_X^{R} defined in* (3.2) *can be computed as*

$$u_X^{\mathrm{R}} = \Big(\max\big(A_X, A_X^T\big)\Big)^{n-1}, \tag{3.76}$$

where the matrix operations are in the dioid algebra $\mathfrak{A}$. Similarly, the nonreciprocal ultrametric u_X^{NR} defined in (3.8) *can be computed as*

$$u_X^{\mathrm{NR}} = \max\Big(A_X^{n-1}, \big(A_X^T\big)^{n-1}\Big). \tag{3.77}$$

Proof: By comparing (3.75) with (2.7), we can see that $A_X^{\dagger} = \tilde{u}_X^*$ from where it follows [cf. (3.8)]

$$u_X^{\mathrm{NR}} = \max\big(A_X^{\dagger}, (A_X^{\dagger})^T\big). \tag{3.78}$$

Similarly, if we consider the quasi-inverse of the symmetrized matrix $\bar{A}_X := \max(A_X, A_X^T)$, expression (3.75) becomes

$$[\bar{A}_X^{\dagger}]_{ij} = \min_{P_{x_i x_j}} \max_{k | x_k \in P_{x_i x_j}} \bar{A}_X(x_k, x_{k+1}). \tag{3.79}$$

From comparing (3.79) and (3.2) it is immediate that

$$u_X^{\mathrm{R}} = \bar{A}_X^{\dagger} = \big(\max(A_X, A_X^T)\big)^{\dagger}. \tag{3.80}$$

If we show that $A_X^{\dagger} = A_X^{n-1}$, then (3.80) and (3.78) imply equations (3.76) and (3.77) respectively, completing the proof.

Notice that in $\mathfrak{A}$, the min or $\oplus$ operation is idempotent, i.e. $a \oplus a = a$ for all a. In this case, it can be shown that [32, Ch.4, Proposition 3.1.1]

$$I \oplus A_X \oplus A_X^2 \oplus ... \oplus A_X^k = (I \oplus A_X)^k, \tag{3.81}$$

for all $k \geq 1$. Recalling that I has zeros in the diagonal and $+\infty$ in the off-diagonal elements, it is immediate that $I \oplus A_X = A_X$. Consequently, (3.81) becomes

$$I \oplus A_X \oplus A_X^2 \oplus ... \oplus A_X^k = A_X^k. \tag{3.82}$$

Taking the limit to infinity in both sides of equality (3.82) and invoking the definition of the quasi-inverse in (3.74), we obtain

$$A_X^{\dagger} = \lim_{k \to \infty} A_X^k. \tag{3.83}$$

Finally, it can be shown [32, Ch. 4, Sec. 3.3, Theorem 1] that $A_X^{n-1} = A_X^n$, proving that the limit in (3.83) exists and, more importantly, that $A_X^\dagger = A_X^{n-1}$, as desired. ■

For the reciprocal ultrametric we symmetrize dissimilarities with a maximization operation and take the $(n-1)$-th power of the resulting matrix on the dioid algebra $\mathfrak{A}$. For the nonreciprocal ultrametric we revert the order of these two operations. We first consider matrix powers A_X^{n-1} and $\left(A_X^T\right)^{n-1}$ of the dissimilarity matrix and its transpose which we then symmetrize with a maximization operator. Besides emphasizing the extreme nature (cf. Theorem 4) of reciprocal and nonreciprocal clustering, Theorem 8 suggests the existence of intermediate methods in which we raise dissimilarity matrices A_X and A_X^T to some power, perform a symmetrization, and then continue applying matrix powers. These procedures yield methods that are not only valid but coincide with the family of semi-reciprocal ultrametrics introduced in Section 3.3.3, as the following proposition asserts.

Proposition 8 *For any network $N = (X, A_X)$ with n nodes the t-th semi-reciprocal ultrametric $u_X^{\mathrm{SR}(t)}$ in (3.52) for every natural $t \geq 2$ can be computed as*

$$u_X^{\mathrm{SR}(t)} = \left(\max\left(A_X^{t-1}, \left(A_X^T\right)^{t-1}\right)\right)^{n-1}, \tag{3.84}$$

where the matrix operations are in the dioid algebra $\mathfrak{A}$.

Proof: By comparison with (3.76), in (3.84) we in fact compute reciprocal clustering on the network (X, A_X^{t-1}). Furthermore, from the definition of matrix multiplication (3.71) in $\mathfrak{A}$, the $(t-1)$-th dioid power A_X^{t-1} is such that its entry $[A_X^{t-1}]_{ij}$ represents the minimum cost of a path containing at most t nodes, i.e.

$$[A_X^{t-1}]_{ij} = \min_{P_{x_i x_j}^t} \max_{k | x_k \in P_{x_i x_j}^t} A_X(x_k, x_{k+1}). \tag{3.85}$$

It is just a matter of notation, when comparing (3.85) and (3.51) to see that $A_X^{t-1} = A_X^{\mathrm{SR}(t)}$. Since semi-reciprocal clustering is equivalent to applying reciprocal clustering to network $(X, A_X^{\mathrm{SR}(t)})$ [cf. (3.52) and (3.2)], the proof concludes. ■

The result in (3.84) is intuitively clear. The powers A_X^{t-1} and $\left(A_X^T\right)^{t-1}$ represent the minimum cost among directed paths of at most $t-1$ links. In the terminology of Section 3.3.3 these are the costs of optimal secondary paths containing at most t nodes. Therefore, the maximization $\max\left(A_X^{t-1}, \left(A_X^T\right)^{t-1}\right)$ computes the cost of joining two nodes with secondary paths of at most t nodes in each direction. This is the definition of $\bar{A}_X^{\mathrm{SR}(t)}$ in (3.52). Applying the $(n-1)$-th dioid power to this new matrix is equivalent to looking for minimizing paths in the network with costs given by the secondary paths. Thus, the outermost dioid power computes the costs of the optimal main paths that achieve the ultrametric values in (3.52).

Observe that we recover (3.76) by making $t = 2$ in (3.84) and that we recover (3.77) when $t = n$. For this latter case note that when $t = n$ in (3.84), comparison with (3.77) shows that $\max(A_X^{t-1}, (A_X^T)^{t-1}) = \max(A_X^{n-1}, (A_X^T)^{n-1}) = u_X^{\mathrm{NR}}$. However, since u_X^{NR} is an ultrametric it is idempotent in the dioid algebra [cf. (3.73)] and the outermost dioid power in (3.84) is moot. This recovery is consistent with the observations in (3.59) and (3.60) that reciprocal and nonreciprocal clustering are particular cases of semi-reciprocal clustering $\mathcal{H}^{\mathrm{SR}(t)}$ such that for $t = 2$ we have $u_X^{\mathrm{SR}(2)} = u_X^{\mathrm{R}}$ and for $t \geq n$ it holds that $u_X^{\mathrm{SR}(t)} = u_X^{\mathrm{NR}}$. The results in Theorem 8 and Proposition 8 emphasize the extreme nature of the reciprocal and nonreciprocal methods and characterize the semi-reciprocal ultrametrics as natural intermediate clustering methods in an algorithmic sense.

This algorithmic perspective allows for a generalization in which the powers of the matrices A_X and A_X^T are different. To be precise consider positive integers $t, t' > 0$ and define the algorithmic intermediate clustering method $\mathcal{H}^{t,t'}$ with parameters t, t' as the one that maps the given network $N = (X, A_X)$ to the ultrametric set $(X, u_X^{t,t'}) = \mathcal{H}^{t,t'}(N)$ given by

$$u_X^{t,t'} := \left(\max\left(A_X^t, \left(A_X^T\right)^{t'}\right)\right)^{n-1}. \tag{3.86}$$

The ultrametric (3.86) can be interpreted as a semi-reciprocal ultrametric where the allowed length of secondary paths varies with the direction. Forward secondary paths may have at most $t + 1$ nodes whereas backward secondary paths may have at most $t' + 1$ nodes. The algorithmic intermediate family $\mathcal{H}^{t,t'}$ encapsulates the semi-reciprocal family since $\mathcal{H}^{t,t} \equiv \mathcal{H}^{\mathrm{SR}(t+1)}$ as well as the reciprocal method since $\mathcal{H}^{\mathrm{R}} \equiv \mathcal{H}^{1,1}$ as it follows from comparison of (3.86) with (3.84) and (3.76), respectively. We also have that $\mathcal{H}^{\mathrm{NR}}(N) = \mathcal{H}^{n-1,n-1}(N)$ for all networks $N = (X, A_X)$ such that $|X| \leq n$. This follows from the comparison of (3.86) with (3.77) and the idempotency of $u_X^{\mathrm{NR}} = \max(A_X^{n-1}, (A_X^T)^{n-1})$ with respect to the dioid algebra. The intermediate algorithmic methods $\mathcal{H}^{t,t'}$ are admissible as we claim in the following proposition.

Proposition 9 *The hierarchical clustering method $\mathcal{H}^{t,t'}$ is valid and admissible. I.e., $u_X^{t,t'}$ defined in* (3.86) *is a valid ultrametric and $\mathcal{H}^{t,t'}$ satisfies axioms (A1)-(A2).*

Proof: Since method $\mathcal{H}^{t,t'}$ is a generalization of $\mathcal{H}^{\mathrm{SR}(t)}$, the proof is almost identical to the one of Proposition 6. The only major difference is that showing symmetry of $u_X^{t,t'}$, i.e. $u_X^{t,t'}(x, x') = u_X^{t,t'}(x', x)$ for all $x, x' \in X$, is not immediate as in the case of $u_X^{\mathrm{SR}(t)}$. In a fashion similar to (3.52), we rewrite the definition of $u_X^{t,t'}$ given an arbitrary network (X, A_X) in terms of minimizing paths,

$$u_X^{t,t'}(x, x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} A_X^{t,t'}(x_i, x_{i+1}) \tag{3.87}$$

where the function $A_X^{t,t'}$ is defined as

$$A_X^{t,t'}(x,x') := \max\Big(A_X^{\mathrm{SR}(t+1)}(x,x'), A_X^{\mathrm{SR}(t'+1)}(x',x)\Big), \tag{3.88}$$

for all $x, x' \in X$ and functions $A_X^{\mathrm{SR}(\cdot)}$ as defined in (3.51). Notice that $A_X^{t,t'}$ is not symmetric in general. Symmetry of $u_X^{t,t'}$, however, follows from the following claim.

Claim 3 *Given any network (X, A_X) and a pair of nodes $x, x' \in X$ such that $u_X^{t,t'}(x,x') = \delta$, then $u_X^{t,t'}(x',x) \leq \delta$.*

Proof: Assuming $u_X^{t,t'}(x,x') = \delta$, we denote by $P_{xx'} = [x = x_0, x_1, ..., x_l = x']$ a minimizing main path achieving the cost δ in (3.87). Thus, we must show that there exists a main path $\hat{P}_{x'x}$ from x' back to x with cost not exceeding δ. From definition (3.88), there must exist secondary paths in both directions between every pair of consecutive nodes x_i, x_{i+1} in $P_{xx'}$ with cost no greater than δ. These secondary paths $P^{t+1}_{x_i x_{i+1}}$ and $P^{t'+1}_{x_{i+1} x_i}$ can have at most $t+1$ nodes in the forward direction and at most $t'+1$ nodes in the opposite direction. Moreover, without loss of generality we may consider the secondary paths as having exactly $t+1$ nodes in one direction and $t'+1$ in the other if we do not require consecutive nodes to be distinct.

Focus on a pair of consecutive nodes x_i, x_{i+1} of the main path $P_{xx'}$. If we can construct a main path from x_{i+1} back to x_i with cost not greater than δ, then we can concatenate these paths for pairs x_{i+1}, x_i for all i and obtain the required path $\hat{P}_{x'x}$ in the opposite direction.

Notice that the secondary paths $P^{t'+1}_{x_{i+1} x_i}$ and $P^{t+1}_{x_i x_{i+1}}$ can be concatenated to form a loop $P_{x_{i+1} x_{i+1}}$, i.e. a path starting and ending at the same node, of $t' + t + 1$ nodes and cost not larger than δ. We rename the nodes in $P_{x_{i+1} x_{i+1}} = [x_{i+1} = x^0, x^1, ..., x^{t'} = x_i, ..., x^{t'+t-1}, x^{t'+t} = x_{i+1}]$ starting at x_{i+1} and following the direction of the loop.

Now we are going to construct a main path $P_{x_{i+1} x_i}$ from x_{i+1} to x_i. We may reinterpret the loop $P_{x_{i+1} x_{i+1}}$ as the concatenation of two secondary paths $[x^0, x^1, \ldots, x^t]$ and $[x^t, x^{t+1}, \ldots, x^{t+t'} = x^0]$ each of them having cost not greater than δ. Thus, we may pick $x^0 = x_{i+1}$ and x^t as the first two nodes of the main path $P_{x_{i+1} x_i}$. With the same reasoning, we may link x^t with $x^{2t \bmod (t+t')}$ with cost not exceeding δ, and we may link $x^{2t \bmod (t+t')}$ with $x^{3t \bmod (t+t')}$ with cost not exceeding δ, and so on. Hence, we construct the main path

$$P_{x_{i+1} x_i} = [x^0, x^t, x^{2t \mod (t+t')}, \ldots, x^{(t+t'-1)t \mod (t+t')}], \tag{3.89}$$

which, by construction, has cost not exceeding δ. In order to finish the proof, we need to verify that the last node in the path in (3.89) is in fact $x^{t'} = x_i$. To do so, we have to show that $(t + t' - 1)\, t \equiv t' \mod (t + t')$, which follows from rewriting the left-hand side as

$(t+t')(t-1)+t'$. ■

Applying Claim 3 to an arbitrary pair of nodes x, x' and then to the pair x', x implies that $u_X^{t,t'}(x,x') = u_X^{t,t'}(x',x)$, as needed to show Proposition 9. ■

Algorithms for the implementation of the rest of the intermediate clustering methods can be derived. Computing ultrametrics associated with the grafting families in Section 3.3.1 entail simple combinations of matrices u_X^{R} and u_X^{NR}. E.g., the ultrametrics in (3.35) corresponding to the grafting method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ can be computed as

$$u_X^{\mathrm{R/NR}}(\beta) = u_X^{\mathrm{NR}} \circ \mathbb{I}\left\{u_X^{\mathrm{R}} \leq \beta\right\} + u_X^{\mathrm{R}} \circ \mathbb{I}\left\{u_X^{\mathrm{R}} > \beta\right\}, \tag{3.90}$$

where $\circ$ denotes the Hadamard matrix product and $\mathbb{I}\{\cdot\}$ is an element-wise indicator function.

In symmetric networks, Theorem 4 states that any admissible method must output an ultrametric equal to the single linkage ultrametric, that we can denote by u_X^{SL}. Thus, all algorithms in this section yield the same output u_X^{SL} when restricted to symmetric matrices A_X. Considering, e.g., the algorithm for the reciprocal ultrametric in (3.76) and noting that for a symmetric network $A_X = \max(A_X, A_X^T)$ we conclude that single linkage can be computed as

$$u_X^{\mathrm{SL}} = A_X^{n-1}. \tag{3.91}$$

Algorithms for the convex combination family in Section 3.3.2 involve computing dioid algebra powers of a convex combination of ultrametric matrices. Given two admissible methods $\mathcal{H}^1$ and $\mathcal{H}^2$ with outputs $(X, u_X^1) = \mathcal{H}^1(N)$ and $(X, u_X^2) = \mathcal{H}^2(N)$, and $\theta \in [0,1]$, the ultrametric in (3.46) corresponding to the method $\mathcal{H}_\theta^{12}$ can be computed as

$$u_X^{12}(\theta) = \Big(\theta\, u_X^1 + (1-\theta)\, u_X^2\Big)^{n-1}. \tag{3.92}$$

The operation $\theta\, u_X^1 + (1-\theta)\, u_X^2$ is just the regular convex combination in (3.45) and the dioid power in (3.92) implements the single linkage operation in (3.46) as it follows from (3.91).

Finally, regarding unilateral clustering, by combining (3.62) and (3.91) we obtain an algorithmic way of computing the unilateral ultrametric output for any network as

$$u_X^{\mathrm{U}} = \Big(\min\left(A_X, A_X^T\right)\Big)^{n-1}. \tag{3.93}$$

Remark 4 It follows from (3.76), (3.77), (3.84), (3.86), (3.90), (3.92), and (3.93) that all methods presented in this section can be computed in a number of operations of order $O(n^4)$ which coincides with the time it takes to compute n matrix products of matrices of size $n \times n$. This complexity can be reduced to $O(n^3 \log n)$ by noting that the dioid matrix power A^n can be computed via the sequence $A, A^2, A^4, \ldots$ which requires $O(\log n)$ matrix products at a cost of $O(n^3)$ each. Complexity can be further reduced using the sub cubic dioid matrix multiplication algorithms in [24, 86] that have complexity $O(n^{2.688})$ for a total complexity of $O(n^{2.688} \log n)$ to compute the n-th matrix power. There are also related methods with even lower complexities. For the case of reciprocal clustering, complexity of order $O(n^2)$ can be achieved by leveraging an equivalence between single linkage and a minimum spanning tree problem [39, 59]. For the case of nonreciprocal clustering, Tarjan's method [79] can be implemented to reduce complexity to $O(n^2 \log n)$.

Remark 5 The relationship between dioid algebras and the structured representation of networks transcends the algorithmic realm discussed in this section. By working at a higher level of algebraic abstraction, in Chapter 11 we show some of the results presented in this first part of the thesis but for arbitrary dioid algebras. In this way, when specializing the dioid algebra to $\mathfrak{A} := (\bar{\mathbb{R}}_+, \min, \max)$ the hierarchical clustering results are recovered. Moreover, when particularizing the results to other algebras, seemingly unrelated domains can be studied under a unified framework. For more details refer to Chapter 11.

Chapter 4

Quasi-clustering

A partition $P = \{B_1, \ldots, B_J\}$ of a set X represents a clustering of X into blocks or groups of nodes $B_1, \ldots, B_J \in P$ that can influence each other more than they can influence or be influenced by the rest. The partition can be interpreted as a reduction in data complexity in which variations between elements of a group are neglected in favor of the larger dissimilarities between elements of different groups. This is natural when clustering datasets endowed with symmetric dissimilarities because the concepts of a node $x \in X$ being close to another node $x' \in X$ and x' being close to x are equivalent. In an asymmetric network these concepts are different and this difference motivates the definition of structures more general than partitions.

Recalling that a partition $P = \{B_1, \ldots, B_J\}$ of X induces and is induced by an equivalence relation $\sim_P$ on X we search for the analogous of an asymmetric partition by removing the symmetry property in the definition of the equivalence relation. Thus, we define a *quasi-equivalence* $\rightsquigarrow$ as a binary relation that satisfies the reflexivity and transitivity properties but is not necessarily symmetric as stated next.

Definition 2 *A binary relation $\rightsquigarrow$ between elements of a set X is a quasi-equivalence if and only if the following properties hold true for all $x, x', x'' \in X$:*

(i) Reflexivity. Points are quasi-equivalent to themselves, $x \rightsquigarrow x$.

(ii) Transitivity. If $x \rightsquigarrow x'$ and $x' \rightsquigarrow x''$ then $x \rightsquigarrow x''$.

Quasi-equivalence relations are more often termed preorders or quasi-orders in the literature [37]. We choose the term quasi-equivalence to emphasize that they are a modified version of an equivalence relation.

We further define a *quasi-partition* of the set X as a directed unweighted graph $\tilde{P} = (P, E)$ where the vertex set P is a partition $P = \{B_1, \ldots, B_J\}$ of the space X and the edge set $E \subset P \times P$ is such that the following properties are satisfied (see Fig. 4.1):

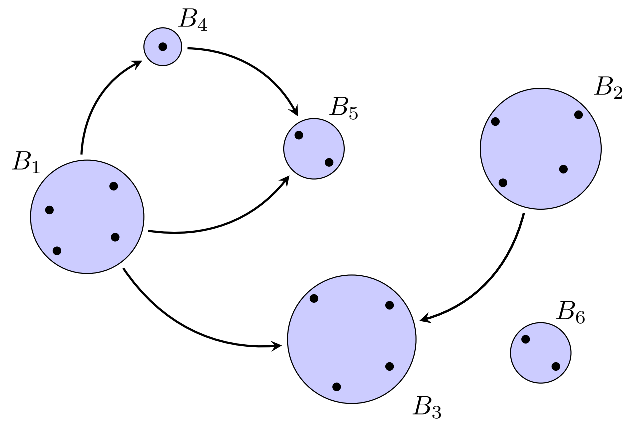


Figure 4.1: Example of a quasi-partition $\tilde{P} = (P, E)$. The vertex set P of the quasi-partition is given by a partition of the nodes $P = \{B_1, B_2, \ldots, B_6\}$. Nodes within the same block of the partition P can influence each other. The edges of the directed graph $\tilde{P} = (P, E)$ represent unidirectional influence between the blocks of the partition. In this case, block B_1 can influence B_3, B_4 and B_5 while block B_2 and B_4 can only influence B_3 and B_5, respectively.

(QP1) Unidirectionality. For any given pair of distinct blocks B_i and $B_j \in P$ we have, at most, one edge between them. Thus, if for some $i \neq j$ we have $(B_i, B_j) \in E$ then $(B_j, B_i) \notin E$.

(QP2) Transitivity. If there are edges between blocks B_i and B_j and between blocks B_j and B_k, then there is an edge between blocks B_i and B_k.

The vertex set P of a quasi-partition $\tilde{P} = (P, E)$ is meant to capture sets of nodes that can influence each other, whereas the edges in E intend to capture the notion of directed influence from one group to the next. In the example in Fig. 4.1, nodes which are drawn close to each other have low dissimilarities between them in both directions. Thus, the nodes inside each block B_i are close to each other but dissimilarities between nodes of different blocks are large in at least one direction. E.g., the dissimilarity from B_1 to B_4 is small but the dissimilarity from B_4 to B_1 is large. This latter fact motivates keeping B_1 and B_4 as separate blocks in the partition whereas the former motivates the addition of the directed influence edge (B_1, B_4). Likewise, dissimilarities from B_1 to B_3, from B_2 to B_3 and from B_4 to B_5 are small whereas those on opposite directions are not. Dissimilarities from the nodes in B_1 to the nodes in B_5 need not be small, but B_1 can influence B_5 through B_4, hence the edge from B_1 to B_5, in accordance with (QP2). All other dissimilarities are large justifying the lack of connections between the other blocks. Further observe that there are no bidirectional edges as required by (QP1).

Requirements (QP1) and (QP2) in the definition of quasi-partition represent the relational structure that emerges from quasi-equivalence relations as we state in the following proposition.

Proposition 10 *Given a node set X and a quasi-equivalence relation $\rightsquigarrow$ on X [cf. Definition 2] define the relation $\leftrightarrow$ on X as*

$$x \leftrightarrow x' \quad \iff \quad x \rightsquigarrow x' \text{ and } x' \rightsquigarrow x, \tag{4.1}$$

for all $x, x' \in X$. Then, $\leftrightarrow$ is an equivalence relation. Let $P = \{B_1, \dots, B_J\}$ be the partition of X induced by $\leftrightarrow$. Define $E \subseteq P \times P$ such that for all distinct $B_i, B_j \in P$

$$(B_i, B_j) \in E \quad \iff \quad x_i \rightsquigarrow x_j, \tag{4.2}$$

for some $x_i \in B_i$ and $x_j \in B_j$. Then, $\tilde{P} = (P, E)$ is a quasi-partition of X. Conversely, given a quasi-partition $\tilde{P} = (P, E)$ of X, define the binary relation $\rightsquigarrow$ on X so that for all $x, x' \in X$

$$x \rightsquigarrow x' \iff [x] = [x'] \;\; or \;\; ([x], [x']) \in E, \tag{4.3}$$

where $[x] \in P$ is the block of the partition P that contains the node x and similarly for $[x']$. Then, $\rightsquigarrow$ is a quasi-equivalence on X.

Proof: See Theorem 4.9, Ch. 1.4 in [37]. ■

In the same way that an equivalence relation induces and is induced by a partition on a given node set X, Proposition 10 shows that a quasi-equivalence relation induces and is induced by a quasi-partition on X. We can then adopt the construction of quasi-partitions as the natural generalization of clustering problems when given asymmetric data. Further, observe that if the edge set E contains no edges, $\tilde{P} = (P, E)$ is equivalent to the regular partition P when ignoring the empty edge set. In this sense, partitions are particular cases of quasi-partitions having the generic form $\tilde{P} = (P, \emptyset)$. To allow generalizations of hierarchical clustering methods with asymmetric outputs we introduce the notion of *quasi-dendrogram* in the following section.

4.1 Quasi-dendrograms and quasi-ultrametrics

Given that a dendrogram is defined as a nested set of partitions, we define a *quasi-dendrogram* $\tilde{D}_X$ of the set X as a collection of nested quasi-partitions $\tilde{D}_X(\delta) = (D_X(\delta), E_X(\delta))$ indexed by a resolution parameter $\delta \geq 0$. Recall the definition of $[x]_\delta$ from Section 2.1. Formally, for $\tilde{D}_X$ to be a quasi-dendrogram we require the following conditions:

($\tilde{D}$1) Boundary conditions. At resolution $\delta = 0$ all nodes are in separate clusters with no influences between them and for some δ_0 sufficiently large all elements of X are in a

single cluster,

$$\tilde{D}_X(0) = \Big(\big\{\{x\},\, x \in X\big\},\ \emptyset\Big), \qquad \tilde{D}_X(\delta_0) = \Big(\{X\}, \emptyset\Big) \quad \text{for some } \delta_0 > 0. \tag{4.4}$$

($\tilde{D}2$) Equivalence hierarchy. For any pair of points x, x' for which $x \sim_{D_X(\delta_1)} x'$ at resolution δ_1 we must have $x \sim_{D_X(\delta_2)} x'$ for all resolutions $\delta_2 > \delta_1$.

($\tilde{D}3$) Influence hierarchy. If there is an influence edge $([x]_{\delta_1}, [x']_{\delta_1}) \in E_X(\delta_1)$ between the equivalence classes $[x]_{\delta_1}$ and $[x']_{\delta_1}$ of nodes x and x' at resolution δ_1, at any resolution $\delta_2 > \delta_1$ we either have $([x]_{\delta_2}, [x']_{\delta_2}) \in E_X(\delta_2)$ or $[x]_{\delta_2} = [x']_{\delta_2}$.

($\tilde{D}4$) Right continuity. For all $\delta \geq 0$ there exists $\epsilon > 0$ such that $\tilde{D}_X(\delta) = \tilde{D}_X(\delta')$ for all $\delta' \in [\delta, \delta + \epsilon]$.

Requirements ($\tilde{\text{D}}$1), ($\tilde{\text{D}}$2), and ($\tilde{\text{D}}$4) are counterparts to the requirements (D1), (D2), and (D3) in the definition of dendrograms. The minor variation in ($\tilde{\text{D}}$1) is to specify that the edge sets at the boundary conditions are empty. For $\delta = 0$ this is because there are no influences at that resolution and for $\delta = \delta_0$ because there is a single cluster and we declared that blocks do not have self-loops. Condition ($\tilde{\text{D}}$3) states for the edge set the analogous requirement that condition (D2), or ($\tilde{\text{D}}$2) for that matter, states for the node set. If there is an edge present at a given resolution δ_1 that edge should persist at coarser resolutions $\delta_2 > \delta_1$ except if the groups linked by the edge merge in a single cluster.

Respective comparison of ($\tilde{\text{D}}$1), ($\tilde{\text{D}}$2), and ($\tilde{\text{D}}$4) to properties (D1), (D2), and (D3) in Section 2.1 implies that given a quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$ on a node set X, the component D_X is a dendrogram on X. I.e, the vertex sets $D_X(\delta)$ of the quasi-partitions $(D_X(\delta), E_X(\delta))$ for varying δ form a nested set of partitions. Hence, if the edge set $E_X(\delta) = \emptyset$ for every resolution parameter $\delta \geq 0$, $\tilde{D}_X$ recovers the structure of the dendrogram D_X. Thus, quasi-dendrograms are a generalization of dendrograms, or, equivalently, dendrograms are particular cases of quasi-dendrograms with empty edge sets. Redefining dendrograms D_X so that they represent quasi-dendrograms $(D_X, \emptyset)$ with empty edge sets and reinterpreting $\mathcal{D}$ as the set of quasi-dendrograms with empty edge sets we have that $\mathcal{D} \subset \tilde{\mathcal{D}}$, where $\tilde{\mathcal{D}}$ is the space of quasi-dendrograms.

A hierarchical clustering method $\mathcal{H} : \tilde{\mathcal{N}} \to \mathcal{D}$ is defined as a map from the space of networks $\tilde{\mathcal{N}}$ to the space of dendrograms $\mathcal{D}$ [cf. (2.3)]. Likewise, we define a hierarchical *quasi-clustering* method as a map

$$\tilde{\mathcal{H}} : \tilde{\mathcal{N}} \to \tilde{\mathcal{D}}, \tag{4.5}$$

from the space of networks to the space of quasi-dendrograms such that the underlying space X is preserved. Since $\mathcal{D} \subset \tilde{\mathcal{D}}$ we have that every clustering method is a quasi-clustering method but not vice versa. Our goal here is to study quasi-clustering methods satisfying suitably modified versions of the Axioms of Value and Transformation introduced in Section 2.2. Before that, we introduce quasi-ultrametrics as asymmetric versions of ultrametrics and show their equivalence to quasi-dendrograms after two pertinent remarks.

Remark 6 If we are given a quasi-equivalence relation and its induced quasi-partition on a node set X, (4.1) implies that all nodes inside the same block of the quasi-partition are quasi-equivalent to each other. If we combine this with the transitivity property in Definition 2, we have that if $x_i \rightsquigarrow x_j$ for some $x_i \in B_i$ and $x_j \in B_j$ or, equivalently, $(B_i, B_j) \in E$ then $x'_i \rightsquigarrow x'_j$ for all $x'_i \in B_i$ and all $x'_j \in B_j$.

Remark 7 Unidirectionality (QP1) ensures that no cycles containing exactly two nodes can exist in any quasi-partition $\tilde{P} = (P, E)$. If there were longer cycles, transitivity (QP2) would imply that every two distinct nodes in a longer cycle would have to form a two-node cycle, contradicting (QP1). Thus, conditions (QP1) and (QP2) imply that every quasi-partition $\tilde{P} = (P, E)$ is directed acyclic graph (DAG). The fact that a DAG represents a partial order shows that our construction of a quasi-partition from a quasi-equivalence relation is consistent with the known set theoretic construction of a partial order on a partition of a set given a preorder on the set [37].

Given a node set X a *quasi-ultrametric* $\tilde{u}_X$ on X is a function $\tilde{u}_X : X \times X \to \mathbb{R}_+$ satisfying the identity property and the strong triangle inequality in (2.12) as we formally define next.

Definition 3 *Given a node set X a quasi-ultrametric $\tilde{u}_X$ is a nonnegative function $\tilde{u}_X$: $X \times X \to \mathbb{R}_+$ satisfying the following properties for all $x, x', x'' \in X$::*

(i) Identity. $\tilde{u}_X(x, x') = 0$ if and only if $x = x'$.

(ii) Strong triangle inequality. $\tilde{u}_X$ satisfies (2.12).

Comparison of Definitions 1 and 3 shows that quasi-ultrametrics may be regarded as ultrametrics where the symmetry property is not imposed. In particular, the space $\tilde{\mathcal{U}}$ of quasi-ultrametric networks, i.e. networks with quasi-ultrametrics as dissimilarity functions, is a superset of the space of ultrametric networks $\mathcal{U} \subset \tilde{\mathcal{U}}$. See [33] for a study of some structural properties of quasi-ultrametrics.

Similar to the claim in Theorem 1 that provides a structure preserving bijection between dendrograms and ultrametrics, the following constructions and theorem establish a structure preserving equivalence between quasi-dendrograms and quasi-ultrametrics.

Consider the map $\Psi : \tilde{\mathcal{D}} \to \tilde{\mathcal{U}}$ defined as follows: for a given quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$ over the set X write $\Psi(\tilde{D}_X) = (X, \tilde{u}_X)$, where we define $\tilde{u}_X(x, x')$ for each $x, x' \in X$ as the smallest resolution δ at which either both nodes belong to the same equivalence class $[x]_\delta = [x']_\delta$, i.e. $x \sim_{D_X(\delta)} x'$, or there exists an edge in $E_X(\delta)$ from the equivalence class $[x]_\delta$ to the equivalence class $[x']_\delta$,

$$\tilde{u}_X(x,x') := \min\Big\{\delta \geq 0 \,\Big|\, [x]_\delta = [x']_\delta \quad \text{or} \quad ([x]_\delta, [x']_\delta) \in E_X(\delta)\Big\}. \tag{4.6}$$

We also consider the map $\Upsilon : \tilde{\mathcal{U}} \to \tilde{\mathcal{D}}$ constructed as follows: for a given quasi-ultrametric $\tilde{u}_X$ on the set X and each $\delta \geq 0$ define the relation $\sim_{\tilde{u}_X(\delta)}$ on X as

$$x \sim_{\tilde{u}_X(\delta)} x' \iff \max\big(\tilde{u}_X(x,x'), \tilde{u}_X(x',x)\big) \leq \delta. \tag{4.7}$$

Define further $D_X(\delta) := \big\{X \mod \sim_{\tilde{u}_X(\delta)}\big\}$ and the edge set $E_X(\delta)$ for every $\delta \geq 0$ as follows: $B_1 \neq B_2 \in D_X(\delta)$ are such that

$$(B_1, B_2) \in E_X(\delta) \iff \min_{\substack{x_1 \in B_1 \\ x_2 \in B_2}} \tilde{u}_X(x_1, x_2) \leq \delta. \tag{4.8}$$

Finally, $\Upsilon(X, \tilde{u}_X) := \tilde{D}_X$, where $\tilde{D}_X := (D_X, E_X)$.

Theorem 9 *The maps $\Psi : \tilde{\mathcal{D}} \to \tilde{\mathcal{U}}$ and $\Upsilon : \tilde{\mathcal{U}} \to \tilde{\mathcal{D}}$ are both well-defined. Furthermore, $\Psi \circ \Upsilon$ is the identity on $\tilde{\mathcal{U}}$ and $\Upsilon \circ \Psi$ is the identity on $\tilde{\mathcal{D}}$.*

Proof: In order to show that Ψ is a well-defined map, we must show that $\Psi(\tilde{D}_X)$ is a quasi-ultrametric network for every quasi-dendrogram $\tilde{D}_X$. Given an arbitrary quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$, for a particular $\delta' \geq 0$ consider the quasi-partition $\tilde{D}_X(\delta')$. Consider the range of resolutions δ associated with such quasi-partition. I.e.,

$$\{\delta \geq 0 \,|\, \tilde{D}_X(\delta) = \tilde{D}_X(\delta')\}. \tag{4.9}$$

Right continuity ($\tilde{\text{D}}$4) of $\tilde{D}_X$ ensures that the minimum of the set in (4.9) is well-defined and hence definition (4.6) is valid. To prove that $\tilde{u}_X$ in (4.6) is a quasi-ultrametric we need to show that it attains non-negative values as well as the identity and strong triangle inequality properties. That $\tilde{u}_X$ attains non-negative values is clear from the definition (4.6). The identity property is implied by the first boundary condition in ($\tilde{\text{D}}$1). Since $[x]_0 = [x]_0$ for all $x \in X$, we must have $\tilde{u}_X(x,x) = 0$. Conversely, since for all $x \neq x' \in X$, $([x]_0, [x']_0) \notin E_X(0)$ and $[x]_0 \neq [x']_0$ we must have that $\tilde{u}_X(x,x') > 0$ for $x \neq x'$ and the identity property is satisfied. To see that $\tilde{u}_X$ satisfies the strong triangle inequality in (2.12), consider nodes x, x', and x'' such that the lowest resolution for which $[x]_\delta = [x'']_\delta$ or $([x]_\delta, [x'']_\delta) \in E_X(\delta)$ is δ_1 and the lowest resolution for which $[x'']_\delta = [x']_\delta$ or $([x'']_\delta, [x']_\delta) \in E_X(\delta)$ is δ_2. Right continuity ($\tilde{\text{D}}$4) ensures that these lowest resolutions are well-defined. According to (4.6) we then have

$$\tilde{u}_X(x, x'') = \delta_1, \qquad \tilde{u}_X(x'', x') = \delta_2. \tag{4.10}$$

Denote by $\delta_0 := \max(\delta_1, \delta_2)$. From the equivalence hierarchy ($\tilde{\text{D}}2$) and influence hierarchy ($\tilde{\text{D}}3$) properties, it follows that $[x]_{\delta_0} = [x'']_{\delta_0}$ or $([x]_{\delta_0}, [x'']_{\delta_0}) \in E_X(\delta_0)$ and $[x'']_{\delta_0} = [x']_{\delta_0}$ or $([x'']_{\delta_0}, [x']_{\delta_0}) \in E_X(\delta_0)$. Furthermore, from transitivity (QP2) of the quasi-partition $\tilde{D}_X(\delta_0)$, it follows that $[x]_{\delta_0} = [x']_{\delta_0}$ or $([x]_{\delta_0}, [x']_{\delta_0}) \in E_X(\delta_0)$. Using the definition in (4.6) for x, x' we conclude that

$$\tilde{u}_X(x, x') \leq \delta_0. \tag{4.11}$$

By definition $\delta_0 := \max(\delta_1, \delta_2)$, hence we substitute this expression in (4.11) and compare with (4.10) to obtain

$$\tilde{u}_X(x, x') \leq \max(\delta_1, \delta_2) = \max\Big(\tilde{u}_X(x, x''), \tilde{u}_X(x'', x')\Big). \tag{4.12}$$

Consequently, $\tilde{u}_X$ satisfies the strong triangle inequality and is therefore a quasi-ultrametric, proving that the map Ψ is well-defined.

For the converse result, we need to show that Υ is a well-defined map. Given a quasi-ultrametric $\tilde{u}_X$ on a node set X and a resolution $\delta \geq 0$, we first define the relation

$$x \rightsquigarrow_{\tilde{u}_X(\delta)} x' \quad \iff \quad \tilde{u}_X(x, x') \leq \delta, \tag{4.13}$$

for all $x, x' \in X$. Notice that $\rightsquigarrow_{\tilde{u}_X(\delta)}$ is a quasi-equivalence relation as defined in Definition 2 for all $\delta \geq 0$. The reflexivity property is implied by the identity property of the quasi-ultrametric $\tilde{u}_X$ and transitivity is implied by the fact that $\tilde{u}_X$ satisfies the strong triangle inequality. Furthermore, definitions (4.7) and (4.8) are just reformulations of (4.1) and (4.2) respectively, for the special case of the quasi-equivalence defined in (4.13). Hence, Proposition 10 guarantees that $\Upsilon(X, \tilde{u}_X) = \tilde{D}_X(\delta) = (D_X(\delta), E_X(\delta))$ is a quasi-partition for every resolution $\delta \geq 0$. In order to show that Υ is well-defined, we need to show that these quasi-partitions are nested, i.e. that $\tilde{D}_X$ satisfies ($\tilde{\text{D}}1$)-($\tilde{\text{D}}4$).

The first boundary condition in ($\tilde{\text{D}}1$) is implied by (4.7) and the identity property of $\tilde{u}_X$. The second boundary condition in ($\tilde{\text{D}}1$) is implied by the fact that $\tilde{u}_X$ takes finite real values on a finite domain since the node set X is finite. Hence, any δ_0 satisfying

$$\delta_0 \geq \max_{x, x' \in X} \tilde{u}_X(x, x'), \tag{4.14}$$

is a valid candidate to show fulfillment of ($\tilde{\text{D}}1$).

To see that $\tilde{D}_X$ satisfies ($\tilde{\text{D}}2$) assume that for a resolution δ_1 we have two nodes $x, x' \in X$

such that $x \sim_{\tilde{u}_X(\delta_1)} x'$ as in (4.7), then it follows that $\max\big(\tilde{u}_X(x,x'), \tilde{u}_X(x',x)\big) \leq \delta_1$. Thus, if we pick any $\delta_2 > \delta_1$ it is immediate that $\max\big(\tilde{u}_X(x,x'), \tilde{u}_X(x',x)\big) \leq \delta_2$ which by (4.7) implies that $x \sim_{\tilde{u}_X(\delta_2)} x'$.

Fulfillment of ($\tilde{\text{D}}$3) can be shown in a similar way as fulfillment of ($\tilde{\text{D}}$2). Given a scalar $\delta_1 \geq 0$ and $x, x' \in X$, if $([x]_{\delta_1}, [x']_{\delta_1}) \in E_X(\delta_1)$ then by (4.8) we have that

$$\min_{x_1 \in [x]_{\delta_1}, x_2 \in [x']_{\delta_1}} \tilde{u}_X(x_1, x_2) \leq \delta_1. \tag{4.15}$$

From property ($\tilde{\text{D}}$2), we know that for all $x \in X$, $[x]_{\delta_1} \subset [x]_{\delta_2}$ for all $\delta_2 > \delta_1$. Hence, two things might happen. Either $\max(\tilde{u}_X(x,x'), \tilde{u}_X(x',x)) \leq \delta_2$ in which case $[x]_{\delta_2} = [x']_{\delta_2}$ or it might be that $[x]_{\delta_2} \neq [x']_{\delta_2}$ but

$$\min_{x_1 \in [x]_{\delta_2}, x_2 \in [x']_{\delta_2}} \tilde{u}_X(x_1, x_2) \leq \delta_1 < \delta_2, \tag{4.16}$$

which implies that $([x]_{\delta_2}, [x']_{\delta_2}) \in E_X(\delta_2)$. Consequently, ($\tilde{\text{D}}$3) is satisfied.

Finally, to see that $\tilde{D}_X$ satisfies the right continuity condition ($\tilde{\text{D}}$4), for each $\delta \geq 0$ such that $\tilde{D}_X(\delta) \neq (\{X\}, \emptyset)$ we may define $\epsilon(\delta)$ as any positive scalar satisfying

$$0 < \epsilon(\delta) < \min_{\substack{x,x' \in X \\ \text{s.t. } \tilde{u}_X(x,x') > \delta}} \tilde{u}_X(x,x') - \delta, \tag{4.17}$$

where the finiteness of X ensures that $\epsilon(\delta)$ is well-defined. Hence, (4.7) and (4.8) guarantee that $\tilde{D}_X(\delta) = \tilde{D}_X(\delta')$ for $\delta' \in [\delta, \delta + \epsilon(\delta)]$. For all other resolutions δ such that $\tilde{D}_X(\delta) = (\{X\}, \emptyset)$, right continuity is trivially satisfied since the quasi-dendrogram remains unchanged for increasing resolutions. Consequently, $\Upsilon(X, \tilde{u}_X)$ is a valid quasi-dendrogram for every quasi-ultrametric network $(X, \tilde{u}_X)$, proving that Υ is well-defined.

To conclude the proof, we need to show that $\Psi \circ \Upsilon$ and $\Upsilon \circ \Psi$ are the identities on $\tilde{\mathcal{U}}$ and $\tilde{\mathcal{D}}$, respectively. To see why the former is true, pick any quasi-ultrametric network $(X, \tilde{u}_X)$ and consider an arbitrary pair of nodes $x, x' \in X$ such that $\tilde{u}_X(x,x') = \delta_0$. Also, consider the ultrametric network $\Psi \circ \Upsilon(X, \tilde{u}_X) := (X, \tilde{u}_X^*)$. From (4.7) and (4.8), in the quasi-dendrogram $\Upsilon(X, \tilde{u}_X)$ there is no influence from x to x' for resolutions $\delta < \delta_0$ and at resolution $\delta = \delta_0$ either an edge appears from $[x]_{\delta_0}$ to $[x']_{\delta_0}$, or both nodes merge into one single cluster. In any case, when we apply Ψ to the resulting quasi-dendrogram, we obtain $\tilde{u}_X^*(x,x') = \delta_0$. Since $x, x' \in X$ were chosen arbitrarily, we have that $\tilde{u}_X = \tilde{u}_X^*$, showing that $\Psi \circ \Upsilon$ is the identity on $\tilde{\mathcal{U}}$. A similar argument can be used to show that $\Upsilon \circ \Psi$ is the identity on $\tilde{\mathcal{D}}$. ■

Theorem 9 implies that every quasi-dendrogram $\tilde{D}_X$ has an equivalent representation as a quasi-ultrametric network defined on the same underlying node set X. This result allows us to reinterpret hierarchical quasi-clustering methods [cf. (4.5)] as maps

$$\tilde{\mathcal{H}} : \tilde{\mathcal{N}} \to \tilde{\mathcal{U}}, \tag{4.18}$$

from the space of networks to the space of quasi-ultrametric networks. Apart from the theoretical importance of Theorem 9, this equivalence result is of practical importance since quasi-ultrametrics are mathematically more convenient to handle than quasi-dendrograms – in the same sense in which regular ultrametrics are easier to handle than regular dendrograms. Quasi-dendrograms are still preferable for data representation as we discuss in the numerical examples in Chapter 6.

Given a quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$, the value $\tilde{u}_X(x, x')$ of the associated quasi-ultrametric for $x, x' \in X$ is given by the minimum resolution δ at which x can influence x'. This may occur when x and x' belong to the same block of $D_X(\delta)$ or when they belong to different blocks $B, B' \in D_X(\delta)$, but there is an edge from the block containing x to the block containing x', i.e. $(B, B') \in E_X(\delta)$. Conversely, given a quasi-ultrametric network $(X, \tilde{u}_X)$, for a given resolution δ the graph $\tilde{D}_X(\delta)$ has as a vertex set the classes of nodes whose quasi-ultrametric is less than δ in both directions. Furthermore, $\tilde{D}_X(\delta)$ contains a directed edge between two distinct equivalence classes if the quasi-ultrametric from some node in the first class to some node in the second is not greater than δ.

In Fig. 4.2 we present an example of the equivalence between quasi-dendrograms and quasi-ultrametric networks stated by Theorem 9. At the top left of the figure, we present a quasi-ultrametric $\tilde{u}_X$ defined on a three-node set $X = \{x_1, x_2, x_3\}$. At the top right, we depict the dendrogram component D_X of the quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$ equivalent to $(X, \tilde{u}_X)$ as given by Theorem 9. At the bottom of the figure, we present graphs $\tilde{D}_X(\delta)$ for a range of resolutions $\delta \geq 0$.

To obtain $\tilde{D}_X$ from $\tilde{u}_X$, we first obtain the dendrogram component D_X by symmetrizing $\tilde{u}_X$ to the maximum [cf. (4.7)], nodes x_1 and x_2 merge at resolution 2 and x_3 merges with $\{x_1, x_2\}$ at resolution 3. To see how the edges in $\tilde{D}_X$ are obtained, at resolutions $0 \leq \delta < 1$, there are no edges since there is no quasi-ultrametric value between distinct nodes in this range [cf. (4.8)]. At resolution $\delta = 1$, we reach the first nonzero values of $\tilde{u}_X$ and hence the corresponding edges appear in $\tilde{D}_X(1)$. At resolution $\delta = 2$, nodes x_1 and x_2 merge and become the same vertex in graph $\tilde{D}_X(2)$. Finally, at resolution $\delta = 3$ all the nodes belong to the same equivalence class and hence $\tilde{D}_X(3)$ contains only one vertex. Conversely, to obtain $\tilde{u}_X$ from $\tilde{D}_X$ as depicted in the figure, note that at resolution $\delta = 1$ two edges $([x_1]_1, [x_2]_1)$ and $([x_3]_1, [x_2]_1)$ appear in $\tilde{D}_X(1)$, thus the corresponding values of the quasi-ultrametric are fixed to be $\tilde{u}_X(x_1, x_2) = \tilde{u}(x_3, x_2) = 1$. At resolution $\delta = 2$, when x_1 and x_2 merge into the same vertex in $\tilde{D}_X(2)$, an edge is generated from $[x_3]_2$ to $[x_1]_2$ the equivalence class of x_1 at resolution $\delta = 2$ which did not exist before, implying that $\tilde{u}_X(x_3, x_1) = 2$.

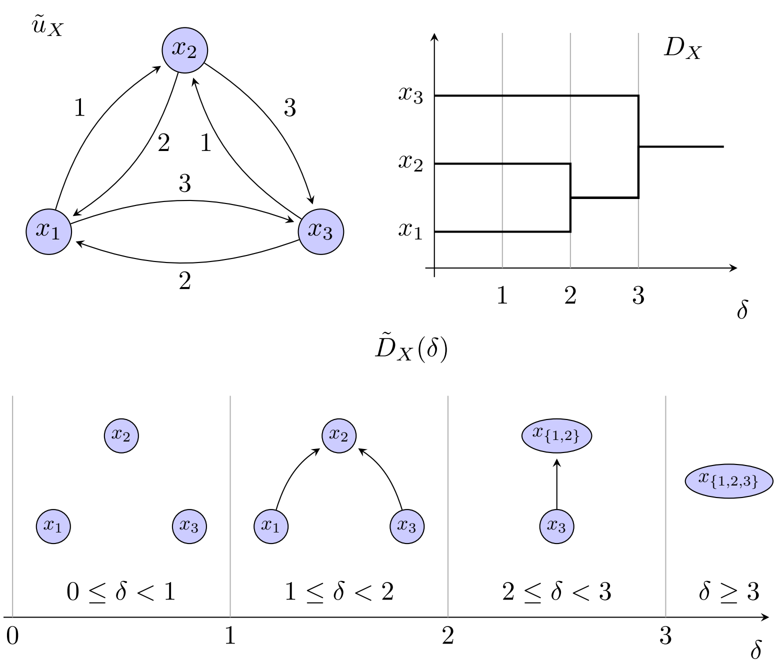


Figure 4.2: Equivalence between quasi-dendrograms and quasi-ultrametrics. A quasi-ultrametric $\tilde{u}_X$ is defined on three nodes $\{x_1, x_2, x_3\}$ and the equivalent quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$ is presented by depicting D_X and graphs $\tilde{D}_X(\delta)$ for every resolution δ.

Moreover, we have that $[x_2]_2 = [x_1]_2$, hence $\tilde{u}_X(x_2, x_1) = 2$. Finally, at $\tilde{D}_X(3)$ there is only one equivalence class, thus the values of $\tilde{u}_X$ that have not been defined so far must equal 3.

4.2 Admissible quasi-clustering methods

Mimicking the development in Section 2.2, we encode desirable properties of quasi-clustering methods into axioms which we use as a criterion for admissibility. The axioms considered are the directed versions of the Axioms of Value (A1) and Transformation (A2) introduced in Section 2.2. The Directed Axiom of Value ($\tilde{\text{A}}1$) and the Directed Axiom of Transformation ($\tilde{\text{A}}2$) winnow the space of quasi-clustering methods by imposing conditions on their output quasi-dendrograms which, given Theorem 9, can be more suitably expressed in terms of quasi-ultrametrics.

($\tilde{A}1$) Directed Axiom of Value. $\tilde{\mathcal{H}}(\vec{\Delta}_2(\alpha, \beta)) = \vec{\Delta}_2(\alpha, \beta)$ for every two-node network $\vec{\Delta}_2(\alpha, \beta)$.

($\tilde{A}2$) Directed Axiom of Transformation. Consider two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ and a dissimilarity reducing map $\phi : X \to Y$, i.e. a map ϕ such that for all $x, x' \in X$ it holds that $A_X(x, x') \geq A_Y(\phi(x), \phi(x'))$. Then, for all $x, x' \in X$, the outputs

$(X, \tilde{u}_X) = \tilde{\mathcal{H}}(X, A_X)$ and $(Y, \tilde{u}_Y) = \tilde{\mathcal{H}}(Y, A_Y)$ satisfy

$$\tilde{u}_X(x, x') \geq \tilde{u}_Y(\phi(x), \phi(x')). \tag{4.19}$$

The Directed Axiom of Transformation (Ã2) is just a restatement of the (regular) Axiom of Transformation (A2) where the ultrametrics u_X and u_Y in (2.14) are replaced by the quasi-ultrametrics $\tilde{u}_X$ and $\tilde{u}_Y$ in (4.19). The axioms are otherwise conceptual analogues. In terms of quasi-dendrograms, (Ã2) states that no influence relation can be weakened by a dissimilarity reducing transformation. The Directed Axiom of Value (Ã1) simply recognizes that in any two-node network, the dissimilarity function is itself a quasi-ultrametric and that there is no valid justification to output a different quasi-ultrametric. In this sense, (Ã1) is similar to the Symmetric Axiom of Value (B1) that also requires two-node networks to be fixed points of (symmetric) hierarchical clustering methods. In terms of quasi-dendrograms, (Ã1) requires the quasi-clustering method to output the quasi-dendrogram equivalent according to Theorem 9 to the dissimilarity function of the two-node network.

4.3 Directed single linkage

We call a quasi-clustering method $\tilde{\mathcal{H}}$ admissible if it satisfies axioms (Ã1) and (Ã2) and, emulating the development in Section 3.1, we want to find methods that are admissible with respect to these axioms. This is can be done in the following way. Recall the definition of the directed minimum path cost $\tilde{u}^*_X$ in (2.7) and define the *directed single linkage* quasi-clustering method $\tilde{\mathcal{H}}^*$ as the one with output quasi-ultrametrics $(X, \tilde{u}^*_X) = \tilde{\mathcal{H}}^*(X, A_X)$ given by the directed minimum path cost function $\tilde{u}^*_X$. The directed single linkage method $\tilde{\mathcal{H}}^*$ is valid and admissible as we show in the following proposition.

Proposition 11 *The hierarchical quasi-clustering method $\tilde{\mathcal{H}}^*$ is valid and admissible. I.e., $\tilde{u}^*_X$ defined by* (2.7) *is a quasi-ultrametric and $\tilde{\mathcal{H}}^*$ satisfies axioms (Ã1)-(Ã2).*

Proof: In order to show that $\tilde{u}^*_X$ is a valid quasi-ultrametric we may apply an argument based on concatenated paths as the one preceding Proposition 1.

To show fulfillment of axiom (Ã1), pick an arbitrary two-node network $\vec{\Delta}_2(\alpha, \beta)$ as defined in Section 2.1 and denote by $(\{p, q\}, \tilde{u}^*_{p,q}) = \tilde{\mathcal{H}}^*(\vec{\Delta}_2(\alpha, \beta))$. Then, we have $\tilde{u}^*_{p,q}(p, q) = \alpha$ and $\tilde{u}^*_{p,q}(q, p) = \beta$ because there is only one possible path selection in each direction [cf. (2.7)]. Satisfaction of the Directed Axiom of Transformation (Ã2) is the intermediate result (3.12) in the proof of Proposition 2. ■

From Proposition 11 we know that $\tilde{u}^*_X$ is a quasi-ultrametric. Its equivalent quasi-dendrogram according to Theorem 9 is related to the nonreciprocal clustering method $\mathcal{H}^{\mathrm{NR}}$ as we show next.

Proposition 12 *Let $\tilde{D}_X^* = (D_X^*, E_X^*)$ be the quasi-dendrogram equivalent to $\tilde{u}_X^*$ according to Theorem 9. Then, for every network $N = (X, A_X)$, $D_X^* = D_X^{\mathrm{NR}}$ where $D_X^{\mathrm{NR}} = \mathcal{H}^{\mathrm{NR}}(N)$ is the output dendrogram of applying nonreciprocal clustering as defined in* (3.8) *to N.*

Proof: Compare (3.8) with (4.7) and conclude that

$$x \sim_{\tilde{u}_X^*(\delta)} x' \iff u_X^{\mathrm{NR}}(x, x') \leq \delta, \tag{4.20}$$

for all $x, x' \in X$. The equivalence relation $\sim_{\tilde{u}_X^*(\delta)}$ defines D_X^* and by Theorem 1 we obtain that the equivalence relation $\sim_{u_X^{\mathrm{NR}}(\delta)}$ defining D_X^{NR} is such that

$$x \sim_{u_X^{\mathrm{NR}}(\delta)} x' \iff u_X^{\mathrm{NR}}(x, x') \leq \delta. \tag{4.21}$$

Comparing (4.20) and (4.21), the result follows. ■

Furthermore, from (2.15) and (3.91) it follows that for every network (X, A_X) with $|X| = n$, the quasi-ultrametric $\tilde{u}_X^*$ can be computed as

$$\tilde{u}_X^* = A_X^{n-1}, \tag{4.22}$$

where the operation $(\cdot)^{n-1}$ denotes the $(n-1)$st matrix power in the dioid algebra $\mathfrak{A} = (\bar{\mathbb{R}}_+, \min, \max)$ with matrix product as defined in (3.71).

Mimicking the developments in sections 3.1 and 3.2, we next ask which other methods satisfy $(\tilde{\mathrm{A}}1)$-$(\tilde{\mathrm{A}}2)$ and what special properties directed single linkage has. As it turns out, directed single linkage is the unique quasi-clustering method that is admissible with respect to $(\tilde{\mathrm{A}}1)$-$(\tilde{\mathrm{A}}2)$ as we assert in the following theorem.

Theorem 10 *Let $\tilde{\mathcal{H}}$ be a valid hierarchical quasi-clustering method satisfying axioms $(\tilde{A}1)$ and $(\tilde{A}2)$. Then, $\tilde{\mathcal{H}} \equiv \tilde{\mathcal{H}}^*$ where $\tilde{\mathcal{H}}^*$ is the directed single linkage method with output quasi-ultrametrics as in* (2.7)*.*

Proof: The proof is similar to the proof of Theorem 4. Given an arbitrary network $N = (X, A_X)$ denote by $(X, \tilde{u}_X) = \tilde{\mathcal{H}}(X, A_X)$ the output quasi-ultrametric resulting from the application of an arbitrary admissible quasi-clustering method $\tilde{\mathcal{H}}$. We will show that for all $x, x' \in X$

$$\tilde{u}_X^*(x, x') \leq \tilde{u}_X(x, x') \leq \tilde{u}_X^*(x, x'). \tag{4.23}$$

To prove the rightmost inequality in (4.23) we begin by showing that the dissimilarity function A_X acts as an upper bound on all admissible quasi-ultrametrics $\tilde{u}_X$, i.e.

$$\tilde{u}_X(x, x') \leq A_X(x, x'), \tag{4.24}$$

for all $x, x' \in X$. To see this, suppose $A_X(x, x') = \alpha$ and $A_X(x', x) = \beta$. Define the two-node network $N_{p,q} = (\{p, q\}, A_{p,q})$ where $A_{p,q}(p, q) = \alpha$ and $A_{p,q}(q, p) = \beta$ and denote by $(\{p, q\}, \tilde{u}_{p,q}) = \tilde{\mathcal{H}}(N_{p,q})$ the output of applying the method $\tilde{\mathcal{H}}$ to the network $N_{p,q}$. From axiom (Ã1), we have $\tilde{\mathcal{H}}(N_{p,q}) = N_{p,q}$, in particular

$$\tilde{u}_{p,q}(p, q) = A_{p,q}(p, q) = A_X(x, x'). \tag{4.25}$$

Moreover, notice that the map $\phi : \{p, q\} \to X$, where $\phi(p) = x$ and $\phi(q) = x'$ is a dissimilarity reducing map, i.e. it does not increase any dissimilarity, from $N_{p,q}$ to N. Hence, from axiom (Ã2), we must have

$$\tilde{u}_{p,q}(p, q) \geq \tilde{u}_X(\phi(p), \phi(q)) = \tilde{u}_X(x, x'). \tag{4.26}$$

Substituting (4.25) in (4.26), we obtain (4.24).

Consider now an arbitrary path $P_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ linking nodes x and x'. Since $\tilde{u}_X$ is a valid quasi-ultrametric, it satisfies the strong triangle inequality (2.12). Thus, we have that

$$\tilde{u}_X(x, x') \leq \max_{i | x_i \in P_{xx'}} \tilde{u}_X(x_i, x_{i+1}) \leq \max_{i | x_i \in P_{xx'}} A_X(x_i, x_{i+1}), \tag{4.27}$$

where the last inequality is implied by (4.24). Since by definition $P_{xx'}$ is an arbitrary path linking x to x', we can minimize (4.27) over all such paths maintaining the validity of the inequality,

$$\tilde{u}_X(x, x') \leq \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} A_X(x_i, x_{i+1}) = \tilde{u}^*_X(x, x'), \tag{4.28}$$

where the last equality is given by the definition of the directed minimum path cost (2.7). Thus, the rightmost inequality in (4.23) is proved.

To prove the leftmost inequality in (4.23), consider an arbitrary pair of nodes $x, x' \in X$ and fix $\delta = \tilde{u}^*_X(x, x')$. Then, by Lemma 1, there exists a partition $P_\delta(x, x') = \{B_\delta(x), B_\delta(x')\}$ of the node space X into blocks $B_\delta(x)$ and $B_\delta(x')$ with $x \in B_\delta(x)$ and $x' \in B_\delta(x')$ such that for all points $b \in B_\delta(x)$ and $b' \in B_\delta(x')$ we have

$$A_X(b, b') \geq \delta. \tag{4.29}$$

Focus on a two-node network $N_{u,v} = (\{u, v\}, A_{u,v})$ with $A_{u,v}(u, v) = \delta$ and $A_{u,v}(v, u) = s$ where $s = \operatorname{sep}(X, A_X)$ as defined in (2.10). Denote by $(\{u, v\}, \tilde{u}_{u,v}) = \tilde{\mathcal{H}}(N_{u,v})$ the output of applying the method $\tilde{\mathcal{H}}$ to the network $N_{u,v}$. Notice that the map $\phi : X \to \{u, v\}$ such that $\phi(b) = u$ for all $b \in B_\delta(x)$ and $\phi(b') = v$ for all $b' \in B_\delta(x')$ is dissimilarity reducing because, from (4.29), dissimilarities mapped to dissimilarities equal to δ in $N_{u,v}$ were originally larger. Moreover, dissimilarities mapped into s cannot have increased due to the definition of separation of a network (2.10). From axiom ($\tilde{\text{A}}$1),

$$\tilde{u}_{u,v}(u,v) = A_{u,v}(u,v) = \delta, \tag{4.30}$$

since $N_{u,v}$ is a two-node network. Moreover, since ϕ is dissimilarity reducing, from ($\tilde{\text{A}}$2) we may assert that

$$\tilde{u}_X(x,x') \geq \tilde{u}_{u,v}(\phi(x),\phi(x')) = \delta, \tag{4.31}$$

where we used (4.30) for the last equality. Recalling that $\tilde{u}^*_X(x,x') = \delta$ and substituting in (4.31) concludes the proof of the leftmost inequality in (4.23).

Since both inequalities in (4.23) hold, we must have $\tilde{u}^*_X(x,x') = \tilde{u}_X(x,x')$ for all $x, x' \in X$. Since this is true for any arbitrary network $N = (X, A_X)$, it follows that the admissible quasi-clustering method must be $\tilde{\mathcal{H}} \equiv \tilde{\mathcal{H}}^*$. ■

As it follows from Theorem 4 there are many different admissible hierarchical clustering algorithms for asymmetric networks. In the case of symmetric networks, Corollary 2 establishes that there is a unique admissible method. Theorem 10 shows that what prevents uniqueness in asymmetric networks is the insistence that the hierarchical clustering method should have a symmetric ultrametric output. If we remove the symmetry requirement there is also a unique admissible hierarchical quasi-clustering method. Furthermore, this unique method is an asymmetric version of single linkage. In that sense we can say that directed single linkage and quasi-ultrametrics are to asymmetric networks what single linkage and ultrametrics are to symmetric networks.

Remark 8 The definition of directed single linkage as a natural extension of single linkage hierarchical clustering to asymmetric networks dates back to [6]. Our contribution is to develop a framework to study hierarchical quasi-clustering that starts from quasi-equivalence relations, builds towards quasi-partitions and quasi-dendrograms, shows the equivalence of the latter to quasi-ultrametrics, and culminates with the proof that directed single linkage is the *unique* admissible method to hierarchically quasi-cluster asymmetric networks. Previously unknown stability and invariance properties of directed single linkage are further established in the ensuing chapter.

Chapter 5

Desirable properties of hierarchical clustering methods

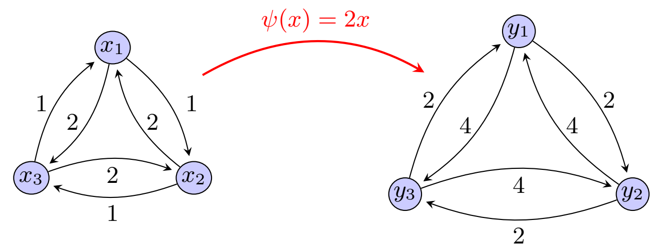
Other than reciprocal and nonreciprocal clustering, in Chapter 3 we introduced four intermediate families of admissible hierarchical clustering methods. These include the grafting methods, the convex combinations, the semi-reciprocal ultrametrics, and the algorithmic intermediate methods. In this chapter we explore additional restrictions to winnow the space of intermediate admissible methods. More specifically, we focus on four desirable features of hierarchical clustering methods: Scale Preservation (Section 5.1), Representability (Section 5.2), Excisiveness (Section 5.3), and Stability (Section 5.4). Furthermore, we show that these properties are not independent of each other and we analyze how they can be combined to obtain a complete characterization of desirable clustering methods.

5.1 Scale preservation

The first desirable property that we introduce is scale preservation, as formally defined next.

> *(P2) Scale Preservation.* Consider a network $N_X = (X, A_X)$ and a nondecreasing function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\psi(0) = 0$, $\psi(z) > 0$ for all $z > 0$ and $\lim_{z\to\infty} \psi(z) = \infty$. Define the network $N_Y = (Y, A_Y)$ with space $Y = X$ and dissimilarities $A_Y = \psi \circ A_X$. A hierarchical clustering method $\mathcal{H}$ is said to be scale preserving if for arbitrary network $N_X = (X, A_X)$ and arbitrary function ψ the outputs $(X, u_X) = \mathcal{H}(X, A_X)$ and $(Y, u_Y) = \mathcal{H}(Y, A_Y)$ satisfy $u_Y = \psi \circ u_X$.

Since ultrametric outcomes vary according to the same function that transforms the dissimilarity function, scale preserving methods are invariant with respect to units. In terms of dendrograms, scale preservation entails that a transformation of dissimilarities with appropriate function ψ results in a dendrogram where the order in which nodes are clustered



$$u_X^{\mathrm{NR}}(x,x') = 1 \xrightarrow{\;= 2\,u_X^{\mathrm{NR}}(x,x')\;} u_Y^{\mathrm{NR}}(y,y') = 2$$

$$u_X^{\mathrm{R}}(x,x') = 2 \xrightarrow{\;= 2\,u_X^{\mathrm{R}}(x,x')\;} u_Y^{\mathrm{R}}(y,y') = 4$$

$$u_X^{\mathrm{R/NR}}(x,x';\beta) = 1 \xrightarrow[\beta = 3]{\;\neq 2\,u_X^{\mathrm{R/NR}}(x,x';\beta)\;} u_Y^{\mathrm{R/NR}}(y,y';\beta) = 4$$

Figure 5.1: Scale preserving and non scale preserving methods. Reciprocal and nonreciprocal clustering are scale preserving. When dissimilarities for the network on the left are doubled, reciprocal and nonreciprocal ultrametrics also double. The grafting method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ [cf. (3.35)] is not scale preserving because in this particular example $u_Y^{R/NR}(y,y';\beta) = u_Y^{\mathrm{R}}(y,y') = 4 \neq 2\,u_X^{R/NR}(x,x';\beta) = 2$.

together is the same while the resolution at which clustering events happen changes according to ψ.

Although scale preservation is a stringent property, admissible methods satisfying (P2) exist. In particular, the following proposition claims that this is the case for the semi-reciprocal family of Section 3.3.3.

Proposition 13 *The semi-reciprocal clustering method $\mathcal{H}^{\mathrm{SR}(t)}$ with ultrametrics as in* (3.52) *satisfies the Property of Scale Preservation (P2) for all integers $t \geq 2$.*

Proof: To prove that the family of semi-reciprocal methods is scale preserving, define two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ with $X = Y$ and $A_Y = \psi \circ A_X$ where ψ is a nondecreasing function satisfying the scale preservation requirements in (P2). We use the fact that the optimal cost of secondary paths is transformed by ψ as we claim next.

Claim 4 *Given two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ with $X = Y$ and $A_Y = \psi \circ A_X$ where ψ is a nondecreasing function satisfying the scale preservation requirements in (P2), then*

$$A_Y^{\mathrm{SR}(t)} = \psi \circ A_X^{\mathrm{SR}(t)}, \tag{5.1}$$

where $A_Y^{\mathrm{SR}(t)}$ and $A_X^{\mathrm{SR}(t)}$ are defined as in (3.51)*.*

Proof: Denote by $P_{xx'}^{t*} = [x = x_0, ..., x_l = x']$ a minimizing secondary path such that we can rewrite (3.51) as

$$A_X^{\mathrm{SR}(t)}(x,x') = \max_{i|x_i \in P^{t*}_{xx'}} A_X(x_i, x_{i+1}) = \delta. \tag{5.2}$$

Consider the path $P^t_{yy'} = [y = y_0, ..., y_l = y']$ in the network N_Y where $y_i = x_i$ for all i. Note that $P^t_{yy'}$ is a particular path joining y and y' and need not be the minimizing one. Hence, we can assert that

$$A_Y^{\mathrm{SR}(t)}(y,y') \leq \max_{i|y_i \in P^t_{yy'}} A_Y(y_i, y_{i+1}) = \psi(\delta). \tag{5.3}$$

From (5.2) and (5.3), we show that $A_Y^{\mathrm{SR}(t)}(y,y') \leq \psi\big(A_X^{\mathrm{SR}(t)}(x,x')\big)$. In the remaining of the proof, we show that this inequality cannot be strict, implying the equality result we want to prove. Suppose $A_Y^{\mathrm{SR}(t)}(y,y') < \psi(\delta)$ for some minimizing path $P^{t*}_{yy'} = [y = y^0, ..., y^{l'} = y']$ such that we can write

$$\max_{i|y_i \in P^{t*}_{yy'}} A_Y(y_i, y_{i+1}) = A_Y(y_s, y_{s+1}) < \psi(\delta). \tag{5.4}$$

Consider the path $P^t_{xx'} = [x = x^0, ..., x^{l'} = x']$ in N_X with $x^i = y^i$ for all i. From (5.2) we can state that,

$$\delta \leq \max_{i|x_i \in P^t_{xx'}} A_X(x_i, x_{i+1}) = A_X(x_s, x_{s+1}), \tag{5.5}$$

where the last equality holds for some s because ψ is a nondecreasing function and, as a consequence, every maximizer in (5.5) must also be a maximizer in (5.4). From (5.4) and the definition $A_Y = \psi \circ A_X$ we can conclude

$$A_Y(y_s, y_{s+1}) = \psi(A_X(x_s, x_{s+1})) < \psi(\delta). \tag{5.6}$$

Notice that according to (5.5) we have $A_X(x_s, x_{s+1}) \geq \delta$ but according to (5.6) it holds that $\psi\big(A_X(x_s, x_{s+1})\big) < \psi(\delta)$. These inequalities are incompatible because ψ is nondecreasing. This contradiction indicates that the path $P^{t*}_{yy'}$ in (5.4) does not exist, implying

$$A_Y^{\mathrm{SR}(t)}(y,y') = \psi(\delta) = \psi\Big(A_X^{\mathrm{SR}(t)}(x,x')\Big), \tag{5.7}$$

proving the claim. ■

Consider one minimizing main path $P^{X*}_{xx'} = [x = x_0, \ldots, x_l = x']$ in definition (3.52) and focus on the path $P^Y_{yy'}$ in N_Y with $x_i = y_i$ for all i. Notice that this is a particular path joining y and y'. Hence, we can state,

$$u_Y^{\mathrm{SR}(t)}(y, y') \leq \max_{y_i \in P_{yy'}^Y} \bar{A}_Y^{\mathrm{SR}(t)}(y_i, y_{i+1}). \tag{5.8}$$

However, from Claim 4

$$u_Y^{\mathrm{SR}(t)}(y, y') \leq \max_{i | x_i \in P_{xx'}^{X*}} \psi\big(\bar{A}_X^{\mathrm{SR}(t)}(x_i, x_{i+1})\big) = \psi\Big(u_X^{\mathrm{SR}(t)}(x, x')\Big), \tag{5.9}$$

where the last equality is enforced because ψ is nondecreasing. To conclude the proof, we can use a contradiction argument analogous to the one used in Claim 4 to show that the inequality in (5.9) cannot be strict, proving that the semi-reciprocal clustering methods are scale preserving. ■

Recall that reciprocal and nonreciprocal clustering are particular cases of semi-reciprocal clustering when the family index is $t = 2$ [(3.59)] and when $t \geq n$ exceeds the number of nodes in the network [cf. (3.60)]. Thus, Proposition 13 implies that, in particular, reciprocal and nonreciprocal clustering are scale preserving. As an example consider the network in Fig. 5.1 and dissimilarities transformed through $\phi(z) = 2z$. The nonreciprocal ultrametrics in the original network are $u_X^{\mathrm{NR}}(x, x') = 1$ for all nodes and the reciprocal ultrametrics are $u_X^{R}(x, x') = 2$. After transformation, the respective ultrametrics are $u_Y^{\mathrm{NR}}(y, y') = 2 = 2\, u_X^{\mathrm{NR}}(x, x')$ and $u_Y^{\mathrm{R}}(y, y') = 4 = 2\, u_X^{\mathrm{R}}(x, x')$, which is consistent with Proposition 13.

Remark 9 The proof of Proposition 13 can be mimicked to show that the directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$ is also scale preserving.

Proposition 13 notwithstanding scale preservation is a condition independent of axioms (A1) and (A2). To see this suffices to note that the grafting method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ with output ultrametric defined in (3.35) – which is admissible with respect to (A1)-(A2) – is not scale preserving. We show the latter through the example in Fig. 5.1 with dissimilarities transformed through $\phi(z) = 2z$. For $\beta = 3$ the original ultrametric value is $u_X^{\mathrm{R}}(x, x') = 2 \leq \beta = 3$ and we therefore select nonreciprocal ultrametrics as the value of $u_X^{R/NR}(x, x'; \beta) = u_X^{NR}(x, x') = 1$. In the transformed network the ultrametric value is $u_Y^{\mathrm{R}}(y, y') = 4 > \beta = 3$ and we therefore select $u_Y^{R/NR}(y, y'; \beta) = u_Y^{R}(y, y') = 4$ which is different from $\phi\big(u_X^{R/NR}(x, x'; \beta)\big) = 2\, u_X^{R/NR}(x, x'; \beta) = 2$. The reason for having outputs that are not scale preserving is that we may have points for which $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ selects the nonreciprocal ultrametrics in the original networks and the reciprocal in the transformed network.

A weaker notion of scale preservation is the concept of linear scale preservation that we introduce next.

(P2') Linear Scale Preservation. A hierarchical clustering method $\mathcal{H}$ is said to be linear scale preserving if it satisfies the Scale Preservation Property (P2) for all linear functions $\psi(z) = az$ with $a > 0$.

The Scale Preservation Property (P2) implies linear scale preservation (P2') but not vice versa. Further note that linear scale preservation is also independent of axioms (A1)-(A2) since the transformation in Fig. 5.1, which is used to argue that the grafting method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ is not scale preserving, is a linear transformation with $a = 2$. Linear scale preservation is central to the relationship between the concepts of representability and excisiveness that we discuss in Sections 5.2 and 5.3 and connect in Theorem 12.

5.1.1 Similarity networks

Define a similarity function $S_X : X \times X \to \mathbb{R}_+ \cup \{+\infty\}$ such that the similarity $S_X(x, x')$ between nodes $x, x' \in X$ is assumed to be non-negative for all pairs (x, x') and $+\infty$ if and only if $x = x'$. In this case, a similarity value equal to 0 represents that the corresponding nodes are not directly related. In some applications, we may be interested in clustering a network of which we are given a similarity function S_X instead of dissimilarity information. In such case, the theory developed can be applied by transforming S_X into dissimilarity $A_X = f \circ S_X$ obtained by composition with a decreasing function f. Given that different transformations could be used, the concern is what is the relation between the clustering outcomes u_X and u'_X resulting from the application of hierarchical clustering method $\mathcal{H}$ to networks with dissimilarities $A_X = f \circ S_X$ and $A'_X = f' \circ S_X$ corresponding to different decreasing functions f and f'. For admissible scale preserving methods, outcomes u_X and u'_X have the same clustering structure as we formally state in the following proposition.

Proposition 14 *Given an admissible scale preserving clustering method $\mathcal{H}$, a node set X, a similarity function $S_X : X \times X \to \mathbb{R}_+ \cup \{+\infty\}$, and two functions $f, f' : \mathbb{R}_+ \cup \{+\infty\} \to \mathbb{R}_+$ with $f(+\infty) = f'(+\infty) = 0$, $f(z) > 0$ and $f'(z) > 0$ for finite z and decreasing in $\mathbb{R}_+$, compute dissimilarity functions $A_X = f \circ S_X$ and $A'_X = f' \circ S_X$. Further, denote by $(X, u_X) = \mathcal{H}(X, A_X)$ and $(X, u'_X) = \mathcal{H}(X, A'_X)$ the ultrametrics resulting from application of $\mathcal{H}$ to the networks (X, A_X) and (X, A'_X), respectively. Then, for all points x_1,x_2, x_3, and x_4 we have*

$$u_X(x_1, x_2) \geq u_X(x_3, x_4) \iff u'_X(x_1, x_2) \geq u'_X(x_3, x_4). \tag{5.10}$$

Proof: Assume that for four arbitrary nodes $x_1, x_2, x_3, x_4 \in X$ the following inequality holds

$$u_X(x_1, x_2) \geq u_X(x_3, x_4). \tag{5.11}$$

Suppose that we can find a function ψ satisfying the conditions in (P2) such that $A'_X = \psi \circ A_X$. Then, since ψ is nondecreasing, we can apply it to (5.11) and the inequality still holds

$$\psi \circ u_X(x_1, x_2) \geq \psi \circ u_X(x_3, x_4). \tag{5.12}$$

However, since the method $\mathcal{H}$ is assumed to be scale preserving (P2), then $u'_X = \psi \circ u_X$. Hence, (5.12) is equivalent to

$$u'_X(x_1, x_2) \geq u'_X(x_3, x_4). \tag{5.13}$$

This means that, assuming the function ψ can be constructed, (5.11) implies (5.13) which is one direction of the assertion (5.10) we are proving. The conditional in the opposite direction follows similarly if we are able to find a function ψ' satisfying conditions in (P2) such that $A_X = \psi' \circ A'_X$.

To complete the proof, we illustrate a possible construction of ψ. The function ψ' can be constructed following an analogous procedure. Consider all the distinct, finite elements in the image of S_X, say that there are m of them, and index them in decreasing order. I.e., sort them such that $s_X^{(1)} > s_X^{(2)} > \ldots > s_X^{(m)}$. Apply function f to the image of S_X to obtain $a_X^{(i)} = f(s_X^{(i)}) > 0$. Further, note that since f is decreasing, we have $a_X^{(i)} < a_X^{(j)}$ for $i < j$. Similarly, apply f' to the image of S_X in order to obtain the increasing sequence $\{a_X^{\prime(i)}\}_{1 \leq i \leq m}$.

We now construct the function ψ as the piecewise linear function

$$\psi(z) = \begin{cases} \frac{a_X^{\prime(i+1)} - a_X^{\prime(i)}}{a_X^{(i+1)} - a_X^{(i)}} (z - a_X^{(i)}) + a_X^{\prime(i)}, & \text{for } a_X^{(i)} \leq z < a_X^{(i+1)}, \text{with } i = 0, \ldots, m-1 \\ \\ (z - a_X^{(m)}) + a_X^{\prime(m)}, & \text{for } z \geq a_X^{(m)} \end{cases} \tag{5.14}$$

where we have defined $a_X^{(0)} = a_X^{\prime(0)} = 0$ for notational consistency. The construction of ψ is illustrated in Fig. 5.2. We map 0 to $\psi(0) = 0$, the smallest value $a_X^{(1)}$ in the image of A_X to $\psi(a_X^{(1)}) = a_X^{\prime(1)}$ and, in general, the ith smallest value $a_X^{(i)}$ in the image of A_X to $\psi(a_X^{(i)}) = a_X^{\prime(i)}$. These points are joined by line segments. The function is also extended linearly for values $z > a_X^{(m)}$.

From the above construction, it follows that ψ fulfills the scale preservation conditions in (P2) and satisfies $\psi \circ A_X = A'_X$, completing the proof. ■

Proposition 14 indicates that regardless of the function f used to transform the similarity function S_X into the dissimilarity function A_X, the resulting dendrograms are qualitatively equivalent. Although nodes may cluster at different resolutions, the clustering order is the same for all functions f. This property is helpful in many practical instances. In, e.g., trust

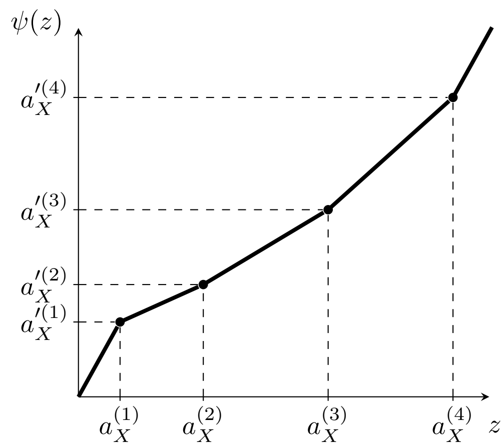


Figure 5.2: Construction of ψ in the proof of Proposition 14. This piecewise linear increasing function is constructed to map $a_X^{(i)}$, the ith smallest value in the image of A_X, to $a_X'^{(i)}$, the ith smallest value in the image of A'_X.

networks it is more natural for subjects to express their trust of neighbors rather than their distrust. Furthermore, in social networks proximity indicators like number of exchanged messages are more common than distance indicators.

5.2 Representability

We build upon the notion of representable methods – introduced for non-hierarchical clustering in [11] – to specify the clustering of arbitrary networks through the clustering of particular examples that we call *representers*. To explain the concept of a representable method we first present an alternative definition for the reciprocal ultrametric (3.2). Start by considering a given asymmetric network $N = (X, A_X)$ and define $\circlearrowleft_2 := \vec{\Delta}_2(1,1)$, i.e. a two-node network with both dissimilarities equal to 1. Define the λ-multiple of the network $\circlearrowleft_2$ as the network $\lambda * \circlearrowleft_2 = (\{p,q\}, \lambda A_{p,q})$ whose underlying set is the same and its dissimilarities are linearly scaled by a given $\lambda > 0$. Further define the Lipschitz constant of a map $\phi : \circlearrowleft_2 \to N$ as

$$L(\phi; \circlearrowleft_2, N) := \max\left(A_X(\phi(p), \phi(q)), A_X(\phi(q), \phi(p))\right), \tag{5.15}$$

i.e., the maximum dissimilarity into which one of the unit dissimilarities in $\circlearrowleft_2$ is mapped. Building upon this notion, for arbitrary nodes $x, x' \in X$ we define the optimal multiple $\lambda_X^{\circlearrowleft_2}(x, x')$ between x and x' with respect to $\circlearrowleft_2$ as

$$\lambda_X^{\circlearrowleft_2}(x,x') = \big\{ L(\phi; \circlearrowleft_2, N) \mid \phi : \{p,q\} \to X,\ x, x' \in \mathrm{Im}(\phi) \big\}. \tag{5.16}$$

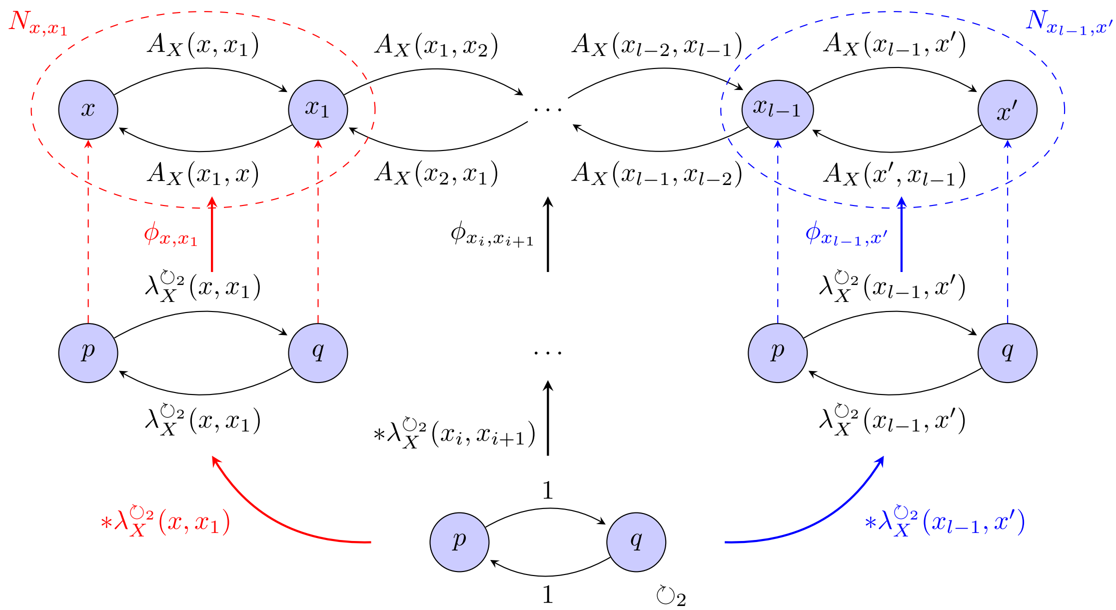


Figure 5.3: Representable method $\mathcal{H}^{\circlearrowright_2}$ with ultrametric output as in (5.17). For every pair x_i, x_{i+1} of consecutive nodes in the path $P_{xx'}$ we multiply the network $\circlearrowright_2$ by the infimal multiple $\lambda_X^{\circlearrowright_2}(x_i, x_{i+1})$ that allows the existence of a dissimilarity reducing map $\phi_{x_i,x_{i+1}}$ containing nodes x_i and x_{i+1} in its image. The maximum among all these multiples determines the cost of the path $P_{xx'}$. The value of $u_X^{\circlearrowright_2}(x,x')$ arises as the minimum of this path cost over all possible paths linking x to x'.

Notice that $\lambda_X^{\circlearrowright_2}(x,x')$ is the minimum multiple needed for the existence of a map with x and x' in its image between a multiple of $\circlearrowright_2$ and N.

To further interpret $\lambda_X^{\circlearrowright_2}(x,x')$ define the subnetwork $N_{x,x'} := (\{x,x'\}, A_{x,x'})$ with dissimilarities $A_{x,x'}(x,x') = A_X(x,x')$ and $A_{x,x'}(x',x) = A_X(x',x)$ which is formed by extracting points x and x' from X. We can then think of $\lambda_X^{\circlearrowright_2}(x,x')$ as the minimum scaling λ of the network $\circlearrowright_2$ that allows us to *fit* the subnetwork $N_{x,x'}$ into the scaled network $\lambda_X^{\circlearrowright_2}(x,x') * \circlearrowright_2$; where the intuitive notion of fitting $N_{x,x'}$ into $\lambda_X^{\circlearrowright_2}(x,x') * \circlearrowright_2$ is formalized by requiring the existence of a dissimilarity reducing map that allows us to bijectively map $\lambda_X^{\circlearrowright_2}(x,x') * \circlearrowright_2$ into $N_{x,x'}$. Observe that the optimal multiples in (5.16) are symmetric.

To define the representable clustering method $\mathcal{H}^{\circlearrowright_2}$ associated with the representer network $\circlearrowright_2$ we consider paths $P_{xx'} = [x = x_0, \dots, x_l = x']$ linking x to x' where the cost of the link x_i, x_{i+1} is given by the optimal multiple $\lambda_X^{\circlearrowright_2}(x_i, x_{i+1})$ that allows us to *fit* N_{x_i,x_i+1} into $\lambda_X^{\circlearrowright_2}(x_i, x_{i+1}) * \circlearrowright_2$. The cost of each path is defined as the maximum of all link costs and the representable clustering method $\mathcal{H}^{\circlearrowright_2}$ is defined as the method with ultrametric output $(X, u_X^{\circlearrowright_2}) = \mathcal{H}^{\circlearrowright_2}(N)$ given by the minimum of these costs across all possible paths linking x to x',

$$u_X^{\circlearrowright_2}(x,x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \lambda_X^{\circlearrowright_2}(x_i, x_{i+1}). \tag{5.17}$$

The definition in (5.17) is illustrated in Fig. 5.3.

It is immediate that the method $\mathcal{H}^{\circlearrowright_2}$ with output ultrametrics as in (5.17) is equivalent to reciprocal clustering $\mathcal{H}^{\mathrm{R}}$ with output ultrametrics as in (3.2). Indeed, given a network $N = (X, A_X)$ and points $x, x' \in X$ there are only two possible maps from $\lambda * \circlearrowright_2$ to N containing nodes x and x' in their images. One map takes p to x and q to x' and the other reverses the images and takes p to x' and q to x. However, either maps have the same Lipschitz constant as defined in (5.15), ensuring that the optimal multiple in (5.16) is well defined. Consequently, we obtain that

$$\lambda_X^{\circlearrowright_2}(x, x') = \max\big(A_X(x, x'), A_X(x', x)\big) = \bar{A}_X(x, x'). \tag{5.18}$$

Comparing (5.17) with (3.2) and using the observation in (5.18) the equivalence $\mathcal{H}^{\mathrm{R}} \equiv \mathcal{H}^{\circlearrowright_2}$ follows. We interpret this equivalence by saying that the reciprocal clustering method $\mathcal{H}^{\mathrm{R}} \equiv \mathcal{H}^{\circlearrowright_2}$ is represented by the network $\circlearrowright_2$. I.e., if we consider $\circlearrowright_2$ as an interaction modality defining a basic clustering unit, the hierarchical clustering of a generic network follows from application of (5.17) or its equivalent (3.2).

The definition in (5.17) is certainly more cumbersome than (3.2). However, the former can be generalized to cases in which we consider arbitrary representers in lieu of $\circlearrowright_2$ as we explain in the following section.

5.2.1 Representable hierarchical clustering methods

Generalizing $\mathcal{H}^{\circlearrowright_2}$ entails redefining the Lipschitz constant of a map and the optimal multiples so that they are calculated with respect to an arbitrary representer network $\omega = (X_\omega, A_\omega)$ instead of $\circlearrowright_2$. In representer networks ω we allow the domain $\mathrm{dom}(A_\omega)$ of the dissimilarity function A_ω to be a proper subset of the product space, i.e., we may have $\mathrm{dom}(A_\omega) \neq X_\omega \times X_\omega$. This relaxation allows for a situation in which dissimilarities may not be defined for some pairs $(x, x') \notin \mathrm{dom}(A_\omega)$ and is a technical modification that allows representer networks to have some dissimilarities that can be interpreted as arbitrarily large. Generalizing (5.15), given an arbitrary network $N = (X, A_X)$ we define the *Lipschitz constant* of a map $\phi : \omega \to N$ as

$$L(\phi; \omega, N) := \max_{\substack{(z,z') \in \mathrm{dom}(A_\omega) \\ z \neq z'}} \frac{A_X(\phi(z), \phi(z'))}{A_\omega(z, z')}. \tag{5.19}$$

Notice that $L(\phi; \omega, N)$ is the minimum multiple of the network ω such that the considered map ϕ is dissimilarity reducing from $L(\phi; \omega, N) * \omega$ to N. In consistency with this interpretation, if no pair of distinct nodes (z, z') is contained in $\mathrm{dom}(A_\omega)$, then we fix $L(\phi; \omega, N) = 0$. Further, observe that (5.19) reduces to (5.15) when $\omega = \circlearrowright_2$. Notice as well that the maximum in (5.19) is computed for pairs (z, z') in the domain of A_ω. Pairs not belonging to the domain could be mapped to any dissimilarity without modifying the value of the Lipschitz constant. Mimicking (5.16), for arbitrary nodes $x, x' \in X$ we define the optimal multiple $\lambda_X^\omega(x, x')$ between x and x' with respect to ω as

$$\lambda_X^\omega(x,x') = \min\left\{L(\phi;\omega,N) \mid \phi : X_\omega \to X,\ x,x' \in \mathrm{Im}(\phi)\right\}. \tag{5.20}$$

This means that $\lambda_X^\omega(x, x')$ is the minimum Lipschitz constant among those maps that have x and x' in its image. Equivalently, it is the minimum multiple needed for the existence of a map from a multiple of ω to N that has x and x' in its image. Observe that (5.20) reduces to (5.16) when $\omega = \circlearrowleft_2$.

To give an interpretation of $\lambda_X^\omega(x, x')$ analogous to the interpretation of $\lambda_X^{\circlearrowleft_2}(x, x')$ define the collection $N_{x,x'}^\omega$ formed by extracting subnetworks from the network N such that each subnetwork contains x and x' and has a number of nodes not larger than the number of nodes in the representer network ω

$$N_{x,x'}^\omega := \left\{(X_{x,x'}, A_{x,x'}) \,\middle|\, x, x' \in X_{x,x'}, |X_{x,x'}| \le |X_\omega|,\ A_{x,x'} = A_X\big|_{X_{x,x'} \times X_{x,x'}}\right\}. \tag{5.21}$$

The notation $A_X|_{X_{x,x'} \times X_{x,x'}}$ refers to the function A_X restricted to the domain $X_{x,x'} \times X_{x,x'}$ and reflects that the subnetwork $(X_{x,x'}, A_{x,x'})$ is a piece of the network N.

With the definition in (5.20) we can interpret $\lambda_X^\omega(x, x')$ as the minimum multiple of the representer network ω that allows us to fit at least one subnetwork $N_{x,x'} \in N_{x,x'}^\omega$ into the scaled network $\lambda_X^\omega(x, x') * \omega$. The intuitive notion of fitting a given subnetwork $N_{x,x'}$ into $\lambda_X^\omega(x, x') * \omega$ is formalized by requiring that there exists a dissimilarity reducing map that allows us to map $\lambda_X^\omega(x, x') * \omega$ into $N_{x,x'}$. For the case where $\omega = \circlearrowleft_2$, the collection $N_{x,x'}^{\circlearrowleft_2}$ contains only one network $N_{x,x'} = (\{x, x'\}, A_{x,x'})$ and we recover the interpretation succeeding (5.16).

Remark 10 Similar to the case where $\omega = \circlearrowleft_2$, it is in general true that we have symmetric costs $\lambda_X^\omega(x, x') = \lambda_X^\omega(x', x)$ for all $x, x' \in X$ due to the symmetry in definition (5.20).

We can now define the representable method $\mathcal{H}^\omega$ associated with a given representer ω by defining the cost of a path $P_{xx'} = [x = x_0, \ldots, x_l = x']$ linking x to x' as the maximum multiple $\lambda_X^\omega(x_i, x_{i+1})$ that allows us to fit some subnetwork in the set $N_{x_i,x_{i+1}}^\omega$ into the scaled network $\lambda_X^\omega(x_i, x_{i+1}) * \omega$. The ultrametric u_X^ω associated with output $(X, u_X^\omega) = \mathcal{H}^\omega(X, A_X)$ is given by the minimum path cost

$$u_X^\omega(x,x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \lambda_X^\omega(x_i, x_{i+1}), \tag{5.22}$$

for all $x, x' \in X$. Representable methods are generalized to cases in which we are given a nonempty set Ω of representer networks ω. In such case we define the function λ_X^Ω by considering the infimum across all representers $\omega \in \Omega$,

$$\lambda_X^\Omega(x, x') \;=\; \inf_{\omega \in \Omega} \; \lambda_X^\omega(x, x'), \tag{5.23}$$

for all $x, x' \in X$. The value $\lambda_X^\Omega(x, x')$ is the infimum across all multiples λ such that there exists a representer network $\omega \in \Omega$ which allows us to fit at least one subnetwork $N_{x,x'} \in N_{x,x'}^\omega$ into the scaled network $\lambda * \omega$. For a given network $N = (X, A_X)$, the representable clustering method $\mathcal{H}^\Omega$ associated with the collection of representers Ω is the one with outputs $(X, u_X^\Omega) = \mathcal{H}^\Omega(X, A_X)$ such that the ultrametric u_X^Ω is given by

$$u_X^\Omega(x, x') = \min_{P_{xx'}} \; \max_{i | x_i \in P_{xx'}} \lambda_X^\Omega(x_i, x_{i+1}), \tag{5.24}$$

for all $x, x' \in X$. The definition in (5.24) is interpreted in Fig. 5.4. Given points $x, x' \in X$ and a path $P_{xx'} = [x = x_0, \ldots, x_l = x']$ we extract the collections of subnetworks N_{x_i, x_i+1}^ω for all $\omega \in \Omega$ and consider scalings of the representer networks ω that allow us to map at least one scaled network $\lambda_X^\Omega(x, x') * \omega$ into at least one subnetwork $N_{x_i, x_i+1} \in N_{x_i, x_i+1}^\omega$ with a dissimilarity reducing map ϕ_{x_i, x_i+1} containing x_i and x_{i+1} in its image. The maximum of all these scalings among all pairs of consecutive nodes x_i, x_{i+1} is the cost of the given path. The ultrametric value $u_X^\Omega(x, x')$ between x and x' is given by the minimum of this cost across all paths linking x and x'. An example application of (5.24) is the representable construction of reciprocal clustering illustrated in Fig. 5.3. A different example is given in Section 5.2.4.

Since not all dissimilarities are necessarily defined in representer networks the issue of whether a representer network is connected or not plays a role in the validity and admissibility of representable methods. We say that a representer network $\omega = (X_\omega, A_\omega)$ is *weakly connected* if for every pair of nodes $x, x' \in X_\omega$ we can find a path $P_{xx'} = [x = x_0, \ldots, x_l = x']$ such that either $(x_i, x_{i+1}) \in \mathrm{dom}(A_\omega)$ or $(x_{i+1}, x_i) \in \mathrm{dom}(A_\omega)$ or both for $i = 0, \ldots, l-1$. The representer network is said to be *strongly connected* if for every pair of points $x, x' \in X_\omega$ there is a path $P_{xx'} = [x = x_0, \ldots, x_l = x']$ such that $(x_i, x_{i+1}) \in \mathrm{dom}(A_\omega)$ for $i = 0, \ldots, l-1$. Notice that strong connectedness implies weak connectedness. When a network is not weakly connected, then it cannot be strongly connected either. For brevity, we refer to any such networks as disconnected or not connected.

Collections that include representers that are not connected or representers for which dissimilarity values $A_\omega(x, x')$ are arbitrarily large for some $x, x' \in X$ – this can happen in sets Ω containing an infinite number of elements and is not to be confused with the possibility of having A_ω undefined for some pairs $(z, z') \notin \mathrm{dom}(A_\omega)$ – may yield outputs

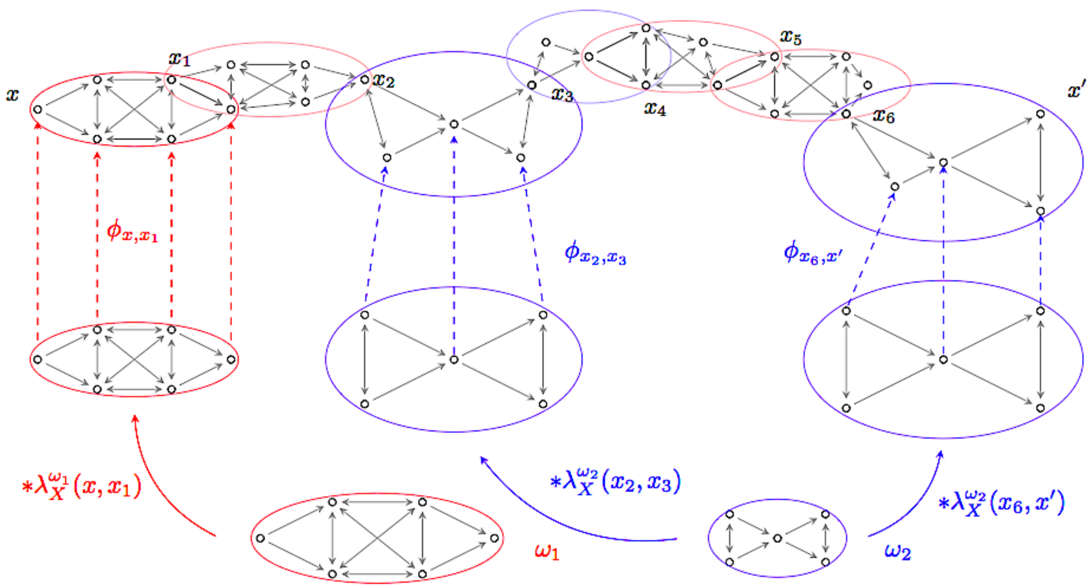


Figure 5.4: Representable method $\mathcal{H}^{\Omega}$ with ultrametric output as in (5.24). The collection of representers $\Omega = \{\omega_1, \omega_2\}$ is shown at the bottom. In order to compute $u_X^{\Omega}(x, x')$ we link x and x' through a path, e.g. $[x, x_1, \ldots, x_6, x']$ in the figure, and link pairs of consecutive nodes with multiples of the representers. The ultrametric value $u_X^{\Omega}(x, x')$ is given by minimizing over all paths joining x and x' the maximum multiple of a representer used to link consecutive nodes in the path (5.24).

that are not valid ultrametrics for some networks. We say that Ω is *uniformly bounded* if and only if there exists a finite $M > 0$ such that

$$\max_{(z,z') \in \mathrm{dom}(A_\omega)} A_\omega(z, z') \leq M, \tag{5.25}$$

for all $\omega = (X_\omega, A_\omega) \in \Omega$. We can now formally define the notion of representability.

> *(P3) Representability.* We say that a clustering method $\mathcal{H}$ is representable if there exists a uniformly bounded collection Ω of weakly connected representers each with finite number of nodes such that $\mathcal{H} \equiv \mathcal{H}^{\Omega}$ where $\mathcal{H}^{\Omega}$ has output ultrametrics as in (5.24).

For every collection of representers Ω satisfying the conditions in property (P3), (5.24) defines a valid ultrametric as stated before the definition and formally claimed by the following proposition.

Proposition 15 *Given a collection of representers Ω satisfying the conditions in (P3), the representable method $\mathcal{H}^{\Omega}$ is valid. I.e., u_X^{Ω} defined in* (5.24) *is an ultrametric for all networks $N = (X, A_X)$.*

Proof: Given a collection Ω of representers $\omega = (X_\omega, A_\omega)$ we need to show that for arbitrary network $N = (X, A_X)$ the output $(X, u_X^{\Omega}) = \mathcal{H}^{\Omega}(X, A_X)$ satisfies the identity, symmetry, and strong triangle inequality properties. To show that the strong triangle inequality in (2.12) is satisfied let $P^*_{xx'}$ and $P^*_{x'x''}$ be minimizing paths for $u^\Omega_X(x,x')$ and $u^\Omega_X(x',x'')$, respectively. Consider the concatenated path $P_{xx''} = P^*_{xx'} \uplus P^*_{x'x''}$ and notice that the maximum over i of the optimal multiples $\lambda^\Omega_X(x_i,x_{i+1})$ in $P_{xx''}$ does not exceed the maximum multiples in each individual path. Thus, the maximum multiple in the path $P_{xx''}$ suffices to bound $u^\Omega_X(x,x'') \leq \max\big(u^\Omega_X(x,x'), u^\Omega_X(x',x'')\big)$ by (5.24) as in (2.12).

To show the symmetry property, $u^\Omega_X(x,x') = u^\Omega_X(x',x)$ for all $x,x' \in X$, recall from Remark 10 that we have $\lambda^\omega_X(x,x') = \lambda^\omega_X(x',x)$ for every representer ω. From (5.23) we then obtain that λ^Ω_X is symmetric. But having symmetric costs in (5.24) implies that if a given path $P_{xx'}$ is a minimizing path when going from x to x', the path traversed in the opposite direction $P_{x'x}$ has to be minimizing when going from x' to x. Symmetry of u^Ω_X follows.

For the identity property, i.e. $u^\Omega_X(x,x') = 0$ if and only if $x = x'$, we first show that if $x = x'$ we must have $u^\Omega_X(x,x') = 0$. Pick any $x \in X$, let $x' = x$ and pick the path $P_{xx} = [x,x]$ starting and ending at x with no intermediate nodes as a candidate minimizing path in (5.24). While this particular path need not be optimal in (5.24) it nonetheless holds that

$$0 \leq u^\Omega_X(x,x) \leq \lambda^\Omega_X(x,x), \tag{5.26}$$

where the first inequality holds because all costs $\lambda^\Omega_X(x_i,x_{i+1})$ in (5.24) are nonnegative since they correspond to the Lipschitz constant of some map [cf. (5.19)]. Notice that for the cost $\lambda^\omega_X(x,x)$ in (5.20) we minimize the Lipschitz constant among maps $\phi_{x,x}$ that are only required to have node x in its image. Thus, consider the map that takes all the nodes in any representer $\omega \in \Omega$ into node $x \in X$. From (5.19), the Lipschitz constant of this map is zero which implies by (5.20) that $\lambda^\omega_X(x,x) = 0$ for all $\omega \in \Omega$. Combining this result with (5.23) we then get

$$\lambda^\Omega_X(x,x) = 0. \tag{5.27}$$

Substituting (5.27) in (5.26) we conclude that $u^\Omega_X(x,x) = 0$.

In order to show that if $u^\Omega_X(x,x') = 0$ we must have $x = x'$ we prove that if $x \neq x'$ we must have $u^\Omega_X(x,x') > \alpha > 0$ for some strictly positive constant α. In the proof we make use of the following claim.

Claim 5 *Given a network $N = (X, A_X)$, a weakly connected representer $\omega = (X_\omega, A_\omega)$, and a dissimilarity reducing map $\phi : X_\omega \to X$ whose image satisfies $|Im(\phi)| \geq 2$, there exists a pair of points $(z,z') \in dom(A_\omega)$ for which $\phi(z) \neq \phi(z')$.*

Proof: Suppose that $\phi(z^1) = x^1$ and $\phi(z^2) = x^2$, with $x^1 \neq x^2 \in X$. These nodes can always be found since $|\mathrm{Im}(\phi)| \geq 2$. By our hypothesis, the network is weakly connected. Hence, there must be a path $P_{z^1z^2} = [z^1 = z_0, z_1, \ldots, z_l = z^2]$ linking z^1 and z^2 for which either $(z_i, z_{i+1}) \in \mathrm{dom}(A_\omega)$ or $(z_{i+1}, z_i) \in \mathrm{dom}(A_\omega)$ for all $i = 0, \ldots, l-1$. Focus on the image of this path under the map ϕ, $P_{x^1x^2} = [x^1 = \phi(z_0), \phi(z_1), \ldots, \phi(z_l) = x^2]$. Notice that not all the nodes are necessarily distinct, however, since the extreme nodes are different by construction, at least one pair of consecutive nodes must differ, say $\phi(z_p) \neq \phi(z_{p+1})$. Due to ω being weakly connected, in the original path we must have either (z_p, z_{p+1}) or $(z_{p+1}, z_p) \in \mathrm{dom}(A_\omega)$. Hence, either $z = z_p$ and $z' = z_{p+1}$ or vice versa must fulfill the statement of the claim. ■

Returning to the main proof, observe that since pairwise dissimilarities in all networks $\omega \in \Omega$ are uniformly bounded, the maximum dissimilarity across all links of all representers

$$d_{\max} = \sup_{\omega \in \Omega} \max_{(x,x') \in \mathrm{dom}(A_\omega)} A_\omega(x, x'), \tag{5.28}$$

is guaranteed to be finite. Recalling the definition of separation of a network $\mathrm{sep}(X, A_X)$ in (2.10), pick any real α such that $0 < \alpha < \mathrm{sep}(X, A_X)/d_{\max}$. Then for all $(z, z') \in \mathrm{dom}(A_\omega)$ and all $\omega \in \Omega$ we have

$$\alpha\ A_\omega(z, z') < \mathrm{sep}(X, A_X). \tag{5.29}$$

Claim 5 implies that independently of the map ϕ chosen, this map transforms some defined dissimilarity in ω, i.e. $A_\omega(z, z')$ for some $(z, z') \in \mathrm{dom}(A_\omega)$, into a dissimilarity in N. Moreover, every positive dissimilarity in N is greater than or equal to the network separation $\mathrm{sep}(X, A_X)$. Hence, (5.29) implies that there cannot be any dissimilarity reducing map ϕ with $|\mathrm{Im}(\phi)| \geq 2$ from $\alpha * \omega$ to N for any $\omega \in \Omega$. From (5.20), this implies that for all $x \neq x' \in X$ and for all ω

$$\lambda_X^\omega(x, x') > \alpha > 0. \tag{5.30}$$

Substituting (5.30) in (5.23) we conclude that

$$\lambda_X^\Omega(x, x') > \alpha > 0, \tag{5.31}$$

for all $x \neq x'$. Hence, substituting (5.31) in definition (5.24) we have that the ultrametric value between two different nodes $u_X^\Omega(x, x') \geq \min_{x \neq x'} \lambda_X^\Omega(x, x') > \alpha > 0$ must be strictly positive. ■

Remark 11 The condition in (P3) that a valid representable method is defined by a set of weakly connected representers is necessary and sufficient. Indeed, consider the network $\omega = (\{p, q\}, A_\omega)$ composed of two isolated nodes, i.e. (p, q) and $(q, p) \notin \mathrm{dom}(A_\omega)$, which is therefore not connected. By (5.19) we have that $L(\phi; \omega, N) = 0$ for any map $\phi : \omega \to N$ into any arbitrary network N. Hence, the ultrametric generated for any network $N = (X, A_X)$ is $u_X^\omega(x, x') = 0$ for nodes $x \neq x'$ violating the identity property. If we have an arbitrary network with at least two disconnected components we can generalize the argument by considering a map such that the image of all the elements of one component is x and the image of all the elements of the other component is x'. In this case, the Lipschitz constant is also null. Thus, any set Ω that contains at least one disconnected network yields the invalid outcome $u_X^\Omega \equiv 0$.

Remark 12 The condition in (P3) that Ω be uniformly bounded is sufficient but not necessary for $\mathcal{H}^\Omega$ to output a valid ultrametric as there are methods induced by representers with arbitrarily large dissimilarities that nonetheless output valid ultrametrics. Infinite collections of representers each with bounded dissimilarities may fail for a reason similar to the one why disconnected representers fail. Consider, e.g., the method defined by the countable collection of representers $\Omega = \{n* \circlearrowright_2 = \big(\{p,q\}, nA_{p,q}\big)\}_{n\in\mathbb{N}}$ with $nA_{p,q}(p,q) = nA_{p,q}(q,p) = n$. Clearly, Ω is not uniformly bounded. For an arbitrary network $N = (X, A_X)$ consider a pair $x, x' \in X$ such that $x \neq x'$. Then, from (5.19) we have that $L(\phi; n* \circlearrowright_2, N) = (1/n) \max\big(A_X(x,x'), A_X(x',x)\big)$ for any map ϕ with x, x' in its image. Then, (5.20) impies that $\lambda_X^{n*\circlearrowright_2}(x,x') = 1/n \max\big(A_X(x,x'), A_X(x',x)\big)$, and since in (5.23) we minimize over all n, we have the (invalid) outcome $u_X^\Omega \equiv 0$. Intuitively, the network $n* \circlearrowright_2$ is indistinguishable from a disconnected network for sufficiently large n. If we modify the class of representers to $\Omega' = \{\big(\{p,q\}, A_n\big)\}_{n\in\mathbb{N}}$ with $A_n(p,q) = 1$ and $A_n(q,p) = n$ the resulting method can be seen to output a valid ultrametric. Indeed, the situation is akin to having one two-node representer that is weakly, but not strongly, connected by an edge of unit dissimilarity. This case is considered in Proposition 15 and hence the represented method must output a valid ultrametric. Thus, Ω' is an example of a non uniformly bounded collection of representers whose associated method outputs a valid ultrametric. In fact, the method associated with such collection is unilateral clustering $\mathcal{H}^U$, introduced in Section 3.4.1.

Regarding admissibility with respect to axioms (A1) and (A2) there are representable methods that do not satisfy (A1) but all representable methods abide by the Axiom of Transformation (A2). We claim the latter in the following proposition and discuss representable methods that do not comply with (A1) after that.

Proposition 16 *If a hierarchical clustering method $\mathcal{H}$ is representable as defined in (P3) then it satisfies the Axiom of Transformation (A2).*

Proof: The proof is a generalization of the argument used to show that reciprocal clustering satisfies (A2) in the proof of Proposition 1. Let Ω be a collection of representers $\omega = (X_\omega, A_\omega)$ characterizing the representable method $\mathcal{H}^\Omega$. Consider networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ and a dissimilarity reducing map $\phi : X \to Y$, i.e. a map ϕ such that for all $x, x' \in X$ it holds that $A_X(x,x') \geq A_Y(\phi(x), \phi(x'))$. Further denote by $(X, u_X^\Omega) = \mathcal{H}^\Omega(N_X)$

and $(Y, u_Y^\Omega) = \mathcal{H}^\Omega(N_Y)$ the outputs of applying $\mathcal{H}^\Omega$ to networks N_X and N_Y, respectively. We want to prove that the Axiom of Transformation holds for $\mathcal{H}^\Omega$, which according to (2.14) requires showing that

$$u_X^\Omega(x, x') \geq u_Y^\Omega(\phi(x), \phi(x')), \tag{5.32}$$

holds for all $x, x' \in X$. To see that (5.32) is true pick two arbitrary nodes $x, x' \in X$ and denote by $P^*_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ any minimizing path achieving the minimum in the definition of $u_X^\Omega(x, x')$ in (5.24). Consider the path $P_{\phi(x)\phi(x')} = [\phi(x) = y_0, \phi(x_1) = y_1, \ldots, \phi(x') = y_l]$ obtained by transforming the path $P^*_{xx'}$ via the map ϕ. For every dissimilarity reducing map $\phi_{x,x'} : X_\omega \to X$ between $\lambda_X^\omega(x, x') * \omega$ and N_X containing x and x' in its image, we can define the composition map $\phi \circ \phi_{x,x'} : X_\omega \to Y$ from $\lambda_X^\omega(x, x') * \omega$ to N_Y containing $\phi(x)$ and $\phi(x')$ in its image. Since the map ϕ is dissimilarity reducing, we obtain that

$$\begin{aligned} L(\phi \circ \phi_{x,x'}; \omega, N_Y) &= \max_{\substack{(z,z') \in \text{dom}(A_\omega) \\ z \neq z'}} \frac{A_Y(\phi \circ \phi_{x,x'}(z), \phi \circ \phi_{x,x'}(z'))}{A_\omega(z, z')} \\ &\leq \max_{\substack{(z,z') \in \text{dom}(A_\omega) \\ z \neq z'}} \frac{A_X(\phi_{x,x'}(z), \phi_{x,x'}(z'))}{A_\omega(z, z')} = L(\phi_{x,x'}; \omega, N_X) = \lambda_X^\omega(x, x') \end{aligned} \tag{5.33}$$

While the map $\phi \circ \phi_{x,x'}$ need not be the one achieving the infimum in (5.20), it suffices to bound

$$\lambda_Y^\omega(\phi(x), \phi(x')) \leq \lambda_X^\omega(x, x'), \tag{5.34}$$

for all $x, x' \in X$ and representers $\omega \in \Omega$. Substituting (5.34) in (5.23) we obtain that $\lambda_X^\Omega(x, x') \geq \lambda_Y^\Omega(\phi(x), \phi(x'))$ for all $x, x' \in X$. Substituting the latter conclusion in (5.24) we recover (5.32). ■

Representable methods that violate the Axiom of Value (A1) can be constructed. E.g., consider the representable method $\mathcal{H}^\Omega$ associated with the set $\Omega = \{2* \circlearrowright_2\}$ composed of the single representer $2* \circlearrowright_2$ and recall the definition of the generic two-node network $\vec{\Delta}_2(\alpha, \beta)$ in (2.1). Further denote by $\big(\{p, q\}, u_{p,q}^\Omega\big) = \mathcal{H}^\Omega(\vec{\Delta}_2(\alpha, \beta))$ the output of applying the clustering method $\mathcal{H}^\Omega$ to the network $\vec{\Delta}_2(\alpha, \beta)$. By (5.19) we have that $L(\phi; 2* \circlearrowright_2, \vec{\Delta}_2(\alpha, \beta)) = \max(\alpha, \beta)/2$. Consequently, we obtain $u_{p,q}^\Omega(p, q) = \max(\alpha, \beta)/2$. This contradicts (A1), which requires admissible ultrametrics to be $u_{p,q}(p, q) = \max(\alpha, \beta)$.

Remark 13 Representability is a mechanism for defining universal hierarchical clustering methods from given representative examples [11]. Each representer $\omega \in \Omega$ can be interpreted as defining a particular structure that is to be considered a cluster unit. The scaling of this unit structure [cf. (5.20)] and its replication through the network [cf. (5.22)] indicate the resolution at which nodes become part of a cluster. For nodes x and x' to cluster together at resolution δ we need to construct a path from x to x' by pasting versions of the representer network scaled by parameters not larger than δ. When we have multiple networks we can use any of them to build these paths [cf. (5.23) and (5.24)]. The interest in representability is that it is easier to state desirable clustering structures for particular networks rather than for arbitrary ones. E.g., reciprocal clustering is defined in Section 3.1 by specifying a methodology to determine the resolution at which nodes cluster together. In this section we redefined the method by defining the network $\circlearrowleft_2$ in Fig. 5.3 as a basic cluster unit. We refer the reader to Section 5.2.4 for particular examples of representer networks that give rise to intuitively appealing clustering methods.

5.2.2 Decomposition of representable methods

Comparing the ultrametric definition for representable methods u_X^Ω in (5.24) with the definition of the single linkage ultrametric u_X^{SL} in (2.15) we see that they coincide except that the multiples λ_X^Ω in (5.24) take the place of the original dissimilarities A_X in (2.15). Further recall that multiples $\lambda_X^\Omega(x,x') = \lambda_X^\Omega(x',x)$ in (5.24) are symmetric due to the fact that the multiples $\lambda_X^\omega(x,x') = \lambda_X^\omega(x',x)$ associated with each individual representer $\omega \in \Omega$ are symmetric as we discussed in Remark 10. Therefore, we can think of representable methods as the composition of a transformation from the original dissimilarities A_X into the symmetric multiples λ_X^Ω followed by application of single linkage. We assert the existence of this decomposition in the following proposition.

Proposition 17 *Every representable clustering method $\mathcal{H}^\Omega : \tilde{\mathcal{N}} \to \mathcal{U}$ having ultrametric outputs as given in* (5.24) *admits a decomposition of the form*

$$\mathcal{H}^\Omega \equiv \mathcal{H}^{\mathrm{SL}} \circ \Lambda^\Omega, \tag{5.35}$$

where $\Lambda^\Omega : \tilde{\mathcal{N}} \to \mathcal{N}$ is a map from the space of asymmetric networks $\tilde{\mathcal{N}}$ to the space of symmetric networks $\mathcal{N}$ and $\mathcal{H}^{\mathrm{SL}} : \mathcal{N} \to \mathcal{U}$ is the single linkage clustering method for symmetric networks, with ultrametrics as defined in (2.15).

Proof: The proof is just a matter of identifying elements in (5.24). Define the function Λ^Ω as the one that maps the network $N = (X, A_X)$ into

$$\Lambda^\Omega(X, A_X) = (X, \lambda_X^\Omega), \tag{5.36}$$

where the dissimilarity function λ_X^Ω has values $\lambda_X^\Omega(x,x')$ as given by (5.23) for all $x, x' \in X$. We want to show that (X, λ_X^Ω) is a symmetric network. From Remark 10 and (5.23) it follows that the function λ_X^Ω is symmetric. To show that λ_X^Ω is a dissimilarity function we need to prove that $\lambda_X^\Omega(x,x') = 0$ if and only if $x = x'$. This is shown in (5.27) and (5.31).

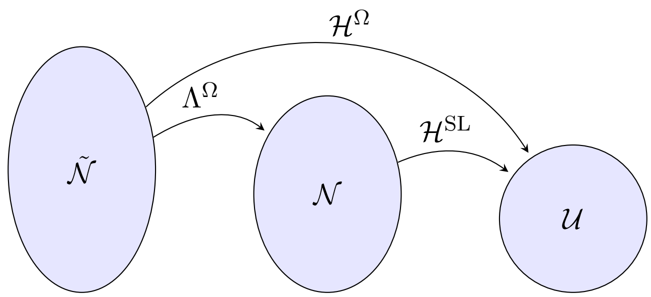


Figure 5.5: Decomposition of representable hierarchical clustering methods. Any representable method can be decomposed into a map Λ^{Ω} from the space of asymmetric networks $\tilde{\mathcal{N}}$ into the space of symmetric networks composed with the single linkage map into the space of ultrametrics.

Comparing the definitions of the output ultrametric of the representable method $\mathcal{H}^{\Omega}$ in (5.24) and the output ultrametric of the single linkage method for symmetric networks $\mathcal{H}^{\mathrm{SL}}$ in (2.15) we conclude

$$\mathcal{H}^{\Omega}(X, A_X) = \mathcal{H}^{\mathrm{SL}}(X, \lambda_X^{\Omega}) = \mathcal{H}^{\mathrm{SL}}\big(\Lambda^{\Omega}(X, A_X)\big), \tag{5.37}$$

where the last equality follows from (5.36). Since (5.37) is valid for every network (X, A_X) it follows that $\mathcal{H}^{\Omega} \equiv \mathcal{H}^{\mathrm{SL}} \circ \Lambda^{\Omega}$ for arbitrary Ω. ■

As a particular case of Proposition 17 consider the case when $\Omega = \{\circlearrowright_2\}$ which we have already seen yields the method $\mathcal{H}^{\circlearrowright_2} \equiv \mathcal{H}^{\mathrm{R}}$ equivalent to reciprocal clustering. Inspecting the definition of the reciprocal ultrametric in (3.2) and (3.1) we see that the method $\mathcal{H}^{\mathrm{R}}$ can indeed be written as $\mathcal{H}^{\mathrm{SL}} \circ \Lambda^{\circlearrowright_2}$ by defining the map $\Lambda^{\circlearrowright_2}$ to be $\Lambda^{\circlearrowright_2}(X, A_X) = (X, \max(A_X, A_X^T)) := (X, \bar{A}_X)$.

Representable clustering methods, as all other hierarchical clustering methods, are maps from the space of asymmetric networks $\tilde{\mathcal{N}}$ to the space of ultrametrics $\mathcal{U}$; see Fig. 5.5. Proposition 17 allows the decomposition of these maps into two components with definite separate roles. The first element of the composition is the function Λ^{Ω} whose objective is to symmetrize the original, possibly asymmetric, dissimilarity function. This transformation is followed by an application of single linkage $\mathcal{H}^{\mathrm{SL}}$ with the goal of inducing an ultrametric structure on this symmetric, but not necessarily ultrametric, intermediate network. Proposition 17 attests that there may be many different ways of inducing a symmetric structure depending on the selection of the representer set Ω but that there is a unique method to induce ultrametric structure. This unique method is $\mathcal{H}^{\mathrm{SL}}$, single linkage hierarchical clustering. This result is consistent, but not equivalent, with the observation in Corollary 2 that $\mathcal{H}^{\mathrm{SL}}$ is the unique map from the space $\mathcal{N}$ of symmetric networks into the space $\mathcal{U}$ of ultrametric networks that satisfies axioms (B1) and (A2).

From an algorithmic perspective, (5.35) implies that computation of ultrametrics from representable methods requires a symmetrization operation that depends on the representer set Ω followed by application of the algorithm for computation of single linkage in (3.91). A similar decomposition result is derived in [11] for non-hierarchical clustering in metric spaces. In [11, Theorem 6.3], the authors show that every representable non-hierarchical clustering method for metric spaces arises as the composition of a non-hierarchical clustering equivalent of single linkage and a change of metric. Proposition 17 extends this result to hierarchical clustering and moves away of metric spaces to consider asymmetric networks.

5.2.3 Representability, scale preservation, and admissibility

Do all representable clustering methods yield reasonable outcomes? The answer depends on the definition of reasonable outcome, and, as has been the approach in earlier sections, we seek to characterize methods that satisfy some desired properties that we deem reasonable. In particular, we consider methods that are admissible with respect to the Axioms of Value and Transformation (A1) and (A2) as well as scale preserving in the sense of (P2).

In characterizing admissible, representable, and scale preserving methods, the concept of structure representer appears naturally. We say that a representer network $\omega = (X_\omega, A_\omega)$ is a *structure representer* if and only if $|X_\omega| \geq 2$ and

$$A_\omega(z, z') = 1, \text{ for all } z \neq z' \text{ s.t. } (z, z') \in \text{dom}(A_\omega). \tag{5.38}$$

The requirement in (5.38) implies that structure representers define the relationships that are necessary in a cluster unit but do not distinguish between different levels of influence. In the following theorem we claim that admissible, representable, and scale preserving hierarchical clustering methods are those represented by a collection Ω of strongly connected, structure representers.

Theorem 11 *A representable clustering method $\mathcal{H}$ satisfies axioms (A1)-(A2) and scale preservation (P2) if and only if $\mathcal{H} \equiv \mathcal{H}^\Omega$ where Ω is a collection of strongly connected, structure representers as defined by the condition in* (5.38).

Proof: See Appendix A.1. ■

Recalling the interpretation of representability as the extension of clustering defined for particular cases, Theorem 11 entails that the definitions of particular cases cannot present dissimilarity degrees if we require scale preservation. That is, the dissimilarity between every pair of distinct nodes in the representers must be either 1 or undefined. The edges with value 1 imply that the corresponding influence relations are required for the formation of a cluster whereas the influence relations associated with undefined edges are

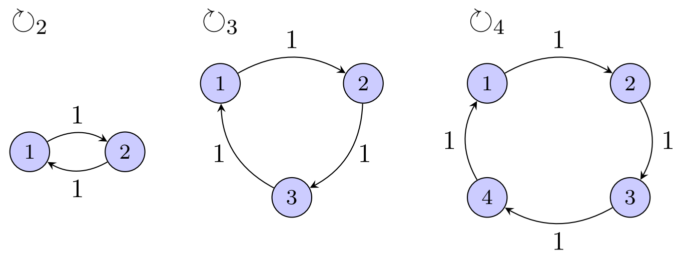


Figure 5.6: First three members of the $\circlearrowright_t$ family of networks. They represent the corresponding cyclic clustering methods $\mathcal{H}^{\circlearrowright_t}$. Undrawn edges correspond to undefined dissimilarities.

not required. Conversely, Theorem 11 states that encoding different degrees of required influence for different pairs of nodes within the representers is impossible if we want the resulting clustering method to be scale preserving.

The result in Theorem 11 is a complete characterization of hierarchical clustering methods that are admissible with respect to the Axioms of Value and Transformation (A1) and (A2) combined with fulfillment of the properties of scale preservation (P2) and representability (P3).

5.2.4 Cyclic clustering methods and algorithms

Let $\circlearrowright_t = (\{1, \ldots, t\}, A_t)$ denote a cycle network with t nodes such that the domain of the dissimilarity function $\mathrm{dom}(A_t) = \{(i, i+1)\}_{i=1}^{t-1} \cup (t, 1)$ contains all pairs of subsequent nodes plus the pair $(t, 1)$. Further require all pairs $(i, j) \in \mathrm{dom}(A_t)$ to have unit dissimilarities,

$$A_t(1, 2) = A_t(2, 3) = \ldots = A_t(t-1, t) = A_t(t, 1) = 1. \tag{5.39}$$

The first three elements of the class of cycle networks, $\circlearrowright_2$, $\circlearrowright_3$, and $\circlearrowright_4$, are illustrated in Fig. 5.6. In this section we study representable methods where the collections of representers contain cycle networks.

We first note that the method defined by a representer collection that contains a finite number of cycle networks is equivalent to the method defined by the singleton collection that contains as representer the longest of the cycles. Indeed, consider a finite collection $\Omega_{t_1,\ldots,t_n}$ of cyclic representers $\Omega_{t_1,\ldots,t_n} = \{\circlearrowright_{t_1}, \circlearrowright_{t_2}, \ldots, \circlearrowright_{t_n}\}$ and assume, without loss of generality, that $t_1 > t_2 > \ldots > t_n$. We can always find a dissimilarity reducing map from $\circlearrowright_{t_1}$ to $\circlearrowright_{t_i}$ for all $i = 2, \ldots, n$. For example, define the map $\phi_{t_1 \to t_i}$ from $\circlearrowright_{t_1}$ to $\circlearrowright_{t_i}$ as

$$\phi_{t_1 \to t_i}(j) = \min(j, t_i), \tag{5.40}$$

for $j \in \{1, \ldots, t_1\}$. The map $\phi_{t_1 \to t_i}$ is dissimilarity reducing since t_i of the unit dissimilarities in $\circlearrowright_{t_1}$ are mapped to unit dissimilarities in $\circlearrowright_{t_i}$ whereas the rest of the unit dissimilarities

$$
\begin{tikzcd}
\circlearrowleft_{t_1} \arrow[d, "\phi_{t_1 \to t_i}"'] \arrow[dr, dashed, "\phi \circ \phi_{t_1 \to t_i}"] & \\
\circlearrowleft_{t_i} \arrow[r, "\phi"'] & N
\end{tikzcd}
$$

Figure 5.7: Given a dissimilarity reducing map ϕ from $\circlearrowleft_{t_i}$ to N, one can always find a dissimilarity reducing map from $\circlearrowleft_{t_1}$ to N.

in $\circlearrowleft_{t_1}$ are mapped to null dissimilarities in $\circlearrowleft_{t_i}$. Thus, given a dissimilarity reducing map ϕ between $\circlearrowleft_{t_i}$ and an arbitrary network N, we may construct the dissimilarity reducing map $\phi \circ \phi_{t_1 \to t_i}$ from $\circlearrowleft_{t_1}$ to N; see Fig. 5.7. Moreover, $\phi \circ \phi_{t_1 \to t_i}$ has the same image as ϕ since $\phi_{t_1 \to t_i}$ is surjective by construction [cf. (5.40)]. The map $\phi_{t_1 \to t_i}$ being dissimilarity reducing ensures that $L(\phi \circ \phi_{t_1 \to t_i}; \circlearrowleft_{t_1}, N) \leq L(\phi; \circlearrowleft_{t_i}, N)$, which, by (5.20) implies that $\lambda_X^{\circlearrowleft_{t_1}}(x, x') \leq \lambda_X^{\circlearrowleft_{t_i}}(x, x')$ and from (5.23) we conclude that $\lambda_X^{\Omega_{t_1, \ldots t_n}}(x, x') = \lambda_X^{\circlearrowleft_{t_1}}(x, x')$. This means that the method represented by $\Omega_{t_1, \ldots, t_n}$ is equivalent to the method represented by the longest cycle $\circlearrowleft_{t_1}$.

Therefore, any method defined by a finite collection of cycle representers is equivalent to a method that is defined by a single cycle representer. Consider then the singleton collections $\{\circlearrowleft_t\}$ and denote the corresponding method as $\mathcal{H}^{\circlearrowleft_t} := \mathcal{H}^{\{\circlearrowleft_t\}}$. The method $\mathcal{H}^{\circlearrowleft_t}$ is referred to as the tth cyclic method. Cyclic methods $\mathcal{H}^{\circlearrowleft_t}$ for all $t \geq 2$ are admissible and scale preserving as shown in the following corollary of Theorem 11.

Corollary 3 *Cyclic methods, defined as representable methods $\mathcal{H}^{\circlearrowleft_t}$ associated with the cycle networks $\circlearrowleft_t = (\{1, \ldots, t\}, A_t)$ having dissimilarities as in* (5.39) *satisfy axioms (A1) and (A2) and the Scale Preservation Property (P2).*

Proof: Since networks $\circlearrowleft_t$ are strongly connected and structure representers, the hypotheses of Theorem 11 are satisfied. ■

The first cyclic method $\mathcal{H}^{\circlearrowleft_2}$ was used to introduce the concept of representable clustering in (5.15)-(5.17) and shown to coincide with the reciprocal clustering method $\mathcal{H}^{\mathrm{R}}$ in (5.18). Interpreting $\circlearrowleft_2$ as a basic cluster unit we can then think of reciprocal clustering $\mathcal{H}^{\mathrm{R}} \equiv \mathcal{H}^{\circlearrowleft_2}$ as a method that allows propagation of influence through cycles that contain at most two nodes. Likewise, the method $\mathcal{H}^{\circlearrowleft_3}$ represented by the cycle network $\circlearrowleft_3$ can be interpreted as a method that allows propagation of influence through cycles that contain at most three nodes. To see the consistency of this interpretation consider the application of $\mathcal{H}^{\circlearrowleft_3}$ to determine the ultrametric $u_X^{\circlearrowleft_3}(x, x')$ between points x and x' of the network $N = (X, A_X)$ shown in Fig. 5.8. In the figure, undrawn edges have dissimilarities greater than 5. Due to the scarcity of bidirectional paths linking x and x', a quick inspection reveals that $u_X^{\circlearrowleft_2}(x, x') = 4$

since the minimizing bidirectional path is given by $[x, x_2, x_3, x_4, x']$. In order to compute $u_X^{\circlearrowright_3}(x, x')$ as defined by (5.24) focus on the minimizing path $P_{xx'} = [x, x_2, x_3, x']$. Consider the map ϕ_{x,x_2} from $\circlearrowright_3$ to N such that $\phi_{x,x_2}(1) = x$, $\phi_{x,x_2}(2) = x_2$ and $\phi_{x,x_2}(3) = x_1$. By computing (5.19), we have that $L(\phi_{x,x_2}; \circlearrowright_3, N) = 2$. Moreover, 2 is the minimum multiple that allows the construction of a dissimilarity reducing map that contains in its image the nodes x and x_2. From (5.20), we then have that $\lambda_X^{\circlearrowright_3}(x, x_2) = 2$. Similarly, we can construct maps ϕ_{x_2,x_3} and $\phi_{x_3,x'}$ containing in their images the second and the third pair of consecutive nodes in $P_{xx'}$ respectively. The map ϕ_{x_2,x_3} goes from $\circlearrowright_3$ to N with $\phi_{x_2,x_3}(1) = \phi_{x_2,x_3}(2) = x_2$ and $\phi_{x_2,x_3}(3) = x_3$. The map $\phi_{x_3,x'}$ goes from $\circlearrowright_3$ to N with $\phi_{x_3,x'}(1) = x_3$, $\phi_{x_3,x'}(2) = x_4$ and $\phi_{x_3,x'}(3) = x'$. By computing the corresponding Lipschitz constants as done for ϕ_{x,x_2}, we have that $\lambda_X^{\circlearrowright_3}(x_2, x_3) = 3$ and $\lambda_X^{\circlearrowright_3}(x_3, x') = 1$. From (5.24) the ultrametric value is the maximum of these three multiples, i.e.,

$$u_X^{\circlearrowright_3}(x, x') = \max[\lambda_X^{\circlearrowright_3}(x, x_2), \lambda_X^{\circlearrowright_3}(x_3, x'), \lambda_X^{\circlearrowright_3}(x_2, x_3)] = 3. \tag{5.41}$$

Moving on to the fourth cyclic method consider the ultrametric $u_X^{\circlearrowright_4}(x, x')$ generated by the method $\mathcal{H}^{\circlearrowright_4}$ where influence cycles of up to 4 nodes are allowed. Focus on the same minimizing path as in the previous case $P_{xx'} = [x, x_2, x_3, x']$ and join the pairs x, x_2 and x_3, x' using the same maps as in the case for $\mathcal{H}^{\circlearrowright_3}$, i.e. $\lambda_X^{\circlearrowright_4}(x, x_2) = 2$ and $\lambda_X^{\circlearrowright_4}(x_3, x') = 1$. These maps exist because for every dissimilarity reducing map from $\circlearrowright_3$ to N we may construct a dissimilarity reducing map from $\circlearrowright_4$ to N using (5.40). However, in order to join x_2 and x_3 in $P_{xx'}$ we replace the mapping to the subnetwork formed by x_2 and x_3 by a map from $\circlearrowright_4$ to the four node subnetwork formed by x_2, x_1, x_4, and x_3. In this case, we obtain $\lambda_X^{\circlearrowright_4}(x_2, x_3) = 2$ and the ultrametric value becomes the maximum of the three multiples,

$$u_X^{\circlearrowright_4}(x, x') = \max[\lambda_X^{\circlearrowright_4}(x, x_2), \lambda_X^{\circlearrowright_4}(x_3, x'), \lambda_X^{\circlearrowright_4}(x_2, x_3)] = 2. \tag{5.42}$$

Observe that the third and fourth cyclic methods yield ultrametric distances smaller than the reciprocal ultrametric distance. This is not only consistent with Theorem 4 but also indicative of the status of these methods as relaxations of the condition of direct mutual influence. As we keep allowing for increasingly long cycles of influence the question arises of whether we end up recovering nonreciprocal clustering. This is not true in general for any $\circlearrowright_t$ where t is finite. However, if we define $\mathcal{C}_\infty$ as the following infinite collection of representers

$$\mathcal{C}_\infty := \{\circlearrowright_t\}_{t=1}^{\infty}, \tag{5.43}$$

we can show that the method $\mathcal{H}^{\mathcal{C}_\infty}$ is equivalent to the nonreciprocal clustering method $\mathcal{H}^{\mathrm{NR}}$.

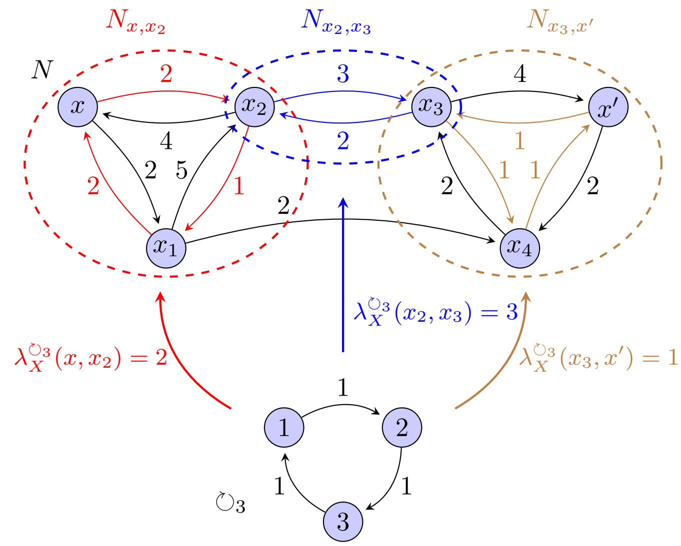


Figure 5.8: Computation of $u_X^{\circlearrowright_3}$. The minimizing path corresponds to $P_{xx'} = [x, x_2, x_3, x']$. Three dissimilarity reducing maps are constructed from multiples of $\circlearrowright_3$ to N. The images of the maps, marked by dashed ellipses, contain pairs of consecutive nodes in $P_{xx'}$. The ultrametric value corresponds to the maximum multiple of $\circlearrowright_3$, i.e. $u_X^{\circlearrowright_3}(x, x') = \max(2, 3, 1) = 3$.

Proposition 18 *The clustering method $\mathcal{H}^{\mathcal{C}_\infty}$ represented by the family of all cycle networks $\mathcal{C}_\infty$ defined in* (5.43) *is equivalent to the nonreciprocal clustering method $\mathcal{H}^{\text{NR}}$ with output ultrametrics as defined in* (3.8).

Proof: In order to show the equivalence $\mathcal{H}^{\mathcal{C}_\infty} \equiv \mathcal{H}^{\text{NR}}$, we have to show that the outputs $(X, u_X^{\mathcal{C}_\infty}) = \mathcal{H}^{\mathcal{C}_\infty}(N)$ and $(X, u_X^{\text{NR}}) = \mathcal{H}^{\text{NR}}(N)$ coincide for every network $N = (X, A_X)$. From Theorem 11 we know that $\mathcal{H}^{\mathcal{C}_\infty}$ is an admissible method since it is represented by a collection of strongly connected structure representers, thus by Theorem 4, we have that

$$u_X^{\text{NR}}(x, x') \leq u_X^{\mathcal{C}_\infty}(x, x'), \tag{5.44}$$

for arbitrary nodes x, x' in any network $N = (X, A_X)$.

Given a network $N = (X, A_X)$, pick any pair of nodes x and x' and define $\delta := u_X^{\text{NR}}(x, x')$ as the nonreciprocal ultrametric between these nodes. From definition (3.8), this implies that we can find a path $P_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ from x to x' and a path $P_{x'x} = [x' = x_0, x_1, ..., x_{l'} = x]$ in the opposite direction both paths of cost not greater than δ. Thus, the loop $P_{xx} = P_{xx'} \uplus P_{x'x}$ generated by the concatenation of the aforementioned paths has a cost not exceeding δ and it contains $l + l'$ nodes. Consequently, we may construct a map ϕ from $\circlearrowright_{l+l'} \in \mathcal{C}_\infty$ to N mapping the nodes in the cycle of $\delta * \circlearrowright_{l+l'}$ to the loop P_{xx}. From (5.20) and (5.24), this implies that

$$u_X^{\mathcal{C}_\infty}(x, x') \leq \delta = u_X^{\text{NR}}(x, x'). \tag{5.45}$$

Combining (5.44) and (5.45) we obtain that

$$u_X^{\mathcal{C}_\infty}(x, x') = u_X^{\mathrm{NR}}(x, x'), \tag{5.46}$$

for every pair of nodes x, x' in any network $N = (X, A_X)$, as wanted. ■

Proposition 18 provides a generative reformulation of nonreciprocal clustering. Moreover, it can be shown that any method represented by a collection of countably infinitely many distinct cycle representers is equivalent to $\mathcal{H}^{\mathcal{C}_\infty}$ as we show next.

Corollary 4 *Given any collection Ω of countably infinitely many distinct cycle representers, the represented method $\mathcal{H}^{\Omega} \equiv \mathcal{H}^{\mathcal{C}_\infty}$ where $\mathcal{C}_\infty$ is defined in* (5.43).

Proof: An analogous proof to the one of Proposition 18 can be done to show $\mathcal{H}^{\Omega} \equiv \mathcal{H}^{\mathrm{NR}}$. Combining this with the result in Proposition 18, it follows that $\mathcal{H}^{\Omega} \equiv \mathcal{H}^{\mathcal{C}_\infty}$. ■

Combining the result in Corollary 4 with the fact that a method represented by a finite collection of cycles is equivalent to the method represented by the longest cycle in the collection, it follows that by considering methods $\mathcal{H}^{\circlearrowright_t}$ for finite t and method $\mathcal{H}^{\mathcal{C}_\infty}$ we are considering every method that can be represented by a countable collection of cyclic representers.

As is intended of representable clustering methods, the reformulation in Proposition 18 expresses nonreciprocal clustering through the consideration of particular cases, namely cycles of arbitrary length. This reformulation uncovers the drawback of nonreciprocal clustering – propagating influence through cycles of arbitrary length is perhaps unrealistic – but also offers alternative formulations that mitigate this limitation – restrict the propagation of influence to cycles of certain length. In that sense, cyclic methods of length t can be interpreted as a tightening of nonreciprocal clustering. This interpretation is complementary of their interpretation as relaxations of reciprocal clustering that we discussed above. Given this dual interpretation, cyclic clustering methods are of practical importance.

Algorithms for the computation of the output ultrametrics associated with cyclic methods follow from matrix operations in the dioid algebra $(\bar{\mathbb{R}}_+, \min, \max)$ defined in Section 3.5. Explicit expressions are given in the following proposition.

Proposition 19 *Consider a given network $N = (X, A_X)$ with n nodes. Denote by $u_X^{\circlearrowright_t}$ the t-th cyclic ultrametric generated by the method $\mathcal{H}^{\circlearrowright_t}$ represented by the cycle network $\circlearrowright_t = (\{1, \ldots, t\}, A_t)$ with dissimilarities A_t as in* (5.39)*. Then, we can compute $u_X^{\circlearrowright_t}$ as*

$$u_X^{\circlearrowright_t} = \left[\bigoplus_{k=1}^{t-1} \max\left(A_X^k, (A_X^T)^{t-k} \right) \right]^{n-1}, \tag{5.47}$$

where the matrix powers are computed in the dioid algebra $(\bar{\mathbb{R}}_+, \min, \max)$ *as defined in* (3.71)*. Equivalently, the expression in* (5.47) *can be simplified to*

$$u_X^{\circlearrowright_t} = \left(\max\left(A_X, (A_X^T)^{t-1}\right)\right)^{n-1}. \tag{5.48}$$

Proof: We begin by showing validity of (5.47). Notice that if we show that

$$\left[\bigoplus_{k=1}^{t-1} \max\left(A_X^k, (A_X^T)^{t-k}\right)\right]_{ij} = \lambda_X^{\circlearrowright_t}(x_i, x_j), \tag{5.49}$$

then we are done since the outmost $(n-1)$ dioid power in (5.47) corresponds to computing single linkage [cf. (3.91)] and, thus, (5.35) completes the proof.

Recall that A_X^k contains the minimum path cost of paths of length at most k nodes, i.e.

$$\left[A_X^k\right]_{ij} = \min_{P_{x_i x_j}^k} \max_{m | x_m \in P_{x_i x_j}^k} A_X(x_m, x_{m+1}), \tag{5.50}$$

and similarly,

$$\left[(A_X^T)^{t-k}\right]_{ij} = \min_{P_{x_j x_i}^{t-k}} \max_{m | x_m \in P_{x_j x_i}^{t-k}} A_X(x_m, x_{m+1}). \tag{5.51}$$

Hence, the maximum of (5.50) and (5.51), i.e. $\max([A_X^k]_{ij}, [(A_X^T)^{t-k}]_{ij})$, gives us the minimum cost of a loop containing x_i and x_j in which the path from x_i to x_j has at most k nodes and the path in the opposite direction has at most $t-k$ nodes. Further, notice that

$$\max\left(\left[A_X^k\right]_{ij}, \left[(A_X^T)^{t-k}\right]_{ij}\right) = \left[\max\left(A_X^k, (A_X^T)^{t-k}\right)\right]_{ij}. \tag{5.52}$$

Finally, by minimizing over all possible k, we find the minimum cost of every loop with at most t nodes that contains x_i and x_j. Moreover,

$$\begin{aligned}
\min_k \left[\max\left(A_X^k, (A_X^T)^{t-k}\right)\right]_{ij} &= \bigoplus_{k=1}^{t-1} \left[\max\left(A_X^k, (A_X^T)^{t-k}\right)\right]_{ij} \\
&= \left[\bigoplus_{k=1}^{t-1} \max\left(A_X^k, (A_X^T)^{t-k}\right)\right]_{ij}.
\end{aligned} \tag{5.53}$$

Hence, we know that the left hand side of (5.49) contains the minimum cost of a loop of at most t nodes containing nodes x_i and x_j and this is exactly the minimum multiple for which we should multiply the cycle with unit dissimilarities $\circlearrowright_t$ such that a dissimilarity reducing map containing x_i and x_j in its image can be formed, which is the definition of $\lambda_X^{\circlearrowright_t}(x_i, x_j)$, showing equality (5.49) and the validity of (5.47).

In order to show (5.48), first observe that the difference with (5.47) is that instead of minimizing for every k – recall that $\oplus$ represents minimization – we only consider the case $k=1$. Thus, the right hand side of (5.48) must be greater than or equal to (5.47). Consequently, if we show that the every element of the right hand side matrix in (5.48) is not greater than its corresponding optimal multiple $\lambda_X^{\circlearrowright_t}(x_i,x_j)$, we are done. By the argument preceding (5.52) we know that $[\max\big(A_X,(A_X^T)^{t-1}\big)]_{ij}$ contains the minimum cost of a loop containing x_i and x_j where the forward path from x_i to x_j consists of only one link and the path in the opposite direction contains at most $t-1$ nodes. Suppose that the minimum cost loop of at most t nodes containing x_i and x_j is formed by the concatenation of $P_{x_ix_j}=[x_i=x_{i_0},x_{i_1},\dots,x_{i_l}=x_j]$ and $P_{x_jx_i}=[x_j=x_{i'_0},x_{i'_1},\dots,x_{i'_{l'}}=x_i]$ and its cost is, by definition, $\lambda_X^{\circlearrowright_t}(x_i,x_j)$. Focus on consecutive pairs of nodes in path $P_{x_ix_j}$. It must be that

$$\big[\max\big(A_X,(A_X^T)^{t-1}\big)\big]_{i_m\, i_{m+1}} \leq \lambda_X^{\circlearrowright_t}(x_i,x_j), \tag{5.54}$$

for $m=0,\dots,l-1$. To see why (5.54) holds, note that the same concatenated loop $P_{x_ix_j}\uplus P_{x_jx_i}$ contains nodes x_{i_m} and $x_{i_{m+1}}$. Moreover, for these two nodes, the aforementioned loop is formed by a forward path from x_{i_m} to $x_{i_{m+1}}$ of just one link and thus its cost must be stored in $\max\big(A_X,(A_X^T)^{t-1}\big)$. Since (5.54) is true for every m, once we apply the $(n-1)$ dioid power in (5.48), we are assured that the strong triangle inequality is satisfied. Hence,

$$\Big[\Big(\max\big(A_X,(A_X^T)^{t-1}\big)\Big)^{n-1}\Big]_{ij} \leq \max_m \big[\max\big(A_X,(A_X^T)^{t-1}\big)\big]_{i_m\, i_{m+1}} \leq \lambda_X^{\circlearrowright_t}(x_i,x_j), \tag{5.55}$$

completing the proof. ∎

Observe that when $t=2$ (5.47) and (5.48) reduce to (3.76). This is as it ought to be given the equivalence $\mathcal{H}^{\mathrm{R}}\equiv\mathcal{H}^{\circlearrowright_2}$ between the reciprocal and the first cyclic method. The reduction of (5.48) to (3.77) which corresponds to the equivalence $\mathcal{H}^{\mathrm{NR}}\equiv\mathcal{H}^{\mathcal{C}_\infty}$ is not as immediate but follows from recalling that in the dioid algebra $\lim_{t\to\infty}(A_X^T)^{t-1}=(A_X^T)^{n-1}$ followed by simple algebraic manipulations. Further observe that the expression in (5.48) is more efficient than (5.47) in terms of number of operations and memory requirements. Implementation of (5.47) requires computing and storing the dioid matrix powers A_X^k and $(A_X^T)^k$ for all $k\leq t-1$ to compute the minimum indicated by the $\oplus$ operation. Implementation of (5.48) requires the computation of $(A_X^T)^{t-1}$ only. The latter expression is therefore preferable for implementation.

Remark 14 Its algorithmic handicap notwithstanding, we present (5.47) as an illustration of Proposition 17 that constructs representable methods as the composition of a symmetrizing operation and single linkage clustering. Indeed, observe that the operation

$\bigoplus_{k=1}^{t-1} \max\left(A_X^k, (A_X^T)^{t-k}\right)$ in (5.47) outputs a symmetric matrix because for any l the terms $\max\left(A_X^l, (A_X^T)^{t-l}\right)$ and its transpose $\max\left(A_X^{t-l}, (A_X^T)^l\right)$ are both part of the dioid sum as it follows from substituting $k = l$ and $k = t - l$ in (5.47). We can then define the symmetrization operation $\Lambda^{\circlearrowright_t} : \tilde{\mathcal{N}} \to \mathcal{N}$ as the one that transforms the possibly asymmetric network $N = (X, A_X)$ into the symmetric network $\Lambda^{\circlearrowright_t}(X, A_X)$ defined as

$$\Lambda^{\circlearrowright_t}(X, A_X) = \left(X, \bigoplus_{k=1}^{t-1} \max\left(A_X^k, (A_X^T)^{t-k}\right)\right). \tag{5.56}$$

Further recalling that by (3.91) the $(n-1)$ dioid power of a symmetric network computes the corresponding single linkage ultrametric it follows that (5.47) is equivalent to

$$(X, u_X^{\circlearrowright_t}) = \mathcal{H}^{\mathrm{SL}} \circ \Lambda^{\circlearrowright_t}(X, A_X), \tag{5.57}$$

which has the form in (5.35) of Proposition 17. The operator $\Lambda^{\circlearrowright_t}$ symmetrizes possibly asymmetric network (X, A_X), which is rendered into the ultrametric $(X, u_X^{\circlearrowright_t})$ by the single linkage operator $\mathcal{H}^{\mathrm{SL}}$, [cf. Fig. 5.5]. Single linkage is the only admissible operator for this latter mapping as shown in Corollary 2.

The intuition behind (5.56) is the following: the matrix A_X^k stores the cost of the optimal forward paths of length at most k nodes and $(A_X^T)^{t-k}$ stores the cost of the optimal backward paths of length at most $t - k$ nodes. Therefore, the componentwise maximum between these two matrices, $\max(A_X^k, (A_X^T)^{t-k})$, corresponds to the optimal cost of cycles of length at most t nodes where the forward paths contain no more than k nodes. We then calculate the componentwise minimum over all k through the $\oplus$ operation in (5.56) to obtain the optimal cost of cycles of length at most t nodes for every length k of forward paths. Consequently, $\Lambda^{\circlearrowright_t}(X, A_X)$ is a symmetric network where the dissimilarity between x and x' coincides with the minimum cost of a loop of length at most t nodes that contains nodes x and x'. The single linkage operation in (5.57) transforms this network of minimum cost loops with at most t nodes into the ultrametric $u_X^{\circlearrowright_t}$.

Remark 15 The family of cyclic methods includes reciprocal clustering $\mathcal{H}^{\mathrm{R}} \equiv \mathcal{H}^{\circlearrowright_2}$ and nonreciprocal clustering $\mathcal{H}^{\mathrm{NR}} \equiv \mathcal{H}^{\mathcal{C}_\infty}$ as its extreme methods. The semi-reciprocal family of clustering methods introduced in Section 3.3 also harbored these methods as extreme. Despite this similarity, semi-reciprocal and cyclic methods are not equivalent in general. Semi-reciprocal methods are a family of intermediate methods from an *algorithmic* point of view whereas cyclic methods are an intermediate family from a *generative* perspective. In terms of influence propagation, the semi-reciprocal method $\mathcal{H}^{\mathrm{SR}(t)}$ allows influence to propagate from node x to x' by concatenating intermediate *paths* $P_{x_i x_{i+1}}$ and $P_{x_{i+1} x_i}$ so that each of the two paths contains no more than t nodes. In contrast, the cyclic method $\mathcal{H}^{\circlearrowleft_t}$ allows influence to propagate from node x to x' by concatenating intermediate *loops* of at most t nodes containing x and x'. I.e., for each of the intermediate paths $P_{x_i x_{i+1}}$ and $P_{x_{i+1} x_i}$ there is an upper bound on the total number of nodes in both paths but no bound on the number of nodes in each of them.

5.3 Excisiveness

The outcome of applying a hierarchical clustering method $\mathcal{H}$ to network $(X, A_X) \in \tilde{\mathcal{N}}$ is a finite ultrametric space $(X, u_X) = \mathcal{H}(X, A_X) \in \mathcal{U} \subset \tilde{\mathcal{N}}$. Since finite ultrametric spaces are particular cases of networks we can study the result of repeated applications of a clustering method $\mathcal{H}$. We expect that clustering a network that has been already clustered should not alter the outcome. This is formally stated as the requirement that the map $\mathcal{H} : \tilde{\mathcal{N}} \to \mathcal{U}$ be idempotent, i.e., that for every network $N = (X, A_X) \in \tilde{\mathcal{N}}$ we have

$$\mathcal{H}\big(\mathcal{H}(X, A_X)\big) = \mathcal{H}(X, A_X). \tag{5.58}$$

Alternatively, (5.58) is true if when we restrict the map $\mathcal{H}$ to the space of finite ultrametric spaces $\mathcal{U}$, the method is equivalent to an identity mapping,

$$\mathcal{H}(X, u_X) = (X, u_X), \text{ for all } (X, u_X) \in \mathcal{U}. \tag{5.59}$$

Idempotency is not a stringent requirement. In fact, any method that satisfies the Axioms of Value and Transformation is idempotent as we show in the following proposition.

Proposition 20 *Every admissible clustering method $\mathcal{H}$ is idempotent in the sense of* (5.58).

Proof: We prove that any admissible method $\mathcal{H}$ is idempotent by showing that is satisfies (5.59). Consider the application of admissible methods $\mathcal{H}$ to the ultrametric network $U_X = (X, u_X) \in \mathcal{U}$. The ultrametric network is, in particular, symmetric and it thus follows from Corollary 2 that applying $\mathcal{H}$ to U_X is equivalent to applying single linkage $\mathcal{H}^{\text{SL}}$ to U_X. Denoting $(X, u'_X) = \mathcal{H}(X, u_X)$ and $(X, u^{\text{SL}}_X) = \mathcal{H}^{\text{SL}}(X, u_X)$ the corresponding outcomes and invoking the definition of the single linkage ultrametric in (2.15) we can write for all $x, x' \in X$,

$$u'_X(x, x') \;=\; u^{\text{SL}}_X(x, x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} u_X(x_i, x_{i+1}). \tag{5.60}$$

Given a path $P_{xx'}$ and using the fact that u_X is an ultrametric it follows from the strong triangle inequality in (2.12) that $u_X(x, x') \leq \max_{i | x_i \in P_{xx'}} u_X(x_i, x_{i+1})$. Substituting this observation, which is valid for all paths $P_{xx'}$, into (5.60) yields

$$u'_X(x,x') \geq u_X(x,x'), \tag{5.61}$$

for all $x, x' \in X$. But considering the particular path $P_{xx'} = [x, x']$ whose cost is $u_X(x,x')$ it follows from (5.60) that it must be

$$u'_X(x,x') \leq u_X(x,x'), \tag{5.62}$$

for all $x, x' \in X$. Comparing (5.61) and (5.62) it must be $u'_X(x,x') = u_X(x,x')$ for all $x, x' \in X$, showing that the admissible method $\mathcal{H}$ is idempotent. ■

Since, according to Proposition 20, idempotency is implied by (A1)-(A2) it cannot be used as an additional desirable feature. However, idempotency is not the only way to formalize the notion that subsequent re-clusterings with the same method should not alter the outcome of its first application. A similar more stringent condition is the property of *excisiveness*, proposed in [11] for non-hierarchical clustering. To define excisiveness consider a clustering method $\mathcal{H}$ and a given network $N = (X, A_X)$. Denote by $(X, u_X) = \mathcal{H}(X, A_X)$ the ultrametric output, as D_X the dendrogram equivalent to (X, u_X) by Theorem 1, and recall that for given resolution δ the dendrogram's partition is denoted by $D_X(\delta) = \{B_1(\delta), \ldots, B_{J(\delta)}(\delta)\}$. Consider then the subnetworks N_i^δ associated with each block $B_i(\delta)$ of $D_X(\delta)$ defined as

$$N_i^\delta := \Big(B_i(\delta),\ A_X\big|_{B_i(\delta)\times B_i(\delta)}\Big), \tag{5.63}$$

where $A_X|_{B_i(\delta)\times B_i(\delta)}$ denotes the restriction of A_X to the product space $B_i(\delta) \times B_i(\delta)$. In terms of ultrametrics, networks N_i^δ are such that their node space $B_i(\delta)$ satisfy

$$\begin{aligned} u_X(x,x') &\leq \delta, \quad \text{for all } x, x' \in B_i(\delta), \\ u_X(x,x'') &> \delta, \quad \text{for all } x \in B_i(\delta),\ x'' \notin B_i(\delta). \end{aligned} \tag{5.64}$$

Upon extraction of the subnetwork N_i^δ we can compare, on the one hand, the result of restricting the original clustering to the set $B_i(\delta)$ with, on the other hand, the outcome corresponding to applying the clustering method $\mathcal{H}$ to N_i^δ. If the two intervening ultrametrics are the same, then we say the method $\mathcal{H}$ is excisive as we formally define next.

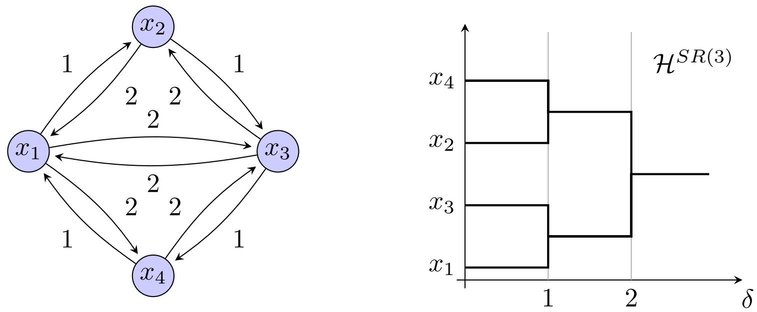


Figure 5.9: A network and its dendrogram when the semi-reciprocal clustering method $\mathcal{H}^{\mathrm{SR}(3)}$ is applied. Edges that are not drawn between x_2 and x_4 take values greater than 2. The subnetwork formed by x_1 and x_3 shows that $\mathcal{H}^{\mathrm{SR}(3)}$ is not excisive since these nodes merge at resolution 1 in the dendrogram but at resolution 2 in their subnetwork.

(P4) Excisiveness. Consider a hierarchical clustering method $\mathcal{H}$, an arbitrary network $N = (X, A_X)$ with ultrametric output $(X, u_X) = \mathcal{H}(X, A_X)$, and the corresponding subnetworks N_i^δ defined in (5.63). We say the method $\mathcal{H}$ is *excisive* if and only if for all subnetworks N_i^δ at all resolutions $\delta > 0$ it holds that

$$\mathcal{H}\left(N_i^\delta\right) = \left(B_i(\delta),\ u_X\big|_{B_i(\delta)\times B_i(\delta)}\right). \tag{5.65}$$

The appeal of excisive methods is that they exhibit local consistency in the following sense. For a given resolution δ, when we cluster the subnetworks as defined in (5.63), we obtain a dendrogram on the node space $B_i(\delta)$ for every i. Excisiveness ensures that when clustering the whole network and cutting the output dendrogram at resolution δ, the branches obtained coincide with the previously computed dendrograms for every subnetwork.

Contrary to idempotency, excisiveness is not implied by admissibility with respect to (A1) and (A2). Consider, e.g., the network in Fig. 5.9 and its dendrogram corresponding to the semi-reciprocal clustering method $\mathcal{H}^{\mathrm{SR}(3)}$ as defined in (3.52), which is admissible with respect to (A1) and (A2). For a resolution $\delta = 1.5$, focus on the subnetwork $N_1^{1.5} = (\{x_1, x_3\}, A_{\{1,3\}})$ with $A_{\{1,3\}}(x_1, x_3) = A_{\{1,3\}}(x_3, x_1) = 2$ and denote by $(X, u_{\{1,3\}}^{\mathrm{SR}(3)}) = \mathcal{H}^{\mathrm{SR}(3)}(N_1^{1.5})$ the output of applying the method $\mathcal{H}^{\mathrm{SR}(3)}$ to the subnetwork $N_1^{1.5}$. Since $N_1^{1.5}$ is a two-node network, the Axiom of Value (A1) implies that $u_{\{1,3\}}^{\mathrm{SR}(3)}(x_1, x_3) = \max(2, 2) = 2$. However, from the dendrogram in Figure 5.9 we see that $u_X^{\mathrm{SR}(3)}(x_1, x_3) = 1$. Hence, the method $\mathcal{H}^{\mathrm{SR}(3)}$ is not excisive.

However, excisiveness is not independent of all conditions introduced. The following theorem states a relationship between representable and excisive methods.

Theorem 12 *Given an admissible hierarchical clustering method $\mathcal{H}$, it is representable (P3) if and only if it is excisive (P4) and linear scale preserving (P2').*

Proof: Linear scale preservation is implied by the definition of representability. To see this, notice that the Lipschitz constants of arbitrary maps (5.19) satisfy

$$L(\phi; \omega, \alpha * N) = \alpha\, L(\phi; \omega, N), \tag{5.66}$$

for an arbitrary positive constant $\alpha > 0$. Consequently, the optimal multiple between any pair of nodes $\lambda_X^{\omega}(x, x')$ with $x, x' \in X$ as defined in (5.20) is multiplied by α when the network N is multiplied by α. Moreover, (5.24) implies that multiplying every optimal multiple by a constant is equivalent to multiplying the original ultrametric value by that same constant and linear scale preservation follows.

To show that representability implies excisiveness, consider a representable clustering method $\mathcal{H}^{\Omega}$, a network $N = (X, A_X)$, a resolution $\delta > 0$ and a subnetwork at such resolution $N_i^{\delta} = \Big(B_i(\delta),\ A_X\big|_{B_i(\delta) \times B_i(\delta)}\Big)$ as defined in (5.63). We want to show that (5.65) is true for the representable clustering method $\mathcal{H}^{\Omega}$. Denote by $(X, u_X^{\Omega}) = \mathcal{H}^{\Omega}(X, A_X)$ the output ultrametric when applying the method $\mathcal{H}^{\Omega}$ on network N and by $(X, u_{N_i^{\delta}}^{\Omega}) = \mathcal{H}^{\Omega}(N_i^{\delta})$ the output ultrametric when applying the method $\mathcal{H}^{\Omega}$ to the subnetwork N_i^{δ}. The identity map from N_i^{δ} to N, i.e. the map that takes $B_i(\delta)$ into the corresponding subset of nodes in X, is dissimilarity reducing. Since the method $\mathcal{H}^{\Omega}$ is assumed to be admissible, the Axiom of Transformation (A2) implies that

$$u_{N_i^{\delta}}^{\Omega}(x, x') \geq u_X^{\Omega}(x, x'), \tag{5.67}$$

for all $x, x' \in B_i(\delta)$. In order to show the opposite inequality, pick arbitrary nodes $x, x' \in B_i(\delta)$ and $x'' \notin B_i(\delta)$. From the definition of subnetwork (5.64), it must be that

$$u_X^{\Omega}(x, x') \leq \delta, \tag{5.68}$$

$$u_X^{\Omega}(x, x'') > \delta. \tag{5.69}$$

From the definition of representability, (5.68) implies that there exists a minimizing path $P_{xx'} = [x = x_0, x_1, ..., x_l = x']$ in definition (5.24) and a series of maps $\phi_{x_j, x_{j+1}}$ for all j determining the optimal multiples $\lambda_X^{\Omega}(x_j, x_{j+1}) \leq \delta$. However, (5.69) implies that $\mathrm{Im}(\phi_{x_j, x_{j+1}}) \subseteq B_i(\delta)$ for all j. Otherwise, if any node $x'' \notin B_i(\delta)$ is such that $x'' \in \mathrm{Im}(\phi_{x_j, x_{j+1}})$, we should have $u_X^{\Omega}(x, x'') \leq \delta$ contradicting (5.69). Hence, the minimizing path $P_{xx'}$ and the image of every optimal dissimilarity reducing map is contained in $B_i(\delta)$. This implies that

$$u_{N_i^{\delta}}^{\Omega}(x, x') \leq u_X^{\Omega}(x, x'), \tag{5.70}$$

for all $x, x' \in B_i(\delta)$. Combining (5.67) with (5.70) we obtain (5.65) completing the necessity proof.

To prove sufficiency, we must show that excisiveness and linear scale preservation imply representability. To do so, consider an arbitrary clustering method $\mathcal{H}$ which is excisive and linear scale preserving. We will construct a representable method $\mathcal{H}^\Omega$ such that $\mathcal{H} \equiv \mathcal{H}^\Omega$, i.e. $\mathcal{H}(N) = \mathcal{H}^\Omega(N)$ for every network N. This would show that $\mathcal{H}$ is represented by Ω and concludes the proof.

Denote by $(X, u_X) = \mathcal{H}(X, A_X)$ the output of the clustering method $\mathcal{H}$ when applied to a given network. Given $\mathcal{H}$, define the collection of representers Ω as follows:

$$\Omega = \Big\{ \omega \;\Big|\; \omega = \frac{1}{\max_{x,x' \in B_i(\delta)} u_X(x,x')} * N_i^\delta,\ |B_i(\delta)| > 1,\ \delta > 0 \Big\}, \tag{5.71}$$

for all possible resolutions $\delta > 0$ and $N_i^\delta := (B_i(\delta), A_X|_{B_i(\delta)\times B_i(\delta)})$ being a subnetwork of all possible networks $N = (X, A_X)$. In other words, we pick as representers the set of all possible subnetworks generated by the method $\mathcal{H}$, each of them scaled by the maximum ultrametric obtained in such subnetwork. Notice that from the definition of subnetwork (5.64) we have that

$$\max_{x,x' \in B_i(\delta)} u_X(x,x') \leq \delta, \tag{5.72}$$

which appears in the denominator of the definition (5.71) for every representer $\omega \in \Omega$.

We show equivalence of methods $\mathcal{H}$ and $\mathcal{H}^\Omega$ by showing that the ultrametric outputs coincide for every network. Pick an arbitrary network $N = (X, A_X)$ and two different nodes $x, x', \in X$ and define $\alpha := u_X(x,x')$. However, Ω was built considering all possible networks, including N. Therefore, there is a representer $\omega \in \Omega$ that corresponds to the subnetwork N_i^α at resolution α that contains x and x'. From (5.72), the map ϕ from $\alpha * \omega$ to N such that $\phi(x) = x$ is dissimilarity reducing and $x, x' \in \mathrm{Im}(\phi)$. From definition (5.20) this implies that $\lambda_X^\omega(x,x') \leq \alpha$. By substituting in (5.23) and further substitution in (5.24) we obtain that $u_X^\Omega(x,x') \leq \alpha$. Recalling that $\alpha = u_X(x,x')$ and that we chose the network N and the pair of nodes x, x' arbitrarily, we may conclude that $u_X^\Omega \leq u_X$, for every network $N = (X, A_X)$.

In order to show the other direction of the inequality, we must first observe that for every representer the ultrametric value given by $\mathcal{H}$ between any pair of nodes in the representer is upper bounded by 1. To see this, given a representer $\omega = (X_\omega, A_{X_\omega})$ associated with the subnetwork N_i^δ in (5.71) we have that

$$\begin{aligned} u_{X_\omega}(\tilde{x}, \tilde{x}') &= \frac{1}{\max_{x,x' \in B_i(\delta)} u_X(x,x')}\ u_{B_i(\delta)}(\tilde{x}, \tilde{x}') \\ &= \frac{1}{\max_{x,x' \in B_i(\delta)} u_X(x,x')}\ u_X|_{B_i(\delta)\times B_i(\delta)}(\tilde{x}, \tilde{x}') \leq 1, \end{aligned} \tag{5.73}$$

for all $\tilde{x}, \tilde{x}' \in X_\omega$. The first equality in (5.73) is implied by the definition of ω in (5.71) and linear scale preservation of $\mathcal{H}$. The second equality is derived from excisiveness of $\mathcal{H}$.

Pick an arbitrary network $N = (X, A_X)$ and a pair of nodes $x, x' \in X$ and define $\beta := u_X^\Omega(x, x')$. This means that there exists a minimizing path $P_{xx'} = [x' = x_0, x_1, ..., x_l = x']$ such that for every consecutive pair of nodes we can find a dissimilarity reducing map $\phi_{x_j,x_{j+1}}$ from $\beta * \omega_j$ to N for some representer $\omega_j \in \Omega$ such that $x_j, x_{j+1} \in \mathrm{Im}(\phi_{x_j,x_{j+1}})$. Focus on a particular pair of consecutive nodes x_j, x_{j+1} and denote by p_j, p_{j+1} their corresponding pre images on $\omega_j = (X_\omega, A_{X_\omega})$ under the map $\phi_{x_j,x_{j+1}}$. Without loss of generality, we can assume that $x_j \neq x_{j+1}$ for all j. The pre images need not be unique. Denote by $\beta * \omega_j = (X_\omega^\beta, A_{X_\omega^\beta})$ the β multiple of the representer ω_j. Since $\phi_{x_j,x_{j+1}}$ is a dissimilarity reducing map from $\beta * \omega_j$ to N, the Axiom of Transformation (A2) implies that

$$u_{X_\omega^\beta}(p_j, p_{j+1}) \geq u_X(x_j, x_{j+1}). \tag{5.74}$$

Moreover, we can assert that

$$u_{X_\omega^\beta}(p_j, p_{j+1}) = \beta\, u_{X_\omega}(p_j, p_{j+1}) \leq \beta, \tag{5.75}$$

where the equality is due to linear scale preservation and the inequality is justified by (5.73). From the combination of (5.74) and (5.75) we obtain that

$$u_X(x_j, x_{j+1}) \leq \beta. \tag{5.76}$$

Since (5.76) was found for an arbitrary pair of consecutive nodes in $P_{xx'}$, from the strong triangle inequality we have that

$$u_X(x, x') \leq \max_j u_X(x_j, x_{j+1}) \leq \beta. \tag{5.77}$$

Recalling that $\beta = u_X^\Omega(x, x')$ and that the network N was arbitrary, we can conclude that $u_X^\Omega \geq u_X$, for every network $N = (X, A_X)$. Combining this inequality with $u_X^\Omega \leq u_X$ the result follows. ■

The relation between representability and excisiveness stated in Theorem 12 originates from the fact that both concepts address the locality of clustering methods. Representability implies that the method can be interpreted as an extension of particular cases or representers. Excisiveness requires the clustering of local subnetworks to be consistent with the clustering of the entire network.

The importance of Theorem 12 resides in relating an implicit property of a method such as excisiveness with a generative model of clustering methods such as representability – as [11] does for non-hierarchical clustering on finite metric spaces. Thus, when designing a clustering method for a particular application, if excisiveness and linear scale preservation are desirable properties then Theorem 12 asserts that representability must be considered as a generative model. Furthermore, Theorem 12 facilitates the analysis of the clustering methods presented throughout this thesis. E.g., in Section 5.1 we showed through a counterexample that grafting methods (Section 3.3.1) are not linear scale preserving (P2'). Thus, from Theorem 12 we may conclude that grafting methods are not representable (P3). Similarly, from the counterexample in Fig. 5.9 we know that the semi-reciprocal clustering methods introduced in Section 3.3 are not excisive (P4). Hence, from Theorem 12 they are not representable (P3) either.

Theorems 11 and 12 may be combined to obtain the following fundamental corollary.

Corollary 5 *A hierarchical clustering method $\mathcal{H}$ satisfies axioms (A1)-(A2), scale preservation (P2) and excisiveness (P4) if and only if $\mathcal{H}$ is representable (P3) by a collection Ω of strongly connected, structure representers as defined in* (5.38).

Proof: To show the first implication, recall that scale preservation (P2) implies linear scale preservation (P2'). Thus, by Theorem 12, scale preservation (P2), excisiveness (P4) and (A2) imply that $\mathcal{H}$ is representable (P3). Consequently, Theorem 11 applies to show that $\mathcal{H}$ can be represented by a collection Ω of strongly connected structure representers. To show the converse implication, if a method can be represented by a collection Ω of strongly connected structure representers then Theorem 11 ensures admissibility and scale preservation whereas Theorem 12 guarantees excisiveness. ■

Corollary 5 is the most complete characterization presented in this first part of the thesis. It shows that there is a unique family of methods that satisfy the Axioms of Value (A1) and Transformation (A2) as well as the Properties of Scale Preservation (P2) and Excisiveness (P4). This unique family consists of representable methods that are represented by strongly connected structure representers as defined by (5.38). If these four properties are desired in a hierarchical clustering method, only methods in this family are admissible.

Excisiveness entails a tangible practical advantage when hierarchically clustering big data. Often in practical applications, one begins by performing a coarse clustering at an exploratory phase. Notice that the computational cost of obtaining this coarse partition, which corresponds to *one* particular resolution, is smaller than that of computing the whole dendrogram. After having done this and having identified blocks in the resulting partition that contain a relevant subset of the original data, one focuses on these blocks – via the subsequent application of the clustering method – in order to reveal the whole hierarchical structure of this subset of the data. It is evident that the computational cost of clustering a subset of the data is smaller than the cost of clustering the whole dataset and then restricting the output to the relevant data subset. However, an excisive method guarantees that the results obtained through both procedures are identical, thus, reducing computational effort with no loss of clustering information. A specific example of the aforementioned computational gain is presented next.

Example 1 (single linkage computation) Focus on the application of single linkage hierarchical clustering to a finite metric space of n points. Single linkage is an excisive clustering method as can be concluded by combining Corollary 5 with the fact that, for finite metric spaces, reciprocal and nonreciprocal clustering coincide with single linkage (cf. Theorem 4). Consider two different ways of computing the output dendrogram for a subspace of the aforementioned finite metric space. The first approach is to hierarchically cluster the whole finite metric space and then extract the relevant branch. The computational cost of single linkage is equivalent to that of finding a minimum spanning tree in an undirected graph which, for a complete graph, is of cost $O(n^2)$ [31]. The second approach consists of first obtaining the partition given by single linkage corresponding to *one* coarse resolution. This is equivalent to finding the connected components in a graph where only the edges of weight smaller than the resolution are present. Assuming that the average degree of each node in this graph is α, the computational cost of finding the connected components is $O(\max(n, n\,\alpha/2)) = O(n\,\alpha/2)$ as long as $\alpha \geq 2$ [38]. After this, we pick the subspace of interest and find its minimum spanning tree. Assuming that the subspace contains $\beta\,n$ nodes, the cost of finding the minimum spanning tree is $O(\beta^2 n^2)$. Consequently, the cost of the first approach is $O(n^2)$ whereas the cost of the second one is $O(n\,\alpha/2) + O(\beta^2 n^2)$. This entails an asymptotic reduction of order β^{-2}. Excisiveness ensures that the output of both approaches coincide, allowing us to follow the second – more efficient – approach.

5.4 Stability

The collection of all compact metric spaces modulo isometry becomes a metric space of its own when endowed with the Gromov-Hausdorff distance [9, Chapter 7.3]. This distance can be generalized to the space of networks $\tilde{\mathcal{N}}$ modulo a properly defined notion of isomorphism. For a given method $\mathcal{H}$ we can then ask the question of whether networks that are close to each other result in dendrograms that are also close to each other. The answer to this question is affirmative for semi-reciprocal methods – of which reciprocal and nonreciprocal methods are particular cases –, admissible scale preserving representable methods, and most other constructions introduced earlier, as we discuss in the following sections.

5.4.1 Gromov-Hausdorff distance for asymmetric networks

Relabeling the nodes of a given network $N_X = (X, A_X)$ results in a network $N_Y = (Y, A_Y)$ that is identical from the perspective of the dissimilarity relationships between nodes. To capture this notion formally, we say that N_X and N_Y are *isomorphic* whenever there exists a bijective map $\phi : X \to Y$ such that for all points $x, x' \in X$ we have

$$A_X(x, x') = A_Y(\phi(x), \phi(x')). \tag{5.78}$$

When networks N_X and N_Y are isomorphic we write $N_X \cong N_Y$. The space where all isomorphic networks are represented by a single point is called the space of networks modulo isomorphism and denoted as $\tilde{\mathcal{N}} \mod \cong$.

To motivate the definition of a distance on the space $\tilde{\mathcal{N}} \mod \cong$ of networks modulo isomorphism, we start considering networks N_X and N_Y with the same number of nodes and assume that a bijective transformation $\phi : X \to Y$ is given. It is then natural to define the *distortion* $\mathrm{dis}(\phi)$ of the map ϕ as

$$\mathrm{dis}(\phi) := \max_{x,x' \in X} \big| A_X(x, x') - A_Y(\phi(x), \phi(x')) \big|. \tag{5.79}$$

Since different maps $\phi : X \to Y$ are possible, we further focus on those maps ϕ that make the networks N_X and N_Y as similar as possible and define the distance d_∞ between networks N_X and N_Y with the same cardinality as

$$d_\infty(N_X, N_Y) := \frac{1}{2} \min_{\phi} \mathrm{dis}(\phi), \tag{5.80}$$

where the factor $1/2$ is added for consistency with the definition of the Gromov-Hausdorff distance for metric spaces [9, Chapter 7.3]. To generalize (5.80) to networks that may have different number of nodes we consider the notion of correspondence between node sets to take the role of the bijective transformation ϕ in (5.79) and (5.80). More specifically, for node sets X and Y consider subsets $R \subseteq X \times Y$ of the Cartesian product space $X \times Y$ with elements $(x, y) \in R$. The set R is a *correspondence* between X and Y if for all $x_0 \in X$ we have at least one element $(x_0, y) \in R$ whose first component is x_0, and for all $y_0 \in Y$ we have at least one element $(x, y_0) \in R$ whose second component is y_0. The distortion of the correspondence R is defined as

$$\mathrm{dis}(R) := \max_{(x,y),(x',y') \in R} \big| A_X(x, x') - A_Y(y, y') \big|. \tag{5.81}$$

In a correspondence R all the elements of X are paired with some point in Y and, conversely, all the elements of Y are paired with some point in X. We can then think of R as a mechanism to superimpose the node spaces on top of each other so that no points are orphaned in either X or Y. As we did in going from (5.79) to (5.80) we now define the distance between networks N_X and N_Y as the distortion associated with the correspondence

R that makes N_X and N_Y as close as possible,

$$d_{\tilde{\mathcal{N}}}(N_X, N_Y) := \frac{1}{2} \min_R \operatorname{dis}(R) = \frac{1}{2} \min_R \max_{(x,y),(x',y') \in R} \big| A_X(x, x') - A_Y(y, y') \big|. \tag{5.82}$$

Notice that (5.82) does not necessarily reduce to (5.80) when the networks have the same number of nodes. Since for networks N_X, N_Y with $|X| = |Y|$, correspondences are more general than bijective maps there may be a correspondence R that results in a distance $d_{\tilde{\mathcal{N}}}(N_X, N_Y)$ smaller than the distance $d_\infty(N_X, N_Y)$.

The definition in (5.82) is a verbatim generalization of the Gromov-Hausdorff distance in [9, Theorem 7.3.25] except that the dissimilarity functions A_X and A_Y are not restricted to be metrics. It is legitimate to ask whether the relaxation of this condition renders $d_{\tilde{\mathcal{N}}}(N_X, N_Y)$ in (5.82) an invalid metric. We prove in the following theorem that this is not the case since $d_{\tilde{\mathcal{N}}}(N_X, N_Y)$ becomes a legitimate metric in the space $\tilde{\mathcal{N}} \mod \cong$ of networks modulo isomorphism.

Theorem 13 *The function* $d_{\tilde{\mathcal{N}}} : \tilde{\mathcal{N}} \times \tilde{\mathcal{N}} \to \mathbb{R}_+$ *defined in* (5.82) *is a metric on the space* $\tilde{\mathcal{N}} \mod \cong$ *of networks modulo isomorphism. I.e., for all networks* $N_X, N_Y, N_Z \in \tilde{\mathcal{N}}$, $d_{\tilde{\mathcal{N}}}$ *satisfies the following properties:*

Nonnegativity: $d_{\tilde{\mathcal{N}}}(N_X, N_Y) \geq 0$.

Symmetry: $d_{\tilde{\mathcal{N}}}(N_X, N_Y) = d_{\tilde{\mathcal{N}}}(N_Y, N_X)$.

Identity: $d_{\tilde{\mathcal{N}}}(N_X, N_Y) = 0$ *if and only if* $N_X \cong N_Y$.

Triangle ineq.: $d_{\tilde{\mathcal{N}}}(N_X, N_Y) \leq d_{\tilde{\mathcal{N}}}(N_X, N_Z) + d_{\tilde{\mathcal{N}}}(N_Z, N_Y)$.

Proof: Proof of nonnegativity and symmetry statements: That the distance $d_{\tilde{\mathcal{N}}}(N_X, N_Y) \geq 0$ is nonnegative follows from the absolute value in the definition of (5.82). The symmetry $d_{\tilde{\mathcal{N}}}(N_X, N_Y) = d_{\tilde{\mathcal{N}}}(N_Y, N_X)$ follows because a correspondence $R \subseteq X \times Y$ with elements $r_i = (x_i, y_i)$ results in the same associations as the correspondence $S \subseteq Y \times X$ with elements $s_i = (y_i, x_i)$. This proves the first two statements. ■

Proof of identity statement: In order to show the identity statement, assume that N_X and N_Y are isomorphic and let $\phi : X \to Y$ be a bijection proving this isomorphism. Then, consider the particular correspondence $R_\phi = \{(x, \phi(x)),\ x \in X\}$. By construction, for all $x_0 \in X$ there is an element $r = (x_0, y) \in R_\phi$ and since ϕ is surjective – indeed, bijective – for all $y_0 \in Y$ there is an element $s = (x, y_0) \in R_\phi$. Thus, R_ϕ is a valid correspondence between X and Y, which, according to (5.78), satisfies

$$A_Y(y, y') = A_Y(\phi(x), \phi(x')) = A_X(x, x'), \tag{5.83}$$

for all $(x, y), (x', y') \in R_\phi$. Since R_ϕ is a particular correspondence while in definition (5.82) we minimize over all possible correspondences it must be that

$$d_{\tilde{\mathcal{N}}}(N_X, N_Y) \leq \frac{1}{2} \max_{(x,y),(x',y') \in R_\phi} |A_X(x, x') - A_Y(y, y')| = 0, \tag{5.84}$$

where the equality follows because $A_X(x, x') - A_Y(y, y') = 0$ for all $(x, y), (x', y') \in R_\phi$ by (5.83). Since we already argued that $d_{\tilde{\mathcal{N}}}(N_X, N_Y) \geq 0$ it must be that $d_{\tilde{\mathcal{N}}}(N_X, N_Y) = 0$ when the networks $N_X \cong N_Y$ are isomorphic.

We now argue that the converse is also true, i.e., $d_{\tilde{\mathcal{N}}}(N_X, N_Y) = 0$ implies that X and Y are isomorphic. If $d_{\tilde{\mathcal{N}}}(N_X, N_Y) = 0$ there is a correspondence R_0 such that $A_X(x, x') = A_Y(y, y')$ for all $(x, y), (x', y') \in R_0$. Define then the function $\phi : X \to Y$ that associates to x any value y among those that form a pair with x in the correspondence R_0,

$$\phi(x) = y_0 \in \left\{ y \,\middle|\, (x, y) \in R_0 \right\}. \tag{5.85}$$

Since R_0 is a correspondence the set $\left\{ y \,\middle|\, (x, y) \in R_0 \right\} \neq \emptyset$ is nonempty implying that (5.85) is defined for all $x \in X$. Moreover, since we know that $(x, \phi(x)) \in R_0$ we must have $A_X(x, x') = A_Y(\phi(x), \phi(x'))$ for all x, x'. From this observation it follows that the function ϕ must be injective. If it were not, there would be a pair of points $x \neq x'$ for which $\phi(x) = \phi(x')$. For this pair of points we can then write,

$$A_X(x, x') = A_Y(\phi(x), \phi(x')) = 0, \tag{5.86}$$

where the first equality follows from the definition of ϕ and the second equality from the fact that $\phi(x) = \phi(x')$ and that dissimilarity functions are such that $A_Y(y, y) = 0$. However, (5.86) is inconsistent with $x \neq x'$ because the dissimilarity function is $A_X(x, x') = 0$ if and only $x = x'$. It then must be $\phi(x) = \phi(x')$ if and only if $x = x'$ implying that ϕ is an injection.

Likewise, define the function $\psi : Y \to X$ that associates to y any value x among those that form a pair with y in the correspondence R_0,

$$\psi(y) = x_0 \in \left\{ x \,\middle|\, (x, y) \in R_0 \right\} \tag{5.87}$$

Since R_0 is a correspondence the set $\left\{ x \,\middle|\, (x, y) \in R_0 \right\} \neq \emptyset$ is nonempty implying that (5.87) is defined for all $y \in Y$ and since we know that $(\psi(y), y) \in R_0$ we must have $A_X(\psi(y), \psi(y')) = A_Y(y, y')$ for all y, y' from where it follows that the function ψ must be injective.

We have then constructed reciprocal injections $\phi : X \to Y$ and $\psi : Y \to X$. The Cantor-Bernstein-Schroeder theorem [48, Chapter 2.6] applies and guarantees that there exists a bijection between X and Y. This forces X and Y to have the same cardinality and, as a consequence, it forces ϕ and ψ to be bijections. Pick the bijection ϕ and recall that since $(x, \phi(x)) \in R_0$ we must have $A_X(x, x') = A_Y(\phi(x), \phi(x'))$ for all x, x' from where it follows that $N_X \cong N_Y$. ■

Proof of triangle inequality: To show the triangle inequality let correspondences R^* between X and Z and S^* between Z and Y be the minimizing correspondences in (5.82) so that we can write

$$\begin{aligned} d_{\tilde{\mathcal{N}}}(N_X, N_Z) &= \frac{1}{2} \max_{(x,z),(x',z') \in R^*} \big| A_X(x, x') - A_Z(z, z') \big|. \\ d_{\tilde{\mathcal{N}}}(N_Z, N_Y) &= \frac{1}{2} \max_{(z,y),(z',y') \in S^*} \big| A_Z(z, z') - A_Y(y, y') \big|. \end{aligned} \tag{5.88}$$

Define now the correspondence T between X and Y as the one induced by pairs (x, z) and (z, y) sharing a common point $z \in Z$,

$$T := \big\{ (x, y) \,\big|\, \exists \ z \in Z \text{ with } (x, z) \in R^*, (z, y) \in S^* \big\}. \tag{5.89}$$

To show that T is a correspondence we have to prove that for every $x \in X$ there exists $y_0 \in Y$ such that $(x, y_0) \in T$ and that for every $y \in Y$ there exists $x_0 \in X$ such that $(x_0, y) \in T$. To see this pick arbitrary $x \in X$. Because R is a correspondence there must exist $z_0 \in Z$ such that $(x, z_0) \in R$. Since S is also a correspondence, there must exist $y_0 \in Y$ such that $(z_0, y_0) \in S$. Hence, there exists $(x, y_0) \in T$ for every $x \in X$. Conversely, pick an arbitrary $y \in Y$. Since S and R are correspondences there must exist $z_0 \in Z$ and $x_0 \in X$ such that $(z_0, y) \in S$ and $(x_0, z_0) \in R$. Thus, there exists $(x_0, y) \in T$ for every $y \in Y$. Therefore, T is a well defined correspondence.

The correspondence T need not be the minimizing correspondence for the distance $d_{\tilde{\mathcal{N}}}(N_X, N_Y)$, but since it is a valid correspondence we can write [cf. (5.82)]

$$d_{\tilde{\mathcal{N}}}(N_X, N_Y) \leq \frac{1}{2} \max_{(x,y),(x',y') \in T} |A_X(x, x') - A_Y(y, y')|. \tag{5.90}$$

According to the definition of T in (5.89) the requirement $(x, y), (x', y') \in T$ is equivalent to requiring $(x, z), (x', z') \in R^*$ and $(z, y), (z', y') \in S^*$. Further adding and subtracting $A_Z(z, z')$ from the maximand and using the triangle inequality on the absolute value yields

$$d_{\tilde{\mathcal{N}}}(N_X, N_Y) \leq \frac{1}{2} \max_{\substack{(x,z),(x',z') \in R^* \\ (z,y),(z',y') \in S^*}} |A_X(x, x') - A_Z(z, z')| + |A_Z(z, z') - A_Y(y, y')|. \tag{5.91}$$

We can further bound (5.91) by maximizing each summand independently so as to write

$$d_{\tilde{\mathcal{N}}}(N_X, N_Y) \leq \frac{1}{2} \max_{(x,z),(x',z')\in R^*} |A_X(x,x') - A_Z(z,z')| + \frac{1}{2} \max_{(z,y),(z',y')\in S^*} |A_Z(z,z') - A_Y(y,y')|. \tag{5.92}$$

Sunbstituting the equalities in (5.88) for the summands on the right hand side of (5.92) yields the triangle inequality. ■

Having shown the four statements in Theorem 13, the global proof concludes. ■

The guarantee offered by Theorem 13 entails that the space $\tilde{\mathcal{N}} \mod \cong$ of networks modulo isomorphism endowed with the distance defined in (5.82) is a metric space. Restriction of (5.82) to symmetric networks shows that the space $\mathcal{N} \mod \cong$ of symmetric networks [cf. Section 3.2.1] modulo isomorphism is also a metric space. Further restriction to metric spaces shows that the space of finite metric spaces modulo isomorphism is metric [9, Chapter 7.3]. A final restriction of (5.82) to finite ultrametric spaces shows that the space $\mathcal{U} \mod \cong$ of ultrametrics modulo isomorphism is a metric space. Having a properly defined metric to measure distances between networks $\tilde{\mathcal{N}}$ and therefore also between ultrametrics $\mathcal{U} \subset \tilde{\mathcal{N}}$ permits the study of stability of hierarchical clustering methods for asymmetric networks that we undertake in the following section.

5.4.2 Stability of clustering methods

Intuitively, a hierarchical clustering method $\mathcal{H}$ is stable if its application to networks that have small distance between each other results in dendrograms that are close to each other. Formally, we require the distance between output ultrametrics to be bounded by the distance between the original networks as we define next.

(P5) Stability. We say that the clustering method $\mathcal{H} : \tilde{\mathcal{N}} \to \mathcal{U}$ is *stable* if

$$d_{\tilde{\mathcal{N}}}(\mathcal{H}(N_X), \mathcal{H}(N_Y)) \leq d_{\tilde{\mathcal{N}}}(N_X, N_Y), \tag{5.93}$$

for all $N_X, N_Y \in \tilde{\mathcal{N}}$.

Remark 16 Note that our definition of a stable hierarchical clustering method $\mathcal{H}$ coincides with the property of $\mathcal{H} : (\tilde{\mathcal{N}}, d_{\tilde{\mathcal{N}}}) \to (\mathcal{U}, d_{\mathcal{N}}|_{\mathcal{U}\times\mathcal{U}})$ being a 1-Lipschitz map between the metric spaces $(\tilde{\mathcal{N}}, d_{\tilde{\mathcal{N}}})$ and $(\mathcal{U}, d_{\mathcal{N}}|_{\mathcal{U}\times\mathcal{U}})$.

Recalling that the space of ultrametrics $\mathcal{U}$ is included in the space of networks $\tilde{\mathcal{N}}$, the distance $d_{\tilde{\mathcal{N}}}(\mathcal{H}(N_X), \mathcal{H}(N_Y))$ in (5.93) is well defined and endows $\mathcal{U}$ with a metric by Theorem 13. The relationship in (5.93) means that a stable hierarchical clustering method is a non-expansive map from the metric space of networks endowed with the distance defined in (5.82) into itself. A particular consequence of (5.93) is that if networks N_X and N_Y are at small distance $d_{\tilde{\mathcal{N}}}(N_X, N_Y) \leq \epsilon$ of each other, the output ultrametrics of the stable method $\mathcal{H}$ are also at small distance of each other $d_{\tilde{\mathcal{N}}}(\mathcal{H}(N_X), \mathcal{H}(N_Y)) \leq d_{\tilde{\mathcal{N}}}(N_X, N_Y) \leq \epsilon$. This latter observation formalizes the idea that nearby networks yield nearby dendrograms when processed with a stable hierarchical clustering method.

Notice that the stability definition in (P5) extends to the hierarchical quasi-clustering methods introduced in Chapter 4, since the space of quasi-ultrametric networks, just like the space of ultrametric networks, is a subset of the space of asymmetric networks. Thus, we begin by showing the stability of the directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$. The reason to start the analysis with $\tilde{\mathcal{H}}^*$ is that the proof of the following theorem can be used to simplify the proof of stability of other clustering methods.

Theorem 14 *The directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$ with outcome quasi-ultrametrics as defined in* (2.7) *is stable in the sense of property (P5).*

Proof: Given two arbitrary networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$, assume $\eta = d_{\mathcal{N}}(N_X, N_Y)$ and let R be a correspondence between X and Y such that $\text{dis}(R) = 2\eta$. Write $(X, \tilde{u}_X) = \tilde{\mathcal{H}}^*(N_X)$ and $(Y, \tilde{u}_Y) = \tilde{\mathcal{H}}^*(N_Y)$. Fix (x, y) and (x', y') in R. Pick any $[x = x_0, x_1, \ldots, x_n = x']$ with $x_i \in X$ such that $\max_i A_X(x_i, x_{i+1}) = \tilde{u}_X(x, x')$. Choose $[y_0, y_1, \ldots, y_n]$ with $y_i \in Y$ so that $(x_i, y_i) \in R$ for all $i = 0, 1, \ldots, n$. Then, by definition of $\tilde{u}_Y(y, y')$ in (2.7) and the definition of η in (5.82):

$$\tilde{u}_Y(y, y') \leq \max_i A_Y(y_i, y_{i+1}) \leq \max_i A_X(x_i, x_{i+1}) + 2\eta = \tilde{u}_X(x, x') + 2\eta. \tag{5.94}$$

By symmetry, one also obtains $\tilde{u}_X(x, x') \leq \tilde{u}_Y(y, y') + 2\eta$, which combined with (5.94) implies that

$$|\tilde{u}_X(x, x') - \tilde{u}_Y(y, y')| \leq 2\eta. \tag{5.95}$$

Since this is true for arbitrary pairs (x, y) and $(x', y') \in R$, it must also be true for the maximum as well. Moreover, R need not be the minimizing correspondence for the distance between the networks $(X, \tilde{u}_X)$ and $(Y, \tilde{u}_Y)$. However, it suffices to obtain an upper bound implying that

$$d_{\tilde{\mathcal{N}}}((X, \tilde{u}_X), (Y, \tilde{u}_Y)) \leq \frac{1}{2} \max_{(x,y),(x',y') \in R} |\tilde{u}_X(x, x') - \tilde{u}_Y(y, y')| \leq \eta = d_{\tilde{\mathcal{N}}}(N_X, N_Y), \tag{5.96}$$

concluding the proof. ■

Moving into the realm of clustering methods, we show that semi-reciprocal methods

$\mathcal{H}^{\mathrm{SR}(t)}$ are stable in the sense of property (P5) in the following theorem.

Theorem 15 *The semi-reciprocal clustering method $\mathcal{H}^{\mathrm{SR}(t)}$ with outcome ultrametrics as defined in* (3.52) *is stable in the sense of property (P5) for every integer $t \geq 2$.*

The following lemma is used to prove Theorem 15.

Lemma 3 *Given $a, \bar{a}, b, \bar{b}, c \in \mathbb{R}_+$ such that $|a-b| \leq c$ and $|\bar{a}-\bar{b}| \leq c$, then $|\max(a,\bar{a}) - \max(b,\bar{b})| \leq c$.*

Proof: Begin by noticing that

$$a = |a-b+b| \leq |a-b| + |b| = |a-b| + b, \tag{5.97}$$

and similarly for $\bar{a}$ and $\bar{b}$. Thus, we may write

$$\max(a,\bar{a}) \leq \max(|a-b|+b, |\bar{a}-\bar{b}|+\bar{b}). \tag{5.98}$$

By using the bounds assumed in the statement of the lemma, we obtain

$$\max(a,\bar{a}) \leq \max(c+b, c+\bar{b}) = c + \max(b,\bar{b}). \tag{5.99}$$

By applying the same reasoning but starting with $\max(b,\bar{b})$, we obtain that

$$\max(b,\bar{b}) \leq c + \max(a,\bar{a}). \tag{5.100}$$

Finally, by combining (5.99) and (5.100) we obtain the result stated in the lemma. ■

Proof of Theorem 15: To facilitate understanding, we first present the proof for the case $t = 2$, which follows similar steps as those in the proof of Proposition 26 of [10]. Recall that from (3.59), we know that $\mathcal{H}^{\mathrm{SR}(2)} \equiv \mathcal{H}^{\mathrm{R}}$. Given two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ denote by $(X, u_X^{\mathrm{R}}) = \mathcal{H}^{\mathrm{R}}(N_X)$ and $(X, u_Y^{\mathrm{R}}) = \mathcal{H}^{\mathrm{R}}(N_Y)$ the outputs of applying the reciprocal clustering method to such networks. Let $\eta = d_{\tilde{\mathcal{N}}}(N_X, N_Y)$ be the distance between N_X and N_Y as defined by (5.82) and R be the associated minimizing correspondence such that

$$|A_X(x,x') - A_Y(y,y')| \leq 2\eta, \tag{5.101}$$

for all (x,y),$(x',y') \in R$. By reversing the order of (x,y) and (x',y') we obtain that

$$|A_X(x',x) - A_Y(y',y)| \leq 2\eta. \tag{5.102}$$

From (5.101), (5.102), and the definition $\bar{A}_X(x,x') = \max(A_X(x,x'), A_X(x',x))$ for all

$x, x' \in X$, we obtain by Lemma 3 that

$$|\bar{A}_X(x, x') - \bar{A}_Y(y, y')| \leq 2\eta, \tag{5.103}$$

for all (x, y),$(x', y') \in R$. By using the same argument applied in the proof of Theorem 14 to go from (5.95) to (5.96), we obtain that

$$d_{\tilde{\mathcal{N}}}((X, \bar{A}_X), (Y, \bar{A}_Y)) \leq d_{\tilde{\mathcal{N}}}(N_X, N_Y). \tag{5.104}$$

By comparing (2.7) with (3.2), or equivalently in terms of algorithms by comparing (4.22) with (3.76) , it follows that

$$(X, u_X^{\mathrm{R}}) = \tilde{\mathcal{H}}^*(X, \bar{A}_X), \tag{5.105}$$

and similarly for (Y, u_Y^{R}). However, since $\tilde{\mathcal{H}}^*$ is stable from Theorem 14, we obtain that

$$d_{\tilde{\mathcal{N}}}((X, u_X^{\mathrm{R}}), (Y, u_Y^{\mathrm{R}})) \leq d_{\tilde{\mathcal{N}}}((X, \bar{A}_X), (Y, \bar{A}_Y)), \tag{5.106}$$

which combined with (5.104) completes the proof.

In order to prove the statement for any $t \geq 2$, we first show that the difference between the costs of secondary paths is bounded as the following claim states.

Claim 6 *Given two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$, let $\eta = d_{\tilde{\mathcal{N}}}(N_X, N_Y)$ and R be the associated minimizing correspondence. Given two pair of nodes (x, y), $(x', y') \in R$ we have*

$$|A_X^{\mathrm{SR}(t)}(x, x') - A_Y^{\mathrm{SR}(t)}(y, y')| \leq 2\eta, \tag{5.107}$$

where $A_X^{\mathrm{SR}(t)}$ and $A_Y^{\mathrm{SR}(t)}$ are defined as in (3.51).

Proof: Let $P^*_{xx'} = [x = x_0, x_1, ..., x_l = x']$ be a minimizing path in the definition (3.51), implying that

$$A_X^{\mathrm{SR}(t)}(x, x') = \max_{i | x_i \in P^*_{xx'}} A_X(x_i, x_{i+1}). \tag{5.108}$$

Construct the path $P_{yy'} = [y = y_0, y_1, ..., y_l = y']$ in N_Y from y to y' such that $(x_i, y_i) \in R$ for all i. This path is guaranteed to exist from the definition of correspondence. Using the definition in (3.51) and the inequality stated in (5.101), we write

$$A_Y^{\mathrm{SR}(t)}(y, y') \leq \max_{i | y_i \in P_{yy'}} A_Y(y_i, y_{i+1}) \leq \max_{i | x_i \in P^*_{xx'}} A_X(x_i, x_{i+1}) + 2\eta. \tag{5.109}$$

Substituting (5.108) in (5.109) we obtain,

$$A_Y^{\mathrm{SR}(t)}(y, y') \leq A_X^{\mathrm{SR}(t)}(x, x') + 2\eta. \tag{5.110}$$

By following an analogous procedure starting with a minimizing path in the network N_Y, we can show that,

$$A_X^{\mathrm{SR}(t)}(x, x') \leq A_Y^{\mathrm{SR}(t)}(y, y') + 2\eta. \tag{5.111}$$

From (5.110) and (5.111), the desired result in (5.107) follows. ■

To complete the proof we first use Lemma 3 to show that (5.107) implies

$$|\bar{A}_X^{\mathrm{SR}(t)}(x, x') - \bar{A}_Y^{\mathrm{SR}(t)}(y, y')| \leq 2\eta, \tag{5.112}$$

where $\bar{A}_X^{\mathrm{SR}(t)}$ and $\bar{A}_Y^{\mathrm{SR}(t)}$ are defined as in (3.53). We then compare (2.7) and (3.52) to see that

$$(X, u_X^{\mathrm{SR}(t)}) = \tilde{\mathcal{H}}^*(X, \bar{A}_X^{\mathrm{SR}(t)}), \tag{5.113}$$

and similarly for $(Y, u_Y^{\mathrm{SR}(t)})$. Finally, as done for the case $t = 2$, by using stability of $\tilde{\mathcal{H}}^*$ [cf. Theorem 14], the result follows. ■

We now consider representable clustering methods. These are not stable in general as it is possible to find representers that induce methods for which (5.93) is not true. Nevertheless, exploiting the characterization Theorem 11 we can prove stability of representable methods that are admissible and scale preserving:

Theorem 16 *If a hierarchical clustering method $\mathcal{H}$ satisfies axioms (A1)-(A2), representability (P3) and scale preservation (P2) then $\mathcal{H}$ is stable in the sense of property (P5).*

Proof: From Theorem 11 it follows that the given method $\mathcal{H}$ can be represented by a collection of strongly connected structure representers Ω, i.e. $\mathcal{H} \equiv \mathcal{H}^\Omega$. Moreover, by Proposition 17, we can decompose $\mathcal{H}^\Omega$ into a symmetrization operation Λ^Ω followed by single linkage $\mathcal{H}^{\mathrm{SL}}$. Since single linkage is a particular case of directed single linkage $\tilde{\mathcal{H}}^*$, Theorem 14 ensures that the second component of $\mathcal{H}^\Omega$ is stable. Thus, if we show that Λ^Ω is a stable map from the space of networks $\tilde{\mathcal{N}}$ into the space of symmetric networks $\mathcal{N}$, stability of $\mathcal{H}^\Omega$ follows.

Given two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ denote by $(X, \lambda_X^\Omega) = \Lambda^\Omega(N_X)$ and $(X, \lambda_Y^\Omega) = \Lambda^\Omega(N_Y)$ where the dissimilarity functions λ_X^Ω and λ_Y^Ω have values as given by (5.23) for all $x, x' \in X$ and all $y, y' \in Y$, respectively. Let $\eta = d_{\tilde{\mathcal{N}}}(N_X, N_Y)$ be the distance between N_X and N_Y as defined by (5.82) associated to the minimizing correspondence R such that

$$|A_X(x, x') - A_Y(y, y')| \leq 2\eta, \tag{5.114}$$

for all (x,y),$(x',y') \in R$.

Fix (x,y) and $(x',y') \in R$ and notice that from definition (5.23) there must exist a structure representer $\omega \in \Omega$ such that it is possible to find a dissimilarity reducing map ϕ from $\lambda_X^\Omega(x,x') * \omega$ to N_X containing x and x' in its image. Construct the map ϕ' from $(\lambda_X^\Omega(x,x') + 2\eta) * \omega$ to N_Y containing y and y' in its image and such that $(\phi(z), \phi'(z)) \in R$ for every node z of the representer ω. If we show that the map ϕ' is dissimilarity reducing, then this would imply that the multiple $(\lambda_X^\Omega(x,x') + 2\eta)$ is an upper bound for the optimal multiple $\lambda_Y^\Omega(y,y')$, i.e.

$$\lambda_Y^\Omega(y,y') \leq \lambda_X^\Omega(x,x') + 2\eta. \tag{5.115}$$

By following an analogous argument but starting with a dissimilarity reducing map from some representer into N_Y, we can show that

$$\lambda_X^\Omega(x,x') \leq \lambda_Y^\Omega(y,y') + 2\eta. \tag{5.116}$$

By combining (5.115) and (5.116), we obtain that

$$|\lambda_X^\Omega(x,x') - \lambda_Y^\Omega(y,y')| \leq 2\eta, \tag{5.117}$$

which is valid for arbitrary pairs of nodes (x,y) and $(x',y') \in R$. Using the same argument that leads to (5.96) from (5.95), it follows that

$$d_{\tilde{\mathcal{N}}}(\Lambda^\Omega(N_X), \Lambda^\Omega(N_Y)) \leq d_{\tilde{\mathcal{N}}}(N_X, N_Y), \tag{5.118}$$

showing that Λ^Ω is stable as we wanted. Hence, to finish the proof we need to show that the constructed map ϕ' is in fact dissimilarity reducing.

We know that the maps ϕ from $\lambda_X^\Omega(x,x') * \omega$ to N_X is dissimilarity reducing and, thus, the unit dissimilarities in ω must be mapped to dissimilarities

$$A_X(\bar{x}, \bar{x}') \leq \lambda_X^\Omega(x,x'). \tag{5.119}$$

The map ϕ' transforms unit dissimilarities between nodes z and z' in ω into $A_Y(\phi'(z), \phi'(z'))$ where, by construction, $(\phi(z), \phi'(z))$ and $(\phi(z'), \phi'(z')) \in R$. Consequently, from (5.114) and (5.119) we have that

$$A_Y(\phi'(z), \phi'(z')) \leq A_X(\phi(z), \phi(z')) + 2\eta \leq \lambda_X^\Omega(x,x') + 2\eta. \tag{5.120}$$

Hence, by multiplying the representer ω by $\lambda_X^\Omega(x,x') + 2\eta$ we are assured that the map ϕ' is dissimilarity reducing. Notice that in the different case where the method $\mathcal{H}$ is not scale preserving (P2) and hence cannot be represented by representers with unit dissimilarities,

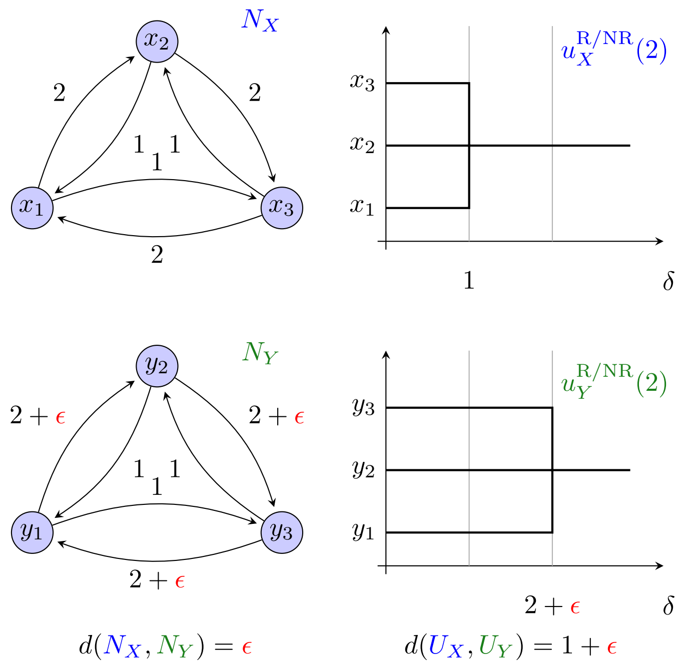


Figure 5.10: Instability of the method $\mathcal{H}^{\mathrm{R/NR}}(2)$. Some dissimilarities in the network N_X are perturbed by an arbitrarily small ϵ to obtain N_Y such that the distance between both networks is ϵ. However, the distance between the output ultrametrics cannot be bounded by the distance between the input networks, violating the definition of stability (5.93).

inequality (5.119) would not be true and map ϕ' need not be dissimilarity reducing. ■

Reciprocal and nonreciprocal clustering are particular cases of admissible scale preserving representable methods. It thus follows from Theorem 16 that these two methods are stable. This result is of sufficient merit so as to be stated separately in the following corollary.

Corollary 6 *The reciprocal and nonreciprocal clustering methods $\mathcal{H}^{\mathrm{R}}$ and $\mathcal{H}^{\mathrm{NR}}$ with output ultrametrics given as in* (3.2) *and* (3.8) *are stable in the sense of property (P5).*

By (5.93), Theorems 15 and 16 show that semi-reciprocal clustering methods and admissible representable scale preserving methods – subsuming the particular cases of reciprocal and nonreciprocal clustering – do not expand distances between pairs of input and their corresponding output networks. In particular, for any method of the above, nearby networks yield nearby dendrograms. This is important when we consider noisy dissimilarity data. Property (P5) ensures that noise has limited effect on output dendrograms.

Theorems 15 and 16 notwithstanding, not all methods that are admissible with respect to axioms (A1) and (A2) are stable. Besides (non scale preserving) admissible representable methods that are not stable in the sense of property (P5) – see last paragraph in the proof of Theorem 16 – the admissible grafting method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ introduced in Section 3.3.1 does not abide by (P5). To see this fix $\beta = 2$ and turn attention to the networks N_X and N_Y shown in Fig. 5.10, where $\epsilon > 0$. For network N_X we have $u_X^{\mathrm{NR}}(x, x') = 1$ and $u_X^{\mathrm{R}}(x, x') = 2$ for all pairs x, x'. Since $u_X^{\mathrm{R}}(x, x') = \beta = 2$ for all x, x', the top condition in definition (3.35) is active and we have $u_X^{\mathrm{R/NR}}(x, x'; 2) = u_X^{\mathrm{NR}}(x, x') = 1$ leading to the top dendrogram in Fig. 5.10. For the network N_Y we have that $u_Y^{\mathrm{R}}(y, y') = 2 + \epsilon > 2 = \beta$ for all y, y'. Thus, the bottom condition in definition (3.35) is active and we have $u_Y^{\mathrm{R/NR}}(y, y'; 2) = u_Y^{\mathrm{R}}(y, y') = 2+\epsilon$ for all y, y'. Given the symmetry in the original network and the output ultrametrics, the correspondence R with $(x_i, y_i) \in R$ for $i = 1, 2, 3$ is an optimal correspondence in the definition in (5.82). It then follows that

$$d_{\tilde{\mathcal{N}}}(\mathcal{H}^{\mathrm{R/NR}}(N_X; 2), \mathcal{H}^{\mathrm{R/NR}}(N_Y; 2)) = 1 + \epsilon > d_{\tilde{\mathcal{N}}}(N_X, N_Y) = \epsilon. \tag{5.121}$$

Comparing (5.121) with (5.93) we conclude that methods in the grafting family $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ are not stable in the sense of property (P5). This observations concurs with our intuition on instability. A small perturbation in the original data results in a large variation in the output ultrametrics. The discontinuity in the grafting method $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ arises due to the switching between reciprocal and nonreciprocal ultrametrics implied by (3.35).

Remark 17 The same tools used in the proofs of Theorems 14, 15, and 16 can be used to show that the unilateral clustering method $\mathcal{H}^{U}$ introduced in Section 3.4.1 is stable and that the convex combination as introduced in Section 3.3.2 of any two stable methods *described in this section* is also stable. However, the respective proofs are omitted to avoid repetition.

Chapter 6

Applications of hierarchical clustering

We apply the hierarchical clustering and quasi-clustering methods developed in this first part of the thesis to determine dendrograms and quasi-dendrograms for two asymmetric network datasets. In Section 6.1 we analyze the internal migration network between states of the United States (U.S.) for the year 2011. In Section 6.2 we analyze the network of interactions between sectors of the U.S. economy for the same year.

6.1 Internal migration in the United States

The number of migrants from state to state, including the District of Columbia (DC) as a separate entity, is published yearly by the geographical mobility section of the U.S. census bureau[1]. We denote by S, with cardinality $|S| = 51$, the set containing every state plus DC and as $M : S \times S \to \mathbb{R}_+$ the migration flow similarity function given by the U.S. census bureau in which $M(s, s')$ is the number of individuals that migrated from state s to s' for all $s, s' \in S$. We then construct the asymmetric network $N_S = (S, A_S)$ with node set S and dissimilarities A_S such that $A_S(s, s) = 0$ for all $s \in S$ and

$$A_S(s, s') = f\left(\frac{M(s, s')}{\sum_i M(s_i, s')}\right), \tag{6.1}$$

for all $s \neq s' \in S$ where $f : [0, 1) \to \mathbb{R}_{++}$ is a given decreasing function. The normalization $M(s, s')/\sum_i M(s_i, s')$ in (6.1) can be interpreted as the probability that an immigrant to state s' comes from state s. The role of the decreasing function f is to transform the similarities $M(s, s')/\sum_i M(s_i, s')$ into corresponding dissimilarities. For the experiments

[1]Available at `http://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html`

here we use $f(x) = 1 - x$. Since the methods we consider are scale preserving [cf. (P2)], the particular form of f is of little consequence to our analysis as it follows from Proposition 14. Dissimilarities $A_S(s, s')$ focus attention on the composition of migration flows rather than on their magnitude. A small dissimilarity from state s to state s' implies that from all the immigrants into s' a high percentage comes from s. E.g., if 85% of the immigration into s' comes from s, then $A_S(s, s') = 1 - 0.85 = 0.15$. The application of hierarchical clustering to migration data has been investigated in the past [76, 77].

Reciprocal clustering $\mathcal{H}^{\mathrm{R}}$ of N_S

The outcome of applying the reciprocal clustering method $\mathcal{H}^{\mathrm{R}}$ defined in (3.2) to the migration network N_S was computed with the algorithmic formula in (3.76). The resulting output dendrogram is shown in Fig. 6.1(a). Moreover, Figs. 6.1(b) through 6.1(e) illustrate the partitions that are obtained at four representative resolutions $\delta_1 = 0.895$, $\delta_2 = 0.921$, $\delta_3 = 0.933$, and $\delta_4 = 0.947$. States marked with the same color other than white are co-clustered at the given resolution whereas states in white are singleton clusters. For a given δ, states that are clustered together in partitions produced by $\mathcal{H}^{\mathrm{R}}$ are those connected by a path of intense bidirectional migration flows in the sense dictated by the resolution under consideration.

The most definite pattern arising from Fig. 6.1 is that migration is highly correlated with geographical proximity. With the exceptions of California, Florida, and Texas that we discuss below, all states merge into clusters with other neighboring states. In particular, the first non-singleton clusters to form are pairs of neighboring states that join together at resolutions smaller than δ_1 with the exception of one cluster formed by three states – New York, New Jersey, and Pennsylvania – as shown in Fig. 6.1(b). In ascending order of resolutions at which they are formed, these pairs are Minnesota and Wisconsin (green, at resolution $\delta = 0.836$), Oregon and Washington (orange, at resolution $\delta = 0.860$), Kansas and Missouri (purple, at resolution $\delta = 0.860$), District of Columbia and Maryland (turquoise, at resolution $\delta = 0.880$), as well as Illinois and Indiana (red, at resolution $\delta = 0.891$). In the group of three states composed of New Jersey, New York, and Pennsylvania, we observe that New York and New Jersey form a cluster (blue) at a smaller resolution ($\delta = 0.853$) than the one at which they merge with Pennsylvania ($\delta = 0.859$). The formation of these clusters can be explained by the fact that these states share respective metropolitan areas. These areas are Minneapolis and Duluth for Minnesota and Wisconsin, Portland for Oregon and Washington, Kansas City for Kansas and Missouri, Washington for the District of Columbia and Maryland, Chicago for Illinois and Indiana, New York City for New York State and New Jersey, as well as Philadelphia for Pennsylvania and New Jersey. Even while crossing state lines, migration within shared metropolitan areas corresponds to people moving to

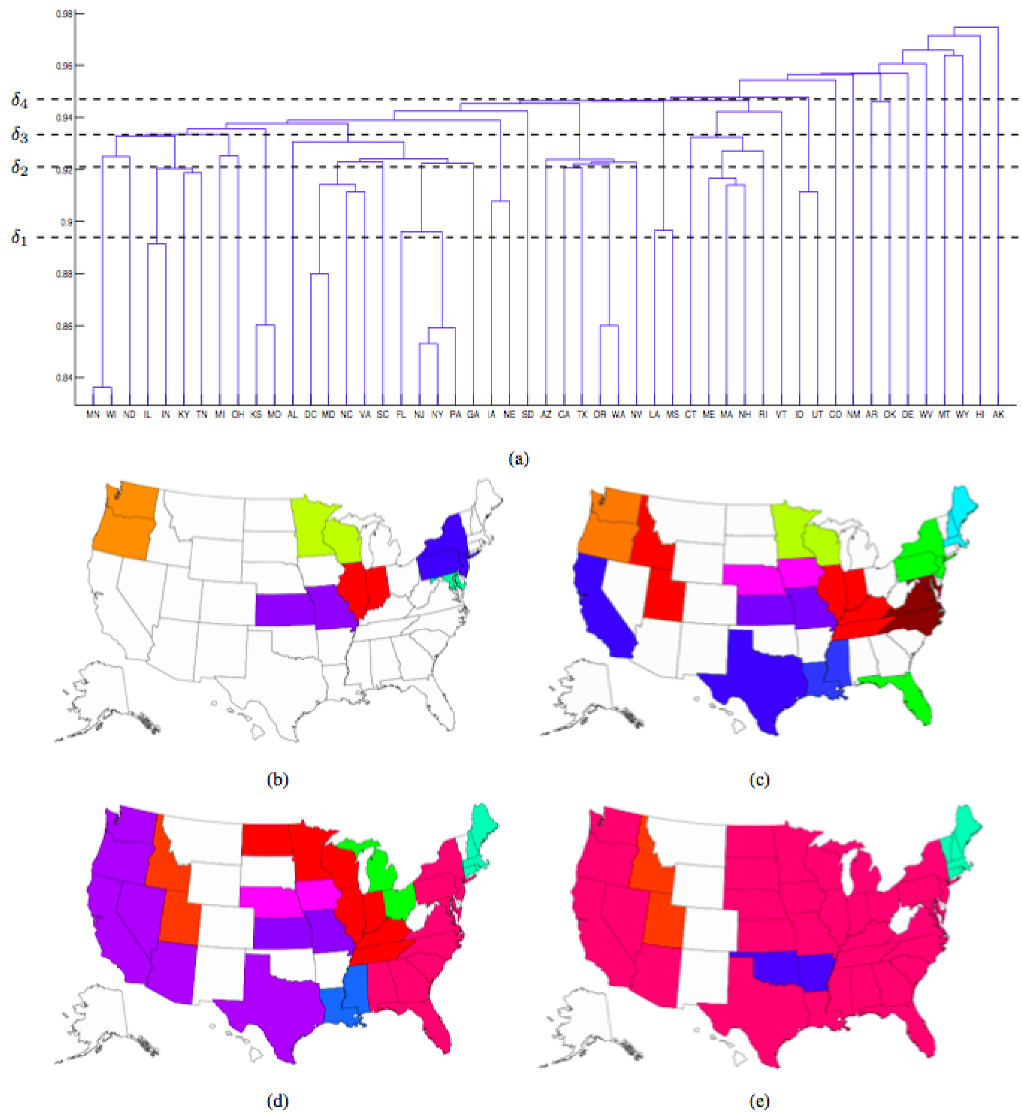


Figure 6.1: (a) Reciprocal dendrogram. Output of clustering method $\mathcal{H}^{\mathrm{R}}$ when applied to the migration network N_S. (b) Clusters at resolution δ_1. States that share urban metropolitan areas merge together first. States in white form singleton clusters at this resolution. (c) Clusters at resolution δ_2. Clusters are highly determined by geographical proximity except for Texas and Florida. (d) Clusters at resolution δ_3. The two coasts form separate clusters. (e) Clusters at resolution δ_4. Most of the nation forms a single cluster. Observe New England's relative isolation.

different neighborhoods or suburbs and occurs frequently enough to suggest it is the reason behind the clusters formed at low resolutions in the reciprocal dendrogram.

As we continue to increase the resolution, clusters formed by pairs of neighboring states continue to appear and a few clusters with multiple states emerge. At resolution δ_2, shown in Fig. 6.1(c), clusters with two adjacent states include Louisiana and Mississippi, Iowa and Nebraska, and Idaho and Utah. Kentucky and Tennessee join Illinois and Indiana to form a midwestern cluster while Maine, Massachusetts, and New Hampshire form a cluster of New England states. The only two exceptions to geographic proximity appear at this resolution. These exceptions are the merging of Florida into the northeastern cluster formed by New Jersey, Pennsylvania, and New York, due to its closeness with the latter, and the formation of a cluster consisting of California and Texas. This anomaly occurs among the four states with the most intense outgoing and incoming migration in the country during 2011. The data analyzed shows that people move from all over the United States to New York, California, Texas, and Florida. For instance, Texas has the lowest standard deviation in the proportion of immigrants from each other state indicating a homogenous migration flow from the whole country. Hence, the proportion of incoming migration from neighboring states is not as significant as for other states. E.g., only 19% of the migration into California comes from its three neighboring states whereas for North Dakota, which also has three neighboring states, these provide 45% of its immigration. Based on the data, we observe that New York, California, Texas, and Florida have a strong influence on the immigration into their neighboring states but, given the mechanics of $\mathcal{H}^{\mathrm{R}}$, the lack of influence in the opposite direction is the reason why Texas joins California and Florida joins New York before forming a cluster with their neighbors. If we require only unidirectional influence, then these four states first join their neighboring states as observed in Fig. 6.3.

Higher resolutions see the appearance of three regional clusters in the Atlantic Coast, Midwest, and New England, as well as a cluster composed of the West Coast states plus Texas. This is illustrated in Fig. 6.1(d) for resolution δ_3. This points towards the fact that people living in a coastal state have a preference to move within the same coast, that people in the midwest tend to stay in the midwest, and that New Englanders tend to stay in New England.

At larger resolutions states start collapsing into a single cluster. At resolution δ_4, shown in Fig. 6.1(e), all states except those in New England and the Mountain West, along with Alaska, Arkansas, Delaware, West Virginia, Hawaii, and Oklahoma are part of a single cluster. The New England cluster includes all six New England states which shows a remarkable degree of migrational isolation with respect to the rest of the country. This indicates that people living in New England tend to move within the region, that people outside New England rarely move into the area, or both. The same observation can be made

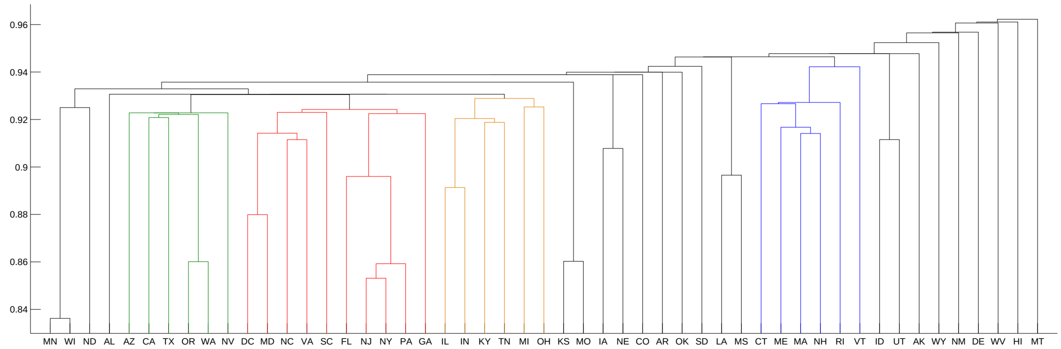


Figure 6.2: Nonreciprocal dendrogram. Dendrogram obtained when applying the nonreciprocal method $\mathcal{H}^{\mathrm{NR}}$ to the state-to-state migration network N_S. The resemblance with the dendrogram in Fig. 6.1(a) indicates that migration cycles are not ubiquitous.

of the pairs Arkansas-Oklahoma and Idaho-Utah. The latter could be partially attributed to the fact that Idaho and Utah are the two states with the highest percentage of mormon population in the country[2]. Four states in the Mountain West – New Mexico, Colorado, Wyoming, and Montana – as well as Delaware, West Virginia, Hawaii and Alaska stay as singleton clusters. Hawaii and Alaska are respectively the next to last, and last state to merge with the rest of the nation further adding evidence to the correlation between geographical proximity and migration clustering.

Nonreciprocal clustering $\mathcal{H}^{\mathrm{NR}}$ of N_S

The outcome of applying the nonreciprocal clustering method $\mathcal{H}^{\mathrm{NR}}$ defined in (3.8) to the migration network N_S is computed with the algorithmic formula in (3.77). The resulting output dendrogram is shown in Fig. 6.2. Comparing the reciprocal and nonreciprocal dendrograms in Figs. 6.1(a) and 6.2 shows that the nonreciprocal clustering method merges any pair of states into a common cluster at a resolution not higher than the resolution at which they are co-clustered by reciprocal clustering. This is as it should be because the uniform dominance of nonreciprocal ultrametrics by reciprocal ultrametrics holds for all networks [cf. (3.14)]. E.g., for the reciprocal method, Colorado and Florida become part of the same cluster at resolution $\delta = 0.954$ whereas for the nonreciprocal case they become part of the same cluster at resolution $\delta = 0.939$. The nonreciprocal resolution need not be strictly smaller, for example, Illinois and Tennessee are merged by both clustering methods at a resolution $\delta = 0.920$.

Further observe that there are many striking similarities between the reciprocal and nonreciprocal dendrograms in Figs. 6.1(a) and 6.2. In both dendrograms, the first three

[2]See `http://religions.pewforum.org/reports`

clusters to emerge are the pair Minnesota and Wisconsin (at resolution $\delta = 0.836$), followed by the pair New York and New Jersey (at resolution $\delta = 0.853$) which are in turn co-clustered with Pennsylvania at resolution $\delta = 0.859$. We then see the emergence of the four pairs: Oregon and Washington (at resolution $\delta = 0.860$), Kansas and Missouri (at resolution $\delta = 0.860$), District of Columbia and Maryland (at resolution $\delta = 0.880$), and Illinois and Indiana (at resolution $\delta = 0.891$). These are the same seven groupings and resolutions at which clusters form in the reciprocal dendrogram that we attributed to the existence of shared metropolitan areas spanning more than one state [cf. Fig. 6.1(b)].

Recall that the difference between the reciprocal and nonreciprocal clustering methods $\mathcal{H}^{\mathrm{R}}$ and $\mathcal{H}^{\mathrm{NR}}$ is that the latter allows influence to propagate through cycles whereas the former requires direct bidirectional influence for the formation of a cluster. In the particular case of the migration network N_S this means that nonreciprocal clustering may be able to detect migration cycles of arbitrary length that are overlooked by reciprocal clustering. E.g., if people in state A tend to move predominantly to B, people in B to move predominantly to C, and people in C move predominantly to A, nonreciprocal clustering merges these three states according to this migration cycle but reciprocal clustering does not. The overall similarity of the reciprocal and nonreciprocal dendrograms in Figs. 6.1(a) and 6.2 suggests that migration cycles are rare in the United States. In particular, the formation of the seven clusters due to shared metropolitan areas indicates that the bidirectional migration flow between these pairs of states is higher than any migration cycle in the country. Notice that highly symmetric data would also correspond to similar reciprocal and nonreciprocal dendrograms. Nevertheless, another consequence of highly symmetric data would be to obtain a unilateral dendrogram similar to the reciprocal and the nonreciprocal ones. Since this is not the case, symmetry cannot be the reason for the similarity observed between the reciprocal and nonreciprocal dendrograms.

However similar, the reciprocal and nonreciprocal dendrograms in Figs. 6.1(a) and 6.2 are not identical. E.g., the last state to merge with the rest of the country in the reciprocal dendrogram is Alaska at resolution $\delta = 0.975$ whereas the last state to merge in the nonreciprocal dendrogram is Montana at resolution $\delta = 0.962$ with Alaska joining the rest of the country at resolution $\delta = 0.948$. Given the mechanics of $\mathcal{H}^{\mathrm{NR}}$, this must occur due to the existence of a cycle of migration involving Alaska which is stronger than the bidirectional exchange between Alaska and any other state, and data confirms this fact.

As we have argued, the areas of the country that cluster together when applying the nonreciprocal method are similar to the ones depicted in Fig. 6.1(d) for the reciprocal clustering method. When we cut the nonreciprocal dendrogram in Fig. 6.2 at resolution $\delta = 0.930$, three major clusters arise – highlighted in green, red, and orange in the dendrogram in Fig. 6.2. The green cluster corresponds to the exact same block containing the West

Coast plus Texas that arises in the reciprocal dendrogram and is depicted in purple in Fig. 6.1(d). The red cluster in the dendrogram corresponds to the East Coast cluster found with the reciprocal method with the exception that Alabama is not included. However, Alabama joins this block at a slightly higher resolution of $\delta = 0.931$, coinciding with the merging of the green, red and orange clusters. The orange cluster in the nonreciprocal dendrogram corresponds to the Midwest cluster found in 6.1(d). However, in contrast with the reciprocal case, Michigan and Ohio join the Midwest cluster before Minnesota, Wisconsin and North Dakota. For the nonreciprocal case, these last three states join the main cluster at resolution $\delta = 0.933$, after the East Coast, West Coast and Midwest become a single block.

The migrational isolation of New England with respect to the rest of the country, which we observed in reciprocal clustering, also arises in the nonreciprocal case. The New England cluster is depicted in blue in the nonreciprocal dendrogram in Fig. 6.2 and joins the main cluster at a resolution of $\delta = 0.946$, which coincides with the merging resolution for the reciprocal case. However, the order in which states become part of the New England cluster varies. In the nonreciprocal case, Connecticut merges with the cluster of Maine-Massachusetts-New Hampshire at resolution $\delta = 0.926$ before Rhode Island which merges at resolution $\delta = 0.927$. However, for the reciprocal case, Rhode Island still merges at the same resolution but Connecticut merges after this at a resolution $\delta = 0.933$. The reason for this is that in the reciprocal case, the states of Connecticut and Rhode Island merge with the cluster Maine-Massachusetts-New Hampshire at the resolution where there exist bidirectional flows with the state of Massachusetts. In the nonreciprocal case, this same situation applies for Rhode Island, but from the data it can be inferred that Connecticut joins the mentioned cluster at a lower resolution due to a migration cycle composed of the path [Connecticut, Maine, New Hampshire, Massachusetts, Connecticut].

Up to this point we see that all the conclusions that we have extracted when applying $\mathcal{H}^{\mathrm{NR}}$ are qualitatively similar to those obtained when applying $\mathcal{H}^{\mathrm{R}}$. This is not surprising because the differences between the reciprocal and nonreciprocal dendrograms either occur at coarse resolutions or are relatively small. In fact, one should expect any conclusion stemming from the application of $\mathcal{H}^{\mathrm{R}}$ and $\mathcal{H}^{\mathrm{NR}}$ to the migration network N_S to be qualitatively similar.

Intermediate clustering methods

From Theorem 4 we know that any clustering method satisfying the Axioms of Value and Transformation applied to the migration network N_S yields an outcome dendrogram such that the resolution at which any pair of states merge in a common cluster is bounded by the resolutions at which the same pair of states is co-clustered in the dendrograms resulting from application of the nonreciprocal and reciprocal clustering methods. Given the similar

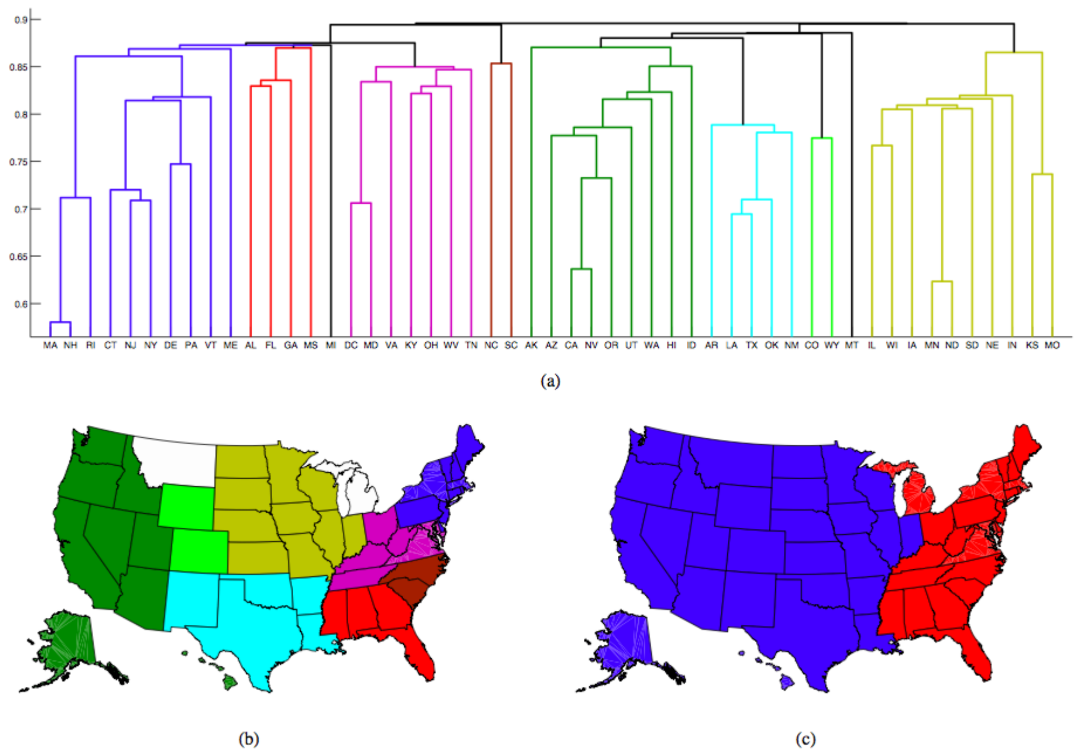


Figure 6.3: Unilateral clustering of the state-to-state migration network. (a) Dendrogram output of applying the unilateral clustering method $\mathcal{H}^{\mathrm{U}}$ to the network of state-to-state migration N_S. Clusters at resolution $\delta_1 = 0.872$ are highlighted in color. (b) Highlighted clusters are identified in a map. Clusters tend to form around high populated states. (c) Map colored according to the partition at resolution $\delta_2 = 0.896$. Two clear clusters, east and west, arise.

conclusions obtained upon analysis of the reciprocal and nonreciprocal clustering outputs we can assert that any other hierarchical clustering method satisfying the Axioms of Value and Transformation would lead to similar conclusions. In particular, this is true for the intermediate methods described in Section 3.3, the algorithmic intermediate of Section 3.5, and the representable methods of Section 5.2.

Unilateral clustering $\mathcal{H}^{\mathrm{U}}$ of N_S

The outcome of applying the unilateral clustering method $\mathcal{H}^{\mathrm{U}}$ defined in (3.62) to the migration network N_S is computed with the algorithmic formula in (3.93). The resulting output dendrogram is shown in Fig. 6.3(a). The colors in the dendrogram correspond to the clusters formed at resolution $\delta_1 = 0.872$ which are also shown in the map in Fig. 6.3(b) with the same color code. States shown in black in Fig. 6.3(a) and white in Fig. 6.3(b) are singleton clusters at this resolution. In Fig. 6.3(c) we show the two clusters that appear when the unilateral dendrogram is cut at resolution $\delta_2 = 0.896$. States that are clustered together in unilateral partitions are those connected by a path of intense unidirectional migration flows in the sense dictated by the resolution under consideration.

In unilateral clustering, the relation between geographical proximity and tendency to form clusters is even more determinant than in reciprocal and nonreciprocal clustering since the exceptions of Texas, California, and Florida do not occur in this case. Indeed, California first merges with Nevada at resolution $\delta = 0.637$, Texas with Louisiana at $\delta = 0.694$, and Florida with Alabama at $\delta = 0.830$, the three pairs of states being neighbors. Moreover, from Fig. 6.3(b) it is immediate that at resolution δ_1 every non-singleton cluster is formed by a set of neighboring states.

Recall that unilateral clustering $\mathcal{H}^{\mathrm{U}}$ abides by the Alternative Axioms of Value and Transformation (A1")-(A2) in contrast to the (regular) Axioms of Value and Transformation satisfied by reciprocal $\mathcal{H}^{\mathrm{R}}$ and nonreciprocal $\mathcal{H}^{\mathrm{NR}}$ clustering. Consequently, unidirectional influence is enough for the formation of a cluster. In the particular case of the migration network N_S this means that unilateral clustering may detect one-way migration flows that are overlooked by reciprocal and nonreciprocal clustering. E.g., if people in state A tend to move to B but people in B rarely move to A either directly or through intermediate states, unilateral clustering merges these two states according to the one-way intense flow from A to B but reciprocal and nonreciprocal clustering do not. The differences between the unilateral dendrogram in Fig. 6.3(a) with the reciprocal and nonreciprocal dendrograms in Figs. 6.1(a) and 6.2 indicate that migration flows which are intense in one way but not in the other are common. E.g., the first two states to merge in the unilateral dendrogram in Fig. 6.3(a) are Massachusetts and New Hampshire at resolution $\delta = 0.580$ because from all the people that moved into New Hampshire, 42% came from Massachusetts, this being the highest value in all the country. The flow in the direction from New Hampshire to Massachusetts is lower, only 9% of the immigrants entering the latter come from the former. This is the reason why these two states are not the first to merge in the reciprocal and nonreciprocal dendrograms. In these previous cases, Minnesota and Wisconsin were the first to merge because the relative flow in both directions is 16% and 19%.

Unilateral clusters tend to form around populous states. In Fig. 6.3(b), the six clusters with more than two states contain the seven states with largest population – California, Texas, New York, Florida, Illinois, Pennsylvania, and Ohio – one in each cluster except for the blue one that contains New York and Pennsylvania. The data suggests that the reason for this is that populous states have a strong influence on the immigration into neighboring states. Indeed, if we focus on the cyan cluster formed around Texas, the proportional immigration into Louisiana, New Mexico, Oklahoma, and Arkansas coming from Texas is 31%, 22%, 29%, and 21% respectively. The opposite is not true, since the immigration into Texas from the four aforementioned neighboring state is of 5%, 3%, 4%, and 3%, respectively. However, this flow in the opposite direction is not required for unilateral clustering to merge the states into one cluster. Between two states with large population, the immigration is more balanced in both directions, thus merging at high resolutions in the unilateral dendrogram. E.g., 11% of the immigration into Texas comes from California and 8% in the opposite direction.

Unilateral clustering detects an east-west division of migration flows in the United States. The last merging in the unilateral dendrogram occurs at resolution $\delta = 0.8958$ and just below the merging resolution, e.g. at resolution δ_2, there are two clusters – east and west – corresponding to the ones depicted in Fig. 6.3(c). The cut at δ_2 corresponds to a migrational flow of 10.45%. This implies that for any two different states within the same cluster we can find a unilateral path where every flow is at least 10.45%. More interestingly, there is no pair of states, one from the east and one form the west, with a flow of 10.45% or more in any direction.

Directed single linkage quasi-clustering $\tilde{\mathcal{H}}^*$ of N_S

The outcome of applying the directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$ with output quasi-ultrametrics defined in (2.7) to the migration network N_S is computed with the algorithmic formula in (4.22). In Fig. 6.4 we show some quasi-partitions of the output quasi-dendrogram $\tilde{D}_S^* = (D_S^*, E_S^*)$ focusing on New England and an extended West Coast including Arizona and Nevada. States represented with the same color are part of the same cluster at the given resolution and states in white form singleton clusters. Arrows between clusters for a given resolution δ represent the edge set $E_S^*(\delta)$ for resolution δ. The resolutions δ at which quasi-partitions are shown in Fig. 6.4 correspond to those 0.001 smaller than those in which mergings in the dendrogram component D_S^* of the output quasi-dendrogram $\tilde{D}_S^*$ occur or, in the case of the last map in each figure, correspond to the resolution of the last merging in the region shown. E.g., Oregon and Washington merge at resolution $\delta = 0.860$, thus, in the first map of the West Coast we look at the quasi-partition at resolution $\delta = 0.859$.

The directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$ captures not only the formation of clusters but also the asymmetric influence between them. E.g. the New England's quasi-partition in Fig. 6.4 for resolution $\delta = 0.913$ is of little interest since every state forms a singleton cluster. The influence structure, however, reveals a highly asymmetric migration pattern. At this resolution Massachusetts has migrational influence over every other state in the region as depicted by the five arrows leaving Massachusetts and entering each of the other five states. No state has influence over Massachusetts at this resolution since this would imply the formation of a non-singleton cluster by the mechanics of $\tilde{\mathcal{H}}^*$. This influence could be explained by the fact that Massachusetts contains Boston, the largest urban

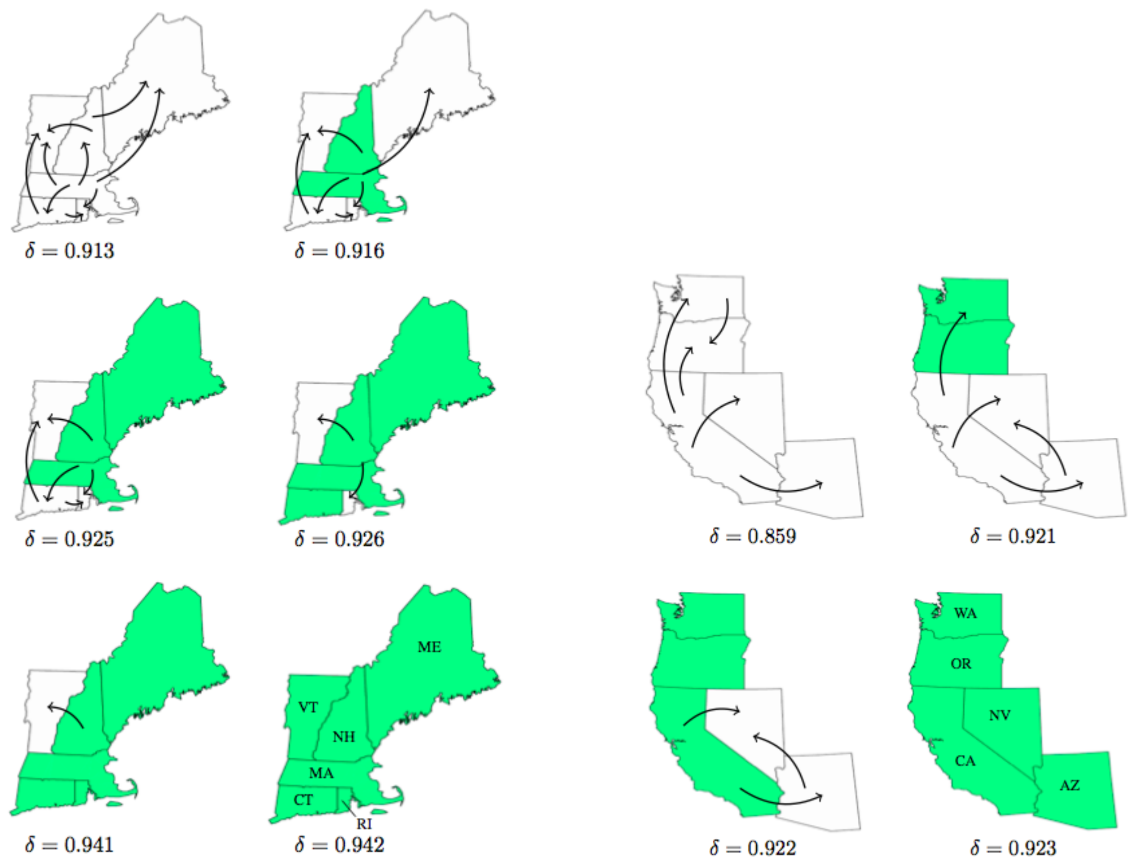


Figure 6.4: Directed single linkage quasi-clustering method applied to New England's (left) and the extended West Coast's (right) migration flows. Quasi-partitions shown for resolutions before every merging and after the last. Massachusetts and California migrational influences over New England and the West Coast, respectively, are represented by the outgoing edges in the quasi-partitions.

area of the region. Hence, Boston attracts immigrants from all over the country reducing the proportional immigration into Massachusetts from its neighbors and generating the asymmetric influence structure observed. This is consistent with the conclusions regarding clustering around populous states that we reached by analyzing the unilateral clusters in Fig. 6.3(b). However, in the quasi-partition analysis, as opposed to the unilateral clustering analysis, the influence of Massachusetts over the other states can be seen clearly as it is formally captured in the edge set $E_S^*(0.913)$. The rest of the influence pattern at this resolution sees Connecticut influencing Rhode Island and Vermont, and New Hampshire influencing Maine and Vermont.

At resolution $\delta = 0.916$, we see that Massachusetts has merged with New Hampshire and this main cluster exerts influence over the rest of the region. Similarly, at resolution $\delta = 0.925$, Maine has joined the cluster formed by Massachusetts and New Hampshire and together they exert influence over the singleton clusters of Connecticut, Rhode Island, and

Vermont. The influence arcs from Connecticut to Rhode Island and Vermont persist in these two diagrams. We know that this has to be the case due to the influence hierarchy property of the the edge sets E_S^* stated in condition ($\tilde{\mathrm{D}}3$) in the definition of quasi-dendrogram in Section 4.1. At resolution $\delta = 0.926$ Connecticut joins the main cluster while Rhode Island joins at resolution $\delta = 0.927$, thus we depict the corresponding maps at resolutions 0.001 smaller than these merging resolutions. The whole region becomes one cluster at resolution $\delta = 0.942$ – which marks the joining of Vermont into the cluster.

For the case of the West Coast in Fig. 6.4, California is the most influential state as expected from its large population. The quasi-partition at resolution $\delta = 0.859$ is such that all states are singleton clusters with California exerting influence onto all other West Coast states and Washington exerting influence on Oregon. The first cluster to form does not involve California but Washington and Oregon merging at resolution $\delta = 0.860$ and the cluster can be observed from the map at resolution $\delta = 0.921$. However, California has influence over this two-state cluster as shown by the arrow going from California to the green cluster in the corresponding figure. The influence over the two other states, Nevada and Arizona, remains. This is as it should be because of the persistence property of the edge set E_S^*. At this resolution we also see an influence arc appearing from Arizona to Nevada. At resolution $\delta = 0.922$ California joins the Washington-Oregon cluster that exerts influence over Arizona and Nevada. The whole region merges in a common cluster at resolution $\delta = 0.923$.

An important property of quasi-dendrograms is that the quasi-partitions at any given resolution define a partial order between the clusters. Recall that slicing a *dendrogram* at certain resolution yields a partition of the node set where there is no defined order between the blocks of the partition. Slicing a *quasi-dendrogram* yields also an edge set $E_S^*(\delta)$ that defines a partial order among the clusters at such resolution. This partial order is useful because it allows us to ascertain the relative importance of different clusters. E.g., in the case of the extended West Coast in Fig. 6.4 one would expect California to be the dominant migration force in the region. The quasi-partition at resolution $\delta = 0.859$ permits asserting this fact formally because the partial order at this resolution has California ranked as more important than any other state. We also see the not unreasonable dominance of Washington over Oregon, while the remaining pairs of the ordering are not defined.

At larger resolutions we can ascertain relative importance of clusters. At resolution $\delta = 0.921$ we can say that California is more important than the cluster formed by Oregon and Washington as well as more important than Arizona and Nevada. We can also see that Arizona precedes Nevada in the migration ordering at this resolution while the remaining pairs of the ordering are undefined. At resolution $\delta = 0.922$ there is an interesting pattern as we can see the cluster formed by the three West Coast states preceding Arizona and

Nevada in the partial order. At this resolution the partial order also happens to be a total order as Arizona is seen to precede Nevada. This is not true in general as we have already seen.

In New England and the West Coast, the respective importance of Massachusetts and California over nearby states acts as an agglutination force towards regional clustering. Indeed, if we delete any of these two states and cluster the remaining states in the corresponding region, the resolution at which the whole region becomes one cluster is increased, showing a decreasing tendency to cluster. E.g., for the case of New England, if we delete Massachusetts and cluster the remaining five states, they become one regional cluster at a resolution of $\delta = 0.979$ whereas if we delete, e.g. Maine or Rhode Island, the remaining five states merge into one single cluster at resolution $\delta = 0.942$ as in the original case [cf. Fig. 6.4].

Further observe that if we limit our attention to the dendrogram component of the quasi-dendrogram depicted in Fig. 6.4, i.e., if we ignore the edge sets $E_S^*(\delta)$, we recover the information in the nonreciprocal dendrogram in Fig. 6.2. In the case of New England the dendrogram part D_S^* of the quasi-dendrogram $\tilde{D}_S^*$ has the mergings occurring at resolutions 0.001 larger than the resolutions used to depict the quasi-partitions, i.e. Massachusetts first merges with New Hampshire ($\delta = 0.914$), then Maine joins this cluster ($\delta = 0.917$), followed by Connecticut ($\delta = 0.926$), Rhode Island ($\delta = 0.927$) and finally Vermont ($\delta = 0.942$). The order and resolutions in which states join the main cluster coincides with the blue part of the nonreciprocal dendrogram in Fig. 6.2. In the case of the extended West Coast in Fig. 6.4 we have Oregon joining Washington ($\delta = 0.860$), which are then joined by California ($\delta = 0.922$), which are then joined by Arizona and Nevada at resolution $\delta = 0.923$. Observe that Arizona and Nevada do not form a separate cluster before joining California, Oregon, and Washington. They both join the rest of the states at the exact same resolution. This is the same order and the same resolutions corresponding to the green part of the nonreciprocal dendrogram in Fig. 6.2. Notice that while Texas appears in the nonreciprocal dendrogram it does not appear in the quasi-partitions. This is only because we decided to show a partial view of the extended West Coast without including Texas. The fact that when we limit our attention to the dendrogram component of the quasi-dendrogram we recover the nonreciprocal dendrogram is not a coincidence. We know from Proposition 12 that the dendrogram component of the quasi-partitions generated by directed single linkage is equivalent to the dendrograms generated by nonreciprocal clustering.

Table 6.1: Code and description of industrial sectors

Code	Industrial Sector	Code	Industrial Sector
AC	Accommodation	AG	Amusements, gambling, and recreation
AH	Ambulatory health care services	AP	Apparel and leather and allied products
AS	Administrative and support services	AT	Air transportation
BT	Broadcasting and telecommunications	CE	Computer and electronic products
CH	Chemical products	CO	Construction
CS	Computer systems design	ED	Educational services
EL	Electrical equip., appliances, and comp.	FA	Farms
FB	Food and beverage and tobacco prod.	FM	Fabricated metal products
FO	Forestry, fishing, and related activities	FP	Food services and drinking places
FR	Federal Reserve banks and credit interm.	FU	Furniture and related products
FT	Funds, trusts, and others	HN	Hospitals and nursing facilities
IC	Insurance carriers and related activities	ID	Information and data process. serv.
LS	Legal services	MA	Machinery
MC	Management of companies and enterprises	MI	Mining, except oil and gas
MM	Miscellaneous manufacturing	MP	Misc. prof., scientific, and tech. serv.
MV	Motor vehicles, bodies and trailers	NM	Nonmetallic mineral products
OG	Oil and gas extraction	OS	Other services, except government
OT	Other transportation and support activities	PA	Paper products
PC	Petroleum and coal products	PE	Performing arts, sports and museums
PL	Plastics and rubber products	PM	Primary metals
PR	Printing and related support activities	PS	Motion picture and sound recording
PT	Pipeline transportation	PU	Publishing industries (incl. software)
RA	Real estate	RE	Retail trade
RL	Rental and leasing of intang. assets	RT	Rail transportation
SA	Social assistance	SC	Securities, commodities, and invest.
SM	Support activities for mining	TE	Textile mills and textile product mills
TG	Transit and ground passenger transportation	TM	Other transportation equipment
TT	Truck transportation	UT	Utilities
WH	Wholesale trade	WM	Waste management and remediation
WO	Wood products	WS	Warehousing and storage
WT	Water transportation		

6.2 Interactions between sectors of the economy

The Bureau of Economic Analysis of the U.S. Department of Commerce publishes a yearly table of input and outputs organized by economic sectors[3]. This table records how economic sectors interact to generate gross domestic product. We focus on a particular section of this table, called *uses*, which shows the inputs to production for year 2011. More precisely, we are given a set I of 61 industrial sectors as defined by the North American Industry Classification System (NAICS) – see Table 6.1 – and a similarity function $U : I \times I \to \mathbb{R}_+$ where $U(i, i')$ for all $i \neq i' \in I$ represents how much of the production of sector i, expressed in dollars, is used as an input of sector i'. Thus, we define the network $N_I = (I, A_I)$ where the dissimilarity function A_I satisfies $A_I(i, i) = 0$ and, for $i \neq i' \in I$, is given by

$$A_I(i, i') = f\left(\frac{U(i, i')}{\sum_k U(i_k, i')}\right), \tag{6.2}$$

where $f : [0, 1) \to \mathbb{R}_{++}$ is a given decreasing function. For the experiments here we use $f(x) = 1 - x$. The normalization $U(i, i')/\sum_k U(i_k, i')$ in (6.2) can be interpreted as the proportion of the input in dollars to productive sector i' that comes from sector i. In this way, we focus on the combination of inputs of a sector rather than the size of the economic sector itself. That is, a small dissimilarity from sector i to sector i' implies that sector i' highly relies on the use of sector i output as an input for its own production. E.g., if 40% of the input into sector i' comes from sector i, we say that sector i has an influence of 40% over i' and the dissimilarity $A_I(i, i') = 1 - 0.40 = 0.60$. Notice that it is common for part of the output of some sector $i \in I$ to be used as input in the same sector. Consequently, if for a given sector we sum the input proportion from every other sector, we obtain a number less than 1. The role of the decreasing function f is to transform the similarities into corresponding dissimilarities. As in the case of the migration matrix in Section 6.1 the particular form of f is of little consequence to the analysis as long as it is a decreasing function since we focus on scale preserving methods, [cf. (P2) and Proposition 14].

Reciprocal clustering $\mathcal{H}^{\mathrm{R}}$ of N_I

The outcome of applying the reciprocal clustering method $\mathcal{H}^{\mathrm{R}}$ defined in (3.2) to the network N_I is computed with the algorithmic formula in (3.76). The resulting output dendrogram is shown in Fig. 6.5(a) where three clusters are highlighted in blue, red and green. These clusters appear at resolutions $\delta_1^{\mathrm{R}} = 0.959$, $\delta_2^{\mathrm{R}} = 0.969$, and $\delta_3^{\mathrm{R}} = 0.977$, respectively. In Fig. 6.5(b) we present the three highlighted clusters with edges representing bidirectional influence between industrial sectors at the corresponding resolution. That is, a double arrow

[3] Available at `http://www.bea.gov/industry/io_annual.htm`

is drawn between two nodes if and only if the dissimilarity between these nodes in both directions is less than or equal to the resolution at which the corresponding cluster appears. In particular, it shows the bidirectional paths of minimum cost between two nodes. E.g., for the blue cluster ($\delta_1^{\mathrm{R}} = 0.959$) the bidirectional path of minimum cost from the sector 'Rental and leasing services of intangible assets' (RL) to 'Computer and electronic products' (CE) goes through 'Management of companies and enterprises' (MC).

According to our analysis, the reciprocal clustering method $\mathcal{H}^{\mathrm{R}}$ tends to cluster sectors that satisfy one of two possible typologies. The first type of clustering occurs among sectors of balanced influence in both directions. E.g., the first two sectors to be merged by $\mathcal{H}^{\mathrm{R}}$ are 'Administrative and support services' (AS) and 'Miscellaneous professional, scientific and technical services' (MP) at a resolution of $\delta = 0.887$. This occurs because 13.2% of the input of AS comes from MP – corresponding to $A_I(\mathrm{MP}, \mathrm{AS}) = 0.868$ – and 11.3% of MP's input comes from AS – implied by $A_I(\mathrm{AS}, \mathrm{MP}) = 0.887$ – both influences being similar in magnitude. It is reasonable that these two sectors hire services from each other in order to better perform their own service. This balanced behavior is more frequently observed among service sectors than between raw material extraction (primary) or manufacturing (secondary) sectors. Notice that for two manufacturing sectors A and B to have balanced bidirectional influence we need the outputs of A to be inputs of B in the same proportion as the outputs of B are inputs of A. This situation is rarer. Further examples of this clustering typology where the influence in both directions is balanced can be found between pairs of service sectors with bidirectional edges in the blue cluster formed at resolution $\delta_1^{\mathrm{R}} = 0.959$. E.g., the participation of RL in the input to MC is of 7.6% – since $A_I(\mathrm{RL}, \mathrm{MC}) = 0.924$ – whereas the influence in the opposite direction is 8.5%. Similarly, 6.5% of the input to the 'Real estate' (RA) sector comes from AS and 6.0% vice versa. This implies that the RA sector hires external administrative and support services and the AS sector depends on the real estate services to, e.g., rent a location for their operation. The second type of clustering occurs between sectors with one natural direction of influence but where the influence in the opposite direction is meaningful. E.g., the second merging in the reciprocal dendrogram in Fig. 6.5(a) occurs at resolution $\delta = 0.893$ between the 'Farm' (FA) sector and the 'Food, beverage and tobacco products' (FB) sector. In this case, one expects a big portion of FB's input to come from FA – 35.2% to be precise – as raw materials for processed food products but there is also a dependency on the opposite direction of 10.7% from, e.g., food supplementation for livestock. This second clustering typology generally occurs between consecutive sectors in the production path of a particular industry, with the strong influence in the natural direction of the material movement and the non-negligible influence in the opposite direction which is particular of each industry. E.g., for the food industry, the primary FA sector precedes in the production process the secondary FB sector. Thus, the influence of FA over FB is clear. However, there is an influence of FB over FA that could be explained by the provision of food supplementation for livestock. Further examples of this interaction between sectors can be found in the textile and metal industries. Representing the textile industry, at resolution $\delta = 0.938$ the sectors 'Textile mills and textile product mills' (TE) and 'Apparel and leather and allied products' (AP) merge. In the garment production process, there is a natural direction of influence from TE that generates fabric from a basic fiber to AP that cuts and sews the fabric to generate garments. Indeed, the influence in this direction is of 17.8% represented by $A_I(\mathrm{TE}, \mathrm{AP}) = 0.822$. However, there is an influence of 6.2% in the opposite direction. This influence can be partially attributed to companies in the TE sector which also manufacture garments and buy intermediate products from companies in the AP sector. For example, a textile mill that produces wool fabric and also manufactures wool garments with some details in leather. This leather comes from a company in the AP sector and represents a movement from AP back to TE. In the metal industry, at resolution $\delta = 0.960$ 'Mining, except oil and gas' (MI) merges with 'Primary metals' (PM). The bidirectional influence between these two sectors can be observed in the red cluster formed at resolution $\delta_2^{\mathrm{R}} = 0.969$ in Fig. 6.5(b). As before, the natural influence is in the direction of the production process, i.e. from MI to PM. Indeed, 9.3% of PM's input comes from MI mainly as ores for metal manufacturing. Moreover, there is an influence of 4.0% in the opposite direction from PM to MI due to, e.g., structural metals for mining infrastructure.

The cluster in Fig. 6.5 that forms at resolution $\delta_1^{\mathrm{R}} = 0.959$ (blue) is mainly composed of services. The first two mergings, described in the previous paragraph, occur between MP-AS and RL-MC representing professional, support, rental and management services, respectively. At resolution $\delta = 0.925$, the sectors 'Federal Reserve banks, credit intermediation, and related activities' (FR) and 'Securities, commodity contracts, and investments' (SC) merge. This is an exception to the described balanced mergings between service sectors. Indeed, 24.1% of FR's input comes from SC whereas only 7.5% of SC's input comes from FR. This is expected since credit intermediation entities in FR have as input investments done in the SC sector. At resolution $\delta = 0.940$, RA joins the MP-AS cluster due to the bidirectional influence between RA and AS described in the previous paragraph. The MP-AS-RA cluster merges with the FR-SC cluster at resolution $\delta = 0.948$ due to the relation between MP and FR. More precisely, MP provides 11.3% of FR input and 5.2% of MP's input comes from FR. At resolution $\delta = 0.957$, CE joins the RL-MC cluster due to its bidirectional influence relation with MC. The sector of electronic products CE is the only sector in the blue cluster formed at resolution $\delta_1^{\mathrm{R}} = 0.959$ that does not represent a service. The 'Insurance carriers and related activities' (IC) sector joins the MP-AS-RA-FR-SC cluster at resolution $\delta = 0.959$ because of its relation with SC. In fact, 4.5% of IC's input

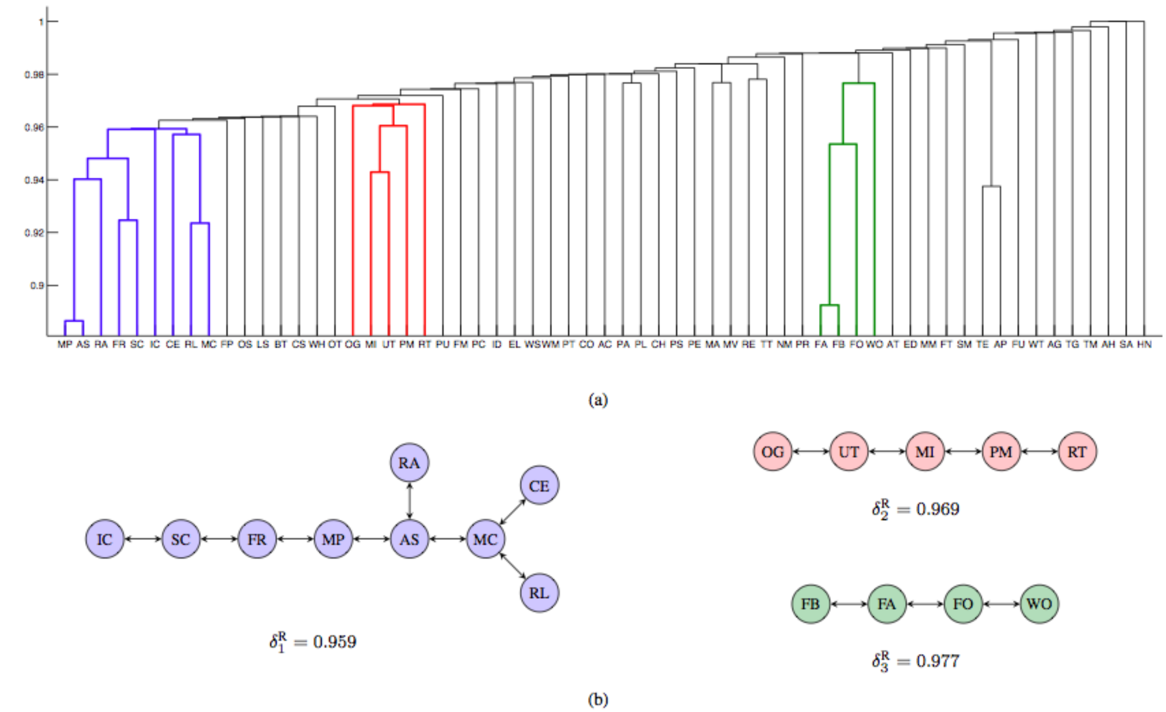


Figure 6.5: (a) Reciprocal dendrogram. Output of the reciprocal clustering method $\mathcal{H}^{\mathrm{R}}$ when applied to the network N_I. Three clusters formed at resolutions $\delta_1^{\mathrm{R}} = 0.959$, $\delta_2^{\mathrm{R}} = 0.969$, and $\delta_3^{\mathrm{R}} = 0.977$ are highlighted in blue, red and green, respectively. (b) Highlighted clusters. Edges between sectors represent bidirectional influence between them at the corresponding resolution.

comes from SC in the form of securities and investments and 4.1% of SC's input comes from IC in the form of insurance policies for investments. Finally, at resolution $\delta_1^{\mathrm{R}} = 0.959$, the clusters MP-AS-RA-FR-SC-IC and CE-RL-MC merge due to the relation between the supporting services AS and the management services MC.

The cluster in Fig. 6.5 that forms at resolution $\delta_2^{\mathrm{R}} = 0.969$ (red) mixes the three levels of the economy: raw material extraction or primary, manufacturing or secondary and services or tertiary. The 'Mining, except oil and gas' sector (MI), which is a primary activity of extraction, merges at resolution $\delta = 0.943$ with the 'Utilities' (UT) sector which extends vertically into the secondary and tertiary industrial sectors since it generates and distributes energy. This merging occurs because 5.7% of UT's input comes from MI and 8.8% vice versa. This pair then merges at resolution $\delta = 0.961$ with the manufacturing sector of 'Primary metals' (PM). PM joins this cluster due to its bidirectional relation with MI previously described. At resolution $\delta = 0.968$, the primary sector of 'Oil and gas extraction' (OG) joins the MI-UT-PM cluster because 3.2% of OG's input comes from UT, mainly as electric power supply, and 57.3% of UT's input comes from OG as natural gas for combustion and distribution. Finally, at resolution $\delta_2^{\mathrm{R}} = 0.969$ the service sector of 'Rail transportation'

(RT) merges with the rest of the cluster due to its influence relation with PM. Indeed, PM provides 7.0% of the input of RT for the construction of railroads – corresponding to $A_I(\text{PM}, \text{RT}) = 0.930$ – and RT provides 3.1% of PM's input as transportation services for final metal products.

The cluster in Fig. 6.5 that forms at resolution $\delta_3^{\text{R}} = 0.977$ (green) is composed of food and wood generation and processing. It starts with the aforementioned merging between FA and FB at $\delta = 0.893$. At resolution $\delta = 0.956$, 'Forestry, fishing, and related activities' (FO) joins the FA-FB cluster due to its relation with FA. The farming sector FA depends 9.2% on FO. The dependence in the opposite direction is of 4.7%. Finally, at $\delta_3^{\text{R}} = 0.977$, 'Wood products' (WO) joins the cluster. Its relation with FO is highly asymmetric and corresponds to the second clustering typology described at the beginning of this section. There is a natural influence in the direction of the material movement from FO to WO. Indeed, 26.2% of WO's input comes from FO whereas the influence is of 2.3% in the opposite direction.

Requiring direct bidirectional influence for clustering generates some cluster which are counter-intuitive. E.g., in the reciprocal dendrogram in Fig. 6.5(a), at resolution $\delta = 0.971$ when the blue and red clusters merge together we have that the oil and gas sector OG in the red cluster joins the insurance sector IC in the blue cluster. However, OG does not merge with 'Petroleum and coal products' (PC), a sector that one would expect to be more closely related, until resolution $\delta = 0.975$. In order to avoid this situation, we may allow nonreciprocal influence as we do next.

Nonreciprocal clustering $\mathcal{H}^{\text{NR}}$ of N_I

The outcome of applying the nonreciprocal clustering method $\mathcal{H}^{\text{NR}}$ defined in (3.8) to the network N_I is computed with the formula in (3.77). The resulting output dendrogram is shown in Fig. 6.6(a). Let us first observe, as we did for the case of the migration matrix, that the nonreciprocal ultrametric distances in Fig. 6.6(a) are not larger than the reciprocal ultrametric distances in Fig. 6.5(a) as it should be the case given the inequality in (3.14). As a test case we have that the mining sector MI and the 'Pipeline transportation' (PT) sectors become part of the same cluster in the reciprocal dendrogram at a resolution $\delta = 0.979$ whereas they merge in the nonreciprocal dendrogram at resolution $\delta' = 0.912 < 0.979$.

A more interesting observation is that, in contrast with the case of the migration matrix, the nonreciprocal dendrogram is qualitatively very different from the reciprocal dendrogram. In the reciprocal dendrogram we tended to see the formation of definite clusters that then merged into larger clusters at coarser resolutions. The cluster formed at resolution $\delta_1^{\text{R}} = 0.959$ (blue) shown in Fig. 6.5(b) grows by merging with singleton clusters (FP, OS, LS, BT, CS, WH, and OT in progressive order of resolution) until it merges at resolution $\delta = 0.971$

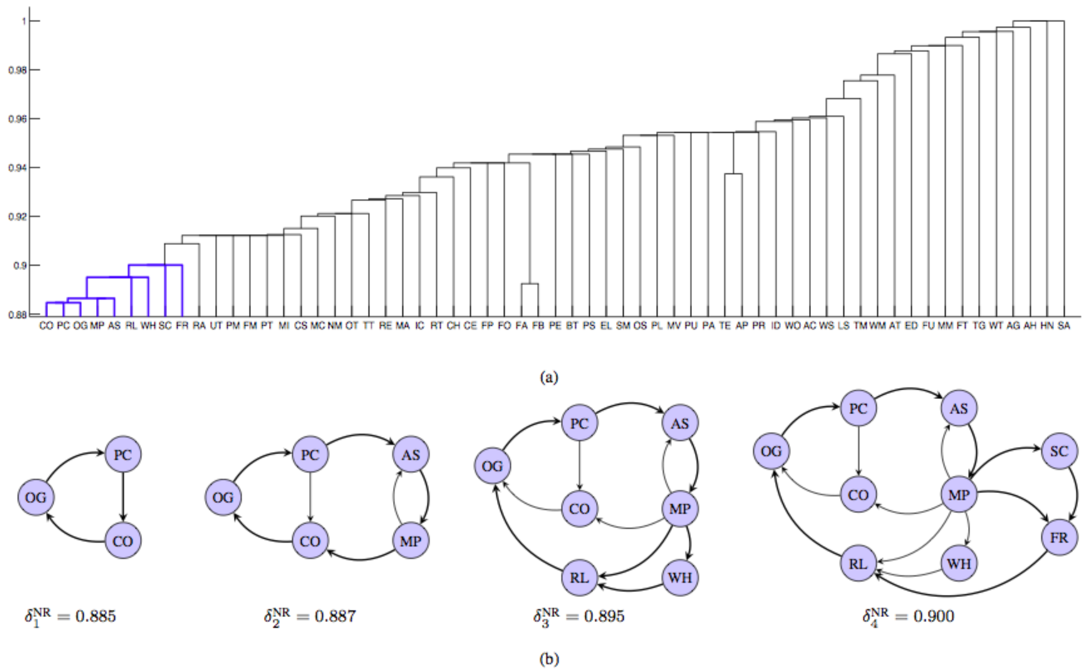


Figure 6.6: (a) Nonreciprocal dendrogram. Output of $\mathcal{H}^{\mathrm{NR}}$ when applied to N_I. One cluster, formed at resolution $\delta_4^{\mathrm{NR}} = 0.900$, is highlighted in blue. (b) Sequential mergings of sectors at resolutions $\delta_1^{\mathrm{NR}} = 0.885$, $\delta_2^{\mathrm{NR}} = 0.887$, $\delta_3^{\mathrm{NR}} = 0.895$, and $\delta_4^{\mathrm{NR}} = 0.900$ are shown. Directed edges between sectors imply unidirectional influence between them at the corresponding resolution.

with a cluster of five nodes which emerges at resolution $\delta_2^{\mathrm{R}} = 0.969$. This whole cluster then grows by adding single nodes and pairs of nodes until it merges at resolution $\delta = 0.988$ with a cluster of four nodes that forms at resolution $\delta_3^{\mathrm{R}} = 0.977$. In the nonreciprocal dendrogram, in contrast, we see the progressive agglutination of economic sectors into a central cluster.

Indeed, the first non-singleton cluster to arise is formed at resolution $\delta_1^{\mathrm{NR}} = 0.885$ by the sectors of oil and gas extraction OG, petroleum and coal products PC, and 'Construction' (CO). For reference, observe that this happens before the first reciprocal merging between AS and MP, which occurs at resolution $\delta = 0.887$ [cf. Fig. 6.5(a)]. The cluster formed by OG, PC, and MP is shown in the leftmost graph in Fig. 6.6(b) where the directed edges represent all the dissimilarities $A_I(i, i') \leq \delta_1^{\mathrm{NR}} = 0.885$ between these three nodes. We see that this cluster forms due to the influence cycle [OG, PC, CO, OG]. Of all the economic input to PC, 82.6% comes from the OG sector – which is represented by the dissimilarity $A_I(\mathrm{OG}, \mathrm{PC}) = 0.174$ – in the form of raw material for its productive processes of which the dominant process is oil refining. In the input to CO a total of 11.5% comes from PC as fuel and lubricating oil for heavy machinery as well as asphalt coating, and 12.3% of OG's input comes from CO mainly from engineering projects to enable extraction such as perforation and the construction of pipelines and their maintenance.

At resolution $\delta_2^{\text{NR}} = 0.887$ this cluster grows by the simultaneous incorporation of the support service sector AS and the professional service sector MP. These sectors join due to the loop [AS, MP, CO, OG, PC, AS]. The three new edges in this loop that involve the new sectors are the ones from PC to AS, from AS to MP and from MP to CO. Of all the economic input to AS, 13.4% comes from the PC sector in the form of, e.g., fuel for the transportation of manpower. Of MP's input, 11.3% comes from AS – given by $A_I(\text{AS}, \text{MP}) = 0.887$ – corresponding to administrative and support services hired by the MP sector for the correct delivery of MP's professional services and in the input to CO a total of 12.8% comes from MP from, e.g., architecture and consulting services for the construction.

We then see the incorporation of the rental service sector RL and 'Wholesale trade' (WH) to the five-node cluster at resolution $\delta_3^{\text{NR}} = 0.895$ given by the loop [WH, RL, OG, PC, AS, MP, WH]. To be more precise, the sector RL joins the main cluster by the aforementioned loop and by another one excluding WH, i.e. [RL, OG, PC, AS, MP, RL]. The formation of both loops is simultaneous since the last edge to appear is the one going from RL to OG at resolution $A_I(\text{RL}, \text{OG}) = \delta_3^{\text{NR}} = 0.895$. This implies that from OG's inputs, 10.5% comes from RL from, e.g., rental and leasing of generators, pumps, welding equipment and other machinery for extraction. The other edges depicted in the cluster at resolution δ_3^{NR} that complete the two mentioned loops are the ones from MP to RL, from MP to WH, and from WH to RL. These edges are associated with the corresponding dissimilarities $A_I(\text{MP}, \text{RL}) = 0.886$, $A_I(\text{MP}, \text{WH}) = 0.836$, and $A_I(\text{WH}, \text{RL}) = 0.894$, all of them less than δ_3^{NR}.

At resolution $\delta_4^{\text{NR}} = 0.900$ the financial sectors SC and FR join this cluster due to the path [SC, FR, RL, OG, PC, AS, SC]. Analogous to RL's merging at resolution δ_3^{NR}, the sector FR merges the main cluster by the aforementioned loop and by the one excluding SC, i.e., [FR, RL, OG, PC, AS, FR]. Both paths are formed simultaneously since the last edge to appear is the one from FR to RL at resolution $A_I(\text{FR}, \text{RL}) = \delta_4^{\text{NR}} = 0.900$. This means that from RL's inputs, 10% comes from FR.

The sole exceptions to this pattern of progressive agglutination are the pairings of the farms FA and the food products FB sectors at resolution $\delta = 0.893$ and the textile mills TE and apparel products AP sectors at resolution $\delta = 0.938$.

The nonreciprocal clustering method $\mathcal{H}^{\text{NR}}$ detects cyclic influences which, in general, lead to clusters that are more reasonable than those requiring the bidirectional influence that defines the reciprocal method $\mathcal{H}^{\text{R}}$. E.g., $\mathcal{H}^{\text{NR}}$ merges OG with PC at resolution $\delta = 0.885$ before they merge with the insurance sector IC at resolution $\delta = 0.923$. As we had already noted in the last paragraph of the preceding section, $\mathcal{H}^{\text{R}}$ merges OG with IC before their common joining with PC. However, the preponderance of cyclic influences in the network

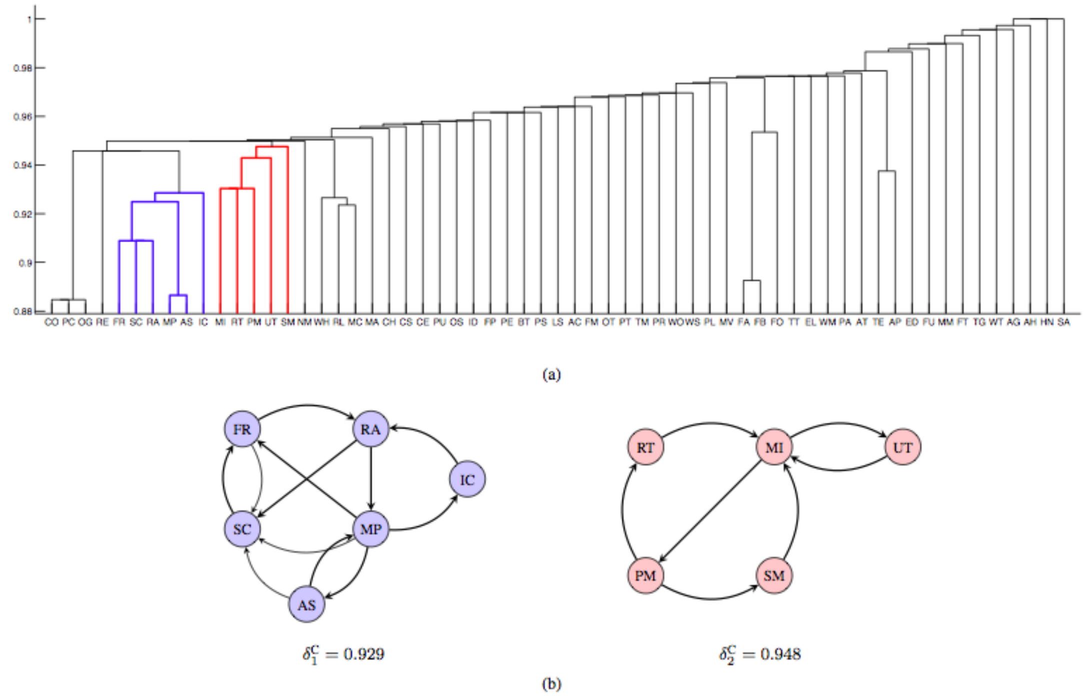


Figure 6.7: (a) Cyclic dendrogram. Output of the cyclic clustering method $\mathcal{H}^{\circlearrowleft_3}$ when applied to the network N_I. Two clusters formed at resolution $\delta_1^{\mathrm{C}} = 0.929$ and $\delta_2^{\mathrm{C}} = 0.948$ are highlighted in blue and red, respectively. (b) Highlighted clusters. Directed edges between sectors imply unidirectional influence between them at the corresponding resolution. Cyclic influences can be observed.

of economic interactions N_I leads to the formation of clusters that look more like artifacts than fundamental features. E.g., the cluster that forms at resolution $\delta_2^{\mathrm{NR}} = 0.887$ has AS and MP joining the three-node cluster CO-PC-OG because of an influence cycle of five nodes composed of [AS, MP, CO, OG, PC, AS]. From our discussion above, it is thus apparent that allowing clusters to be formed by arbitrarily long cycles overlooks important bidirectional influences between co-clustered nodes. If we wanted a clustering method which at resolution $\delta_2^{\mathrm{NR}} = 0.887$ would cluster the nodes PC, CO, and OG into one cluster and AS and MP into another cluster, we should allow influence to propagate through cycles of at most three or four nodes. A family of methods that permits this degree of flexibility is the family of cyclic methods $\mathcal{H}^{\circlearrowleft_t}$ that we discussed in Section 5.2.4 and whose application we exemplify next.

Cyclic clustering $\mathcal{H}^{\circlearrowleft_3}$ of N_I

The outcome of applying the cyclic clustering method $\mathcal{H}^{\circlearrowleft_3}$ defined in Section 5.2.4 to the network N_I is computed with the formula in (5.48). The resulting output dendrogram is shown in Fig. 6.7(a). Two clusters generated at resolutions $\delta_1^{\mathrm{C}} = 0.929$ and $\delta_2^{\mathrm{C}} = 0.948$ are highlighted in blue and red, respectively. These clusters are depicted in Fig. 6.7(b) with directed edges between the nodes representing dissimilarities less than or equal to the corresponding resolution. E.g., for the cluster generated at resolution $\delta_1^{\mathrm{C}} = 0.929$ (blue), we draw an edge from sector i to sector i' if and only if $A_I(i, i') \leq \delta_1^{\mathrm{C}}$. Comparing the cyclic dendrogram in Fig. 6.7(a) with the reciprocal and nonreciprocal dendrograms in Figs. 6.5(a) and 6.6(a), we observe that cyclic clustering merges any pair of sectors into a cluster at a resolution not higher than the resolution at which they are co-clustered by reciprocal clustering and not lower than the one at which they are co-clustered by nonreciprocal clustering. E.g., the sectors of construction CO and 'Fabricated metal products' (FM) become part of the same cluster at resolution $\delta_{\mathrm{R}} = 0.980$ in the reciprocal dendrogram, at resolution $\delta_{\mathrm{C}} = 0.964$ in the cyclic dendrogram and at resolution $\delta_{\mathrm{NR}} = 0.912$ in the nonreciprocal dendrogram, satisfying $\delta_{\mathrm{NR}} \leq \delta_{\mathrm{C}} \leq \delta_{\mathrm{R}}$. The inequalities described among the merging resolutions need not be strict as in the previous example, e.g., the farms (FA) sector merges with the food products FB sector at resolution $\delta = 0.893$ for the reciprocal, nonreciprocal and cyclic clustering methods. This ordering of the merging resolutions is as it should be since the reciprocal and nonreciprocal ultrametrics uniformly bound the output ultrametric of any clustering method satisfying the Axioms of Value and Transformation such as the cyclic clustering method [cf. (3.15)].

The cyclic clustering method $\mathcal{H}^{\circlearrowleft_3}$ allows reasonable cyclic influences and is insensitive to intricate influences described by long cycles. As we pointed out in the two preceding subsections, $\mathcal{H}^{\mathrm{R}}$ does not recognize the obvious relation between the sectors oil and gas extraction OG and the petroleum products PC sectors because it requires direct bidirectional influence whereas $\mathcal{H}^{\mathrm{NR}}$ merges OG and PC at a low resolution but also considers other counter-intuitive cyclic influence structures represented by long loops such as the merging of the service sectors AS and MP with the cluster OG-PC-CO [cf. Fig. 6.6]. The cyclic method $\mathcal{H}^{\circlearrowleft_3}$ combines the desirable features of the reciprocal and nonreciprocal methods. Indeed, as can be seen from the cyclic dendrogram in Fig. 6.7(a), $\mathcal{H}^{\circlearrowleft_3}$ recognizes the heavy industry cluster OG-PC-CO since these three sectors are the first to merge at resolution $\delta = 0.885$. However, the service sectors MP and AS do not merge first with the heavy industry cluster. Instead, they become part of a service cluster formed at resolution $\delta_1^{\mathrm{C}} = 0.929$ and depicted in blue in Fig. 6.7. To be more precise, MP and AS merge at resolution $\delta = 0.887$ due to the bidirectional influence between them. This resolution coincides with the first merging in the reciprocal dendrogram [cf. Fig. 6.5(a)]. At resolution $\delta = 0.909$ the credit intermediation sector FR, the investment sector SC and the real estate sector RA form a three-node cluster given by the influence cycle [RA, SC, FR, RA]. Of all the economic input to SC, 9.1% comes from the RA sector – which is represented by the dissimilarity $A_I(\mathrm{RA}, \mathrm{SC}) = 0.909$ – in the form of, e.g., leasing services related to real estate investment trusts. The sector SC provides 24.1% of FR's input whereas FR represents

35.1% of RA's input. We interpret the relation among these three sectors as follows: the credit intermediation sector FR acts as a vehicle to connect the investments sector SC with the sector that attracts investments RA. When we increase the resolution, at $\delta = 0.925$ the financial and real estate services cluster FR-SC-RA joins the professional and support services cluster MP-AS due to the three-node loop [MP, FR, RA, MP]. We have already explained the relation between FR and RA. The other two edges in the loop correspond to the fact that 7.5% of MP's input comes from RA due to, e.g., service companies renting a location to operate and to the fact that 11.3% of FR's input comes from MP due to, e.g., hiring of financial consulting services. Finally, at resolution $\delta_1^{\mathrm{C}} = 0.929$ the insurance service sector IC merges with the five-node cluster due to the loop [IC, RA, MP, IC]. The two new edges in this loop are the ones from IC to RA and from MP to IC. The insurance sector IC provides 8.6% of RA's input in the form of insurance policies for real estate assets. The professional service sector MP provides 7.1% of IC's input in the form of, e.g., risk assessment consulting. Further, notice that in the reciprocal dendrogram in Fig. 6.6(a) the six service sectors MP-AS-RA-FR-SC-IC form a separate cluster at resolution $\delta = 0.958$. The fact that $\delta_1^{\mathrm{C}} = 0.929 < 0.958$ implies that cyclic influences among the service sectors are more meaningful than corresponding direct bidirectional influences.

The method $\mathcal{H}^{\circlearrowright_3}$, being represented by a three-node cycle $\circlearrowright_3$, captures cyclic interactions among the three sectors of the economy: extraction of raw materials, manufacturing and services. As an example, consider the cluster formed at resolution $\delta_2^{\mathrm{C}} = 0.948$ (red) in Fig. 6.7(b). In the generation of this cluster, there is first a three-node merging between the sectors of mining MI, primary metals PM and rail transportation RT at resolution $\delta = 0.930$ due to the cycle [MI, PM, RT, MI]. In this cycle, there is an influence from the primary sector of extraction of raw materials MI to the secondary sector of manufacturing PM to the tertiary sector of transportation service RT. Indeed, as explained in the reciprocal clustering subsection, 9.3% of PM's input comes from MI , e.g., in the form of ores for metal production and 7.0% of RT's input comes from PM in part due to the metal needed for the construction of railroads. The cycle closes due to the influence of rail transportation RT on mining MI – 7.0% of MI's input comes from RT – that can be attributed to the use of rail transportation for the movement of materials. This cyclic influence seems to be a more natural way of relating the three sectors than the direct bidirectional influence required by reciprocal clustering [cf. Fig. 6.5(b)]. At resolution $\delta = 0.943$, the utilities UT sector joins the cluster due to its bidirectional influence with the mining sector like in the reciprocal case. At resolution $\delta_2^{\mathrm{C}} = 0.948$ the sector 'Support activities for mining' (SM) merges with the rest of the red cluster. This last merging also occurs due to the presence of a three-node cycle formed by [MI, PM, SM, MI]. In this case, we also have influence from the primary sector MI to the secondary sector PM to the tertiary sector SM. Indeed, besides the already mentioned influence from MI to PM, we have that 11.0% of SM's input comes from PM and 5.2% of MI's input comes from SM. The influence from SM to MI closing the cycle is expected since the sector SM has as its main goal to provide support services to the mining industry.

Unilateral clustering $\mathcal{H}^{\mathrm{U}}$ of N_I

The outcome of applying the unilateral clustering method $\mathcal{H}^{\mathrm{U}}$ defined in (3.62) to the network N_I is computed with the algorithmic formula in (3.93). The resulting output dendrogram is shown in Fig. 6.8(a). Four clusters appearing at resolutions $\delta_1^{\mathrm{U}} = 0.775$, $\delta_2^{\mathrm{U}} = 0.831$, $\delta_3^{\mathrm{U}} = 0.854$, and $\delta_4^{\mathrm{U}} = 0.883$ are highlighted in blue, red, orange, and green, respectively. In Fig. 6.8(b) we explicit the highlighted clusters and draw a directed edge between two nodes if and only if the dissimilarity between them is less than or equal to the corresponding resolution at which the clusters are formed. E.g., for the cluster generated at resolution $\delta_1^{\mathrm{U}} = 0.775$ (blue), we draw an edge from sector i to sector i' if and only if $A_I(i, i') \leq \delta_1^{\mathrm{U}}$. Unidirectional influence is enough for clusters to form when applying the unilateral clustering method.

The asymmetry of the original network N_I is put in evidence by the difference between the unilateral dendrogram in Fig. 6.8(a) and the reciprocal dendrogram in Fig. 6.5(a). The last merging in the unilateral dendrogram, i.e. when 'Waste management and remediation services' (WM) joins the main cluster, occurs at $\delta = 0.923$. If, in turn, we cut the reciprocal dendrogram at this resolution, we observe 57 singleton clusters and two pairs of nodes merged together. Recall that if the original network is symmetric, the unilateral and the reciprocal dendrograms must coincide and so must every other method satisfying the agnostic set of axioms in Section 3.4.2 [cf. (3.70)]. Thus, the observed difference between the dendrograms is a manifestation of asymmetries in the network N_I.

Unilateral clustering detects intense one-way influences between sectors. The first two sectors to be merged into a single cluster by the unilateral clustering method $\mathcal{H}^{\mathrm{U}}$ are the financial sectors SC and 'Funds, trusts, and other financial vehicles' (FT) at a resolution $\delta = 0.132$. This occurs because 86.8% of FT's input comes from SC, corresponding to $A_I(\mathrm{SC}, \mathrm{FT}) = 0.132$ the smallest positive dissimilarity in the network N_I. The strong influence of SC over FT is expected since FT is comprised of entities organized to pool securities coming from the SC sector. The next merging when increasing the resolution occurs at $\delta = 0.174$ between oil and gas extraction OG and petroleum and coal products PC since 82.6% of PC's input comes from OG mainly as crude oil for refining. The following three mergings correspond to sequential additions to the OG-PC cluster of the utilities UT, 'Water transportation' (WT), and 'Air transportation' (AT) sectors at resolution $\delta = 0.428$, $\delta = 0.482$, and $\delta = 0.507$, respectively. These mergings occur because 57.2% of UT's input

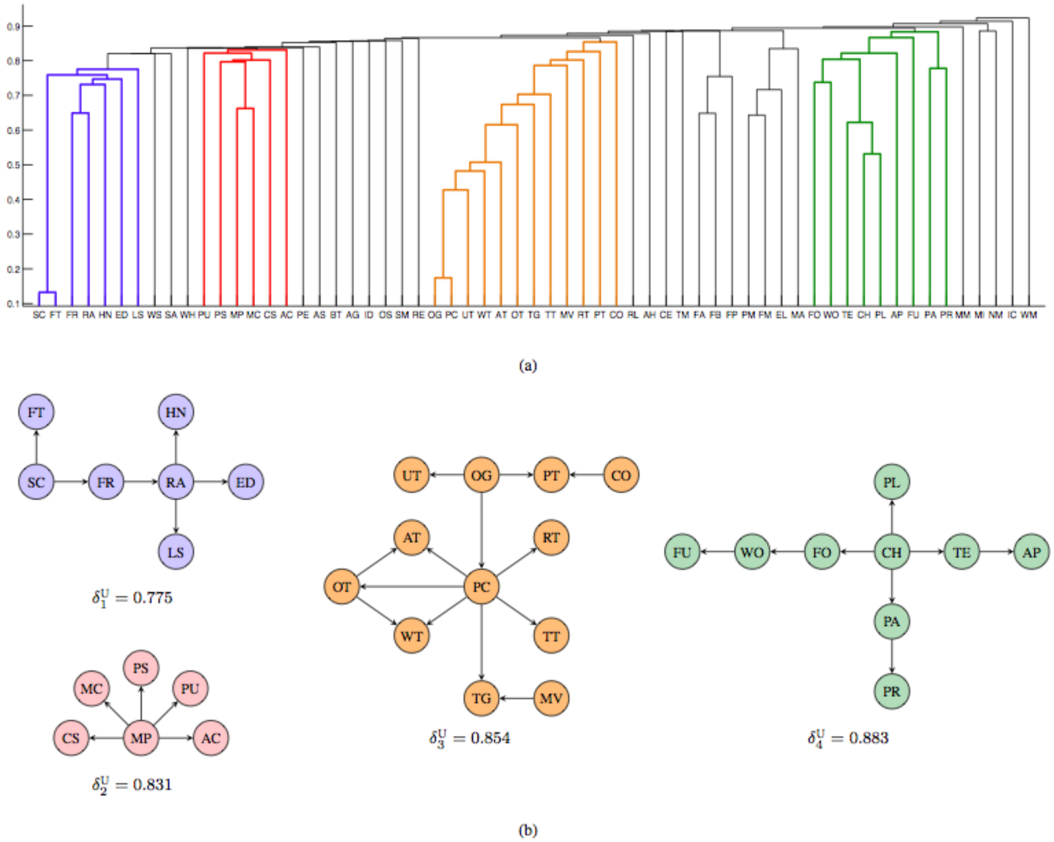


Figure 6.8: (a) Unilateral dendrogram. Output of the unilateral clustering method $\mathcal{H}^{\mathrm{U}}$ when applied to the network N_I. Four clusters formed at resolutions $\delta_1^{\mathrm{U}} = 0.775$, $\delta_2^{\mathrm{U}} = 0.831$, $\delta_3^{\mathrm{U}} = 0.854$, and $\delta_4^{\mathrm{U}} = 0.883$ are highlighted in blue, red, orange, and green, respectively. (b) Highlighted clusters. Directed edges between sectors imply unidirectional influence between them at the corresponding resolution. Cycles are not required for the formation of clusters due to the definition of unilateral clustering $\mathcal{H}^{\mathrm{U}}$.

comes from OG in the form of natural gas for both distribution and fuel for the generation of electricity and for the transportation sectors WT and AT, 51.8% and 49.3% of the respective inputs come from PC as the provision of liquid fuel.

Unilateral clusters tend to form around sectors of intense output. This observation is analogous to the formation of clusters around populous states, hence with intense population movement, that we observed in Section 6.1. Indeed, if for each sector we evaluate the commodity intermediate value in dollars, i.e. the total output not destined to final uses, the professional service MP sector achieves the maximum followed by, in decreasing order, the sectors RA, OG, FR, AS and 'Chemical products' (CH). These top sectors are composed of massively demanded services like professional, support, real estate and financial services plus the core activities of two important industries, namely oil & gas and chemical products. Of these top six sectors, five are contained in the four clusters highlighted in Fig. 6.8(b), with every cluster containing at least one of these sectors and the cluster formed at resolution $\delta_1^{\mathrm{U}} = 0.775$ (blue) containing two, FR and RA. These clusters of intense output have influence, either directly or indirectly, over most of the sectors in their same cluster. E.g., in the cluster formed at resolution $\delta_2^{\mathrm{U}} = 0.831$ (red) in Fig. 6.8(b) there is a directed edge from MP to every other sector in the cluster. This occurs because MP provides professional and technical services that represent, in decreasing order, 33.8%, 20.3%, 19.8%, 17.8%, and 16.9% of the input to the sectors of management of companies MC, 'Motion picture and sound recording industries' (PS), 'Computer systems design and related services' (CS), 'Publishing industries' (PU), and 'Accommodation' (AC), respectively. Consequently, in the unilateral clustering we can observe the MP sector merging with MC at resolution $\delta = 0.662$ followed by a sequential merging of the remaining singleton clusters, i.e. PS at $\delta = 0.797$, CS at $\delta = 0.802$, PU at $\delta = 0.822$ and finally AC joins at resolution $\delta_2^{\mathrm{U}} = 0.831$. As another example consider the cluster formed at resolution $\delta_4^{\mathrm{U}} = 0.833$ (green) containing the influential sector CH. Its influence over four different industries, namely plastics, apparel, paper and wood, is represented by the four directed branches leaving from CH in Fig. 6.8(b). The sector CH first merges with 'Plastics and rubber products' (PL) at resolution $\delta = 0.531$ because 46.9% of PL's input comes from CH as materials needed for the handling and manufacturing of plastics. The textile mills TE sector then merges at resolution $\delta = 0.622$ because 37.8% of TE's input comes from CH as dyes and other chemical products for the fabric manufacturing. At resolution $\delta = 0.804$ the previously formed cluster composed of the forestry FO and wood products WO sectors join the CH-PL-TE cluster due to the dependence of FO on CH for the provision of chemicals for soil treatment and pest control. At resolution $\delta = 0.822$, the apparel sector AP joins the main cluster due to its natural dependence on the fabrics generated by TE. Indeed, 17.8% of AP's input comes from TE. In a similar way, at resolution $\delta = 0.867$, 'Furniture and related products' (FU) joins the cluster due to the influence from the WO sector. Finally, at resolution $\delta_4^{\mathrm{U}} = 0.833$, the previously clustered paper industry comprised of the sectors 'Paper products' (PA) and 'Printing and related support activities' (PR) joins the main cluster due to the intense utilization of chemical products in the paper manufacturing process.

Directed single linkage quasi-clustering $\tilde{\mathcal{H}}^*$ of N_I

The outcome of applying the directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$ with output quasi-ultrametrics defined in (2.7) to the network N_I is computed with the algorithmic formula in (4.22). In Fig. 6.9 we present four quasi-partitions of the output quasi-dendrogram $\tilde{D}_I^* = (D_I^*, E_I^*)$ focusing on ten economic sectors. We limit the view of the quasi-partitions – which were computed for the whole network – to ten sectors to facilitate the interpretation. These ten sectors are the first to cluster in the dendrogram component D_I^* of the quasi-dendrogram $\tilde{D}_I^*$. To see this, recall that from Proposition 12 we have that $D_I^* = \mathcal{H}^{\rm NR}(N_I)$, i.e. the dendrogram component D_I^* coincides with the output dendrogram of applying the nonreciprocal clustering method to the network N_I. Hence, the ten sectors depicted in the quasi-partitions in Fig. 6.9 coincide with the ten leftmost sectors in the dendrogram in Fig. 6.6(a). We present quasi-partitions $\tilde{D}_I^*(\delta)$ for four different resolutions $\delta_1^* = 0.884$, $\delta_2^* = 0.886$, $\delta_3^* = 0.894$, and $\delta_4^* = 0.899$. These resolutions are 0.001 smaller than the first four merging resolutions in the dendrogram component D_I^* or, equivalently, in the nonreciprocal dendrogram [cf. Fig. 6.6(b)].

The edge component E_I^* of the quasi-dendrogram $\tilde{D}_I^*$ captures the asymmetric influence between clusters. E.g. in the quasi-partition in Fig. 6.9 for resolution $\delta_1^* = 0.884$ every cluster is a singleton since the resolution is smaller than that of the first merging. However, the influence structure reveals an asymmetry in the dependence between the economic sectors. At this resolution the professional service sector MP has influence over every other sector except for the rental services RL as depicted by the eight arrows leaving the MP sector. No sector has influence over MP at this resolution since this would imply, except for RL, the formation of a non-singleton cluster. The influence of MP reaches primary sectors as OG, secondary sectors as PC and tertiary sectors as AS or SC. The versatility of MP's influence can be explained by the diversity of services condensed in this economic sector, e.g. civil engineering and architectural services are demanded by CO, production engineering by PC and financial consulting by SC. For the rest of the influence pattern, we can observe an influence of CO over OG mainly due to the construction and maintenance of pipelines, which in turn influences PC due to the provision of crude oil for refining. Thus, from the transitivity (QP2) property of quasi-partitions introduced in Chapter 4 we have an influence edge from CO to PC. The sectors CO, PC and OG influence the support service sector AS. Moreover, the service sectors RA, SC and FR have a totally hierarchical influence structure where SC has influence over the other two and FR has influence over RA. Since these three nodes remain as singleton clusters for the resolutions studied, the influence structure described is preserved for higher resolutions as it should be from the influence hierarchy property of the the edge set $E_S^*(\delta)$ stated in condition ($\tilde{\rm D}$3) in the definition of quasi-dendrogram in Section 4.1.

At resolution $\delta_2^* = 0.886$, we see that the sectors OG-PC-CO have formed a three-node cluster depicted in red that influences AS. At this resolution, the influence edge from MP to RL appears and, thus, MP gains influence over every other cluster in the quasi-partition including the three-node cluster. At resolution $\delta = 0.887$ the service sectors AS and MP join the cluster OG-PC-CO and for $\delta_3^* = 0.894$ we have this five-node cluster influencing the other five singleton clusters plus the mentioned hierarchical structure among SC, FR, and RA and an influence edge from WH to RL. When we increase the resolution to $\delta_4^* = 0.899$ we

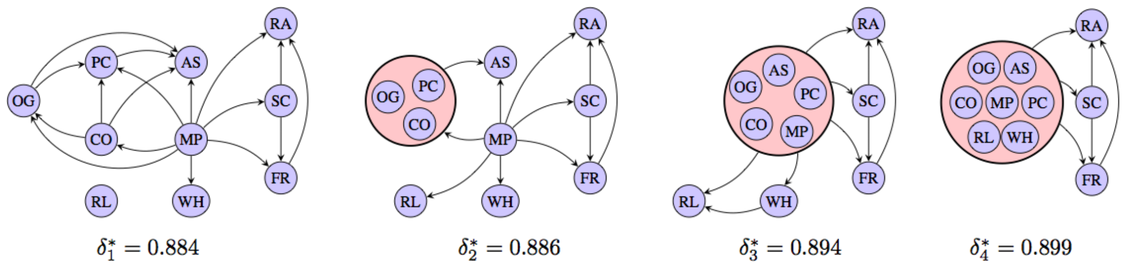


Figure 6.9: Directed single linkage quasi-clustering method applied to a portion of the sectors of the economy. Quasi-partitions shown for resolutions 0.001 smaller than the first four merging resolutions in the dendrogram component D_I^* of the quasi-dendrogram $\tilde{D}_I^*$. The edges define a partial order among the blocks of every quasi-partition.

see that RL and WH have joined the main cluster that influences the other three singleton clusters. If we keep increasing the resolution, we would see at resolution $\delta = 0.900$ the sectors SC and FR joining the main cluster which would have influence over RA the only other cluster in the quasi-partition. Finally, at resolution $\delta = 0.909$ RA joins the main cluster and the quasi-partition contains only one block.

The influence structure between clusters at any given resolution defines a partial order. More precisely, for every resolution δ, the edge set $E_I^*(\delta)$ defines a partial order between the blocks given by the partition $D_I^*(\delta)$. We can use this partial order to evaluate the relative importance of different clusters by stating that more important sectors have influence over less important ones. E.g., at resolution $\delta_1^* = 0.884$ we have that MP is more important than every other sector except for RL, which is incomparable at this resolution. There are three totally ordered paths that have MP as the most important sector at this resolution. The first one contains five sectors which are, in decreasing order of importance, MP, CO, OG, PC, and AS. The second one is comprised of MP, SC, FR, and RA and the last one only contains MP and WH. At resolution $\delta_2^* = 0.886$ we observe that the three-node cluster OG-PC-CO, although it contains more nodes than any other cluster, it is not the most important of the quasi-partition. Instead, the singleton cluster MP has influence over the three-node cluster and, on top of that, is comparable with every other cluster in the quasi-partition. From resolution $\delta_3^* = 0.894$ onwards, after MP joins the red cluster, the cluster with the largest number of nodes coincides with the most important of the quasi-partition. At resolution $\delta_4^* = 0.899$ we have a total ordering among the four clusters of the quasi-partition. This is not true for the other three depicted quasi-partitions.

As a further illustration of the quasi-clustering method $\tilde{\mathcal{H}}^*$, we apply it to the network $N_C = (C, A_C)$ of consolidated industrial sectors of year 2011 where $|C| = 14$ – as given by the Bureau of Economic Analysis; see Table 6.2 – instead of the original 61 sectors. To generate the dissimilarity function A_C from the similarity data we use (6.2). The

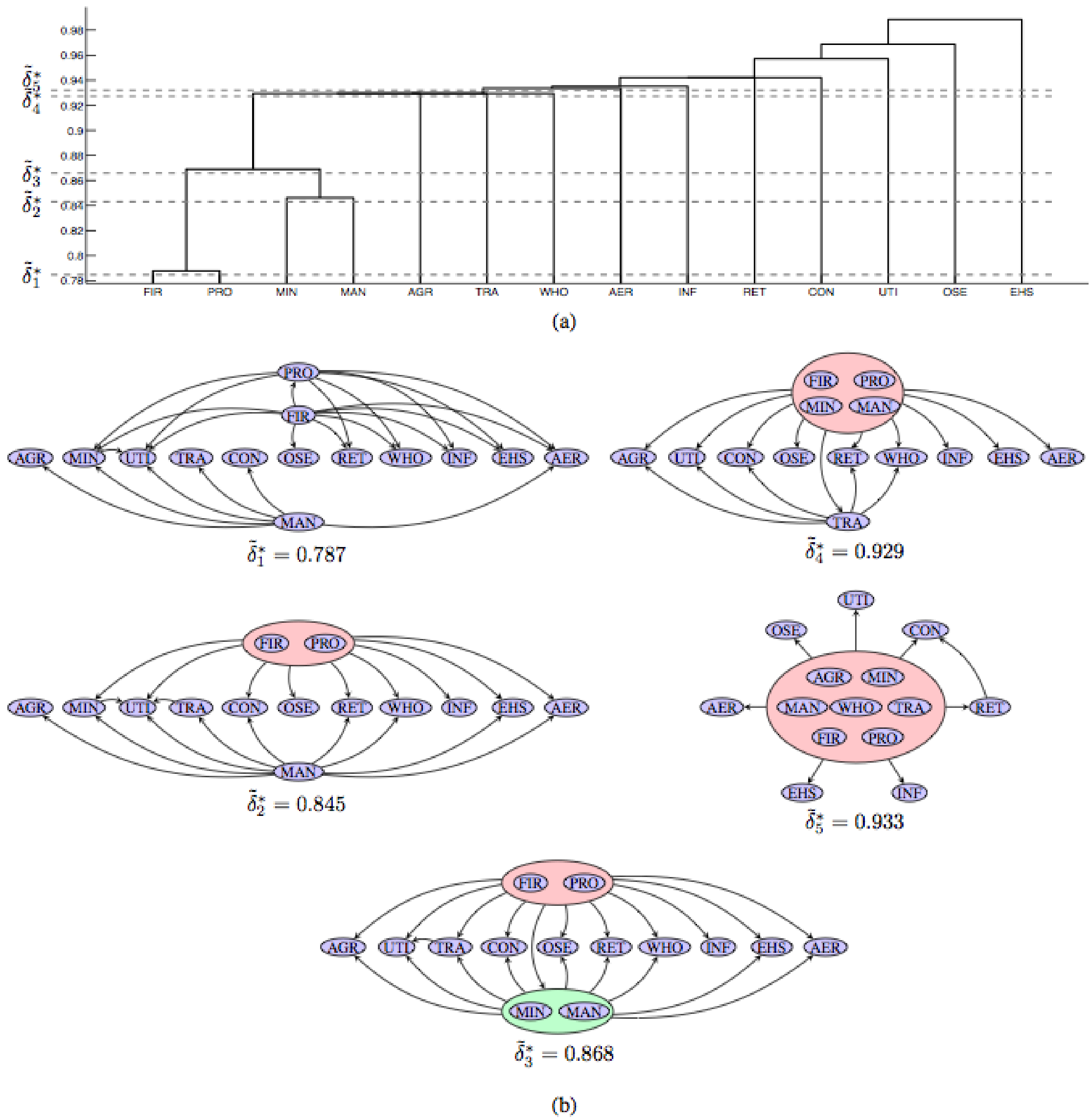


Figure 6.10: (a) Dendrogram component D_C^* of the quasi-dendrogram $\tilde{D}_C^* = (D_C^*, E_C^*)$. Output of the directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$ when applied to the network N_C. (b) Quasi-partitions. Given by the specification of the quasi-dendrogram $\tilde{D}_C^*$ at a particular resolution $\tilde{D}_C^*(\tilde{\delta}_k^*)$ for $k = 1, \ldots, 5$.

Table 6.2: Code and description of consolidated industrial sectors

Code	Consolidated Industrial Sector	Code	Consolidated Industrial Sector
AER	Arts, entertain., accomm., and food serv.	AGR	Agricult., forestry, fishing, and hunting
CON	Construction	EHS	Education, health care, and social assis.
FIR	Finance, insurance, real estate, and rental	INF	Information
MAN	Manufacturing	MIN	Mining
OSE	Other services, except government	PRO	Professional and business services
RET	Retail trade	TRA	Transportation and warehousing
UTI	Utilities	WHO	Wholesale trade

outcome of applying the directed single linkage quasi-clustering method $\tilde{\mathcal{H}}^*$ with output quasi-ultrametrics defined in (2.7) to the network N_C is computed with the algorithmic formula in (4.22). Of the output quasi-dendrogram $\tilde{D}_C^* = (D_C^*, E_C^*)$, in Fig. 6.10(a) we show the dendrogram component D_C^* and in Fig. 6.10(b) we depict the quasi-partitions $\tilde{D}_C^*(\tilde{\delta}_i^*)$ for $\tilde{\delta}_1^* = 0.787$, $\tilde{\delta}_2^* = 0.845$, $\tilde{\delta}_3^* = 0.868$, $\tilde{\delta}_4^* = 0.929$, and $\tilde{\delta}_5^* = 0.933$, corresponding to resolutions 0.001 smaller than mergings in the dendrogram D_C^*. The reason why we use the consolidated network N_C is to facilitate the visualization of quasi-partitions that capture every sector of the economy instead of only ten particular sectors as in the previous application.

The quasi-dendrogram $\tilde{D}_C^*$ captures the asymmetric influences between clusters of industrial sectors at every resolution. E.g., at resolution $\tilde{\delta}_1^* = 0.787$ the dendrogram D_C^* in Fig. 6.10(a) indicates that every industrial sector forms its own singleton cluster. However, this simplistic representation, characteristic of clustering methods, ignores the asymmetric relations between clusters at resolution $\tilde{\delta}_1^*$. These influence relations are formalized in the quasi-dendrogram $\tilde{D}_C^*$ with the introduction of the edge set $E_C^*(\delta)$ for every resolution δ. In particular, for $\tilde{\delta}_1^*$ we see in Fig. 6.10(b) that the sectors of 'Finance, insurance, real estate, rental, and leasing' (FIR) and 'Manufacturing' (MAN) combined have influence over the remaining 12 sectors. More precisely, the influence of FIR is concentrated on the service and commercialization sectors of the economy whereas the influence of MAN is concentrated on primary sectors, transportation, and construction. Furthermore, note that due to the transitivity (QP2) property of quasi-partitions defined in Chapter 4, the influence of FIR over 'Professional and business services' (PRO) implies influence of FIR over every sector influenced by PRO. The influence among the remaining 11 sectors, i.e. excluding MAN, FIR and PRO, is minimal, with the 'Mining' (MIN) sector influencing the 'Utilities' (UTI) sector. This influence is promoted by the influence of the 'Oil and gas extraction' (OG) subsector of MIN over the utilities sector as observed in the cluster formed at resolution $\delta_3^{\mathrm{U}} = 0.854$ (orange) by the unilateral clustering method [cf. Fig. 6.8(b)]. At resolution $\tilde{\delta}_2^* = 0.845$, FIR and PRO form one cluster, depicted in red, and they add an influence to the 'Construction' (CON) sector apart from the previously formed influences that must persist due to the influence hierarchy property of the the edge set $E_C^*(\delta)$ stated in condition ($\tilde{\text{D}}3$) in the definition of quasi-dendrogram in Section 4.1. The manufacturing sector also intensifies its influences by reaching the commercialization sectors 'Retail trade' (RET) and 'Wholesale trade' (WHO) and the service sector 'Educational services, health care, and social assistance' (EHS). The influence among the rest of the sectors is still scarce with the only addition of the influence of 'Transportation and warehousing' (TRA) over UTI. At resolution $\tilde{\delta}_3^* = 0.868$ we see that mining MIN and manufacturing MAN form their own cluster, depicted in green. The previously formed red cluster has influence over every other cluster in the quasi-partition, including the green one. At resolution $\tilde{\delta}_4^* = 0.929$, the red and green clusters become one, composed of four original sectors. Also, the influence of the transportation TRA sector over the rest is intensified with the appearance of edges to the primary sector 'Agriculture, forestry, fishing, and hunting' (AGR), the construction CON sector and the commercialization sectors RET and WHO. Finally, at resolution $\tilde{\delta}_5^* = 0.933$ there is one clear main cluster depicted in red and composed of seven sectors spanning the primary, secondary, and tertiary segments of the economy. This main cluster influences every other singleton cluster. The only other influence in the quasi-partition $\tilde{D}_C^*(0.933)$ is the one of RET over CON. For increasing resolutions, the singleton clusters join the main red cluster until at resolution $\delta = 0.988$ the 14 sectors form one single cluster.

The influence structure at every resolution induces a partial order in the blocks of the corresponding quasi-partition. As done in previous examples, we can interpret this partial order as an ordering of relative importance of the elements within each block. E.g., we can say that at resolution $\tilde{\delta}_1^* = 0.787$, MAN is more important that MIN which in turn is more important than UTI which is less important that PRO. However, PRO and MAN are not comparable at this resolution. At resolution $\tilde{\delta}_4^* = 0.929$, after the red and green clusters have merged together at resolution $\delta = 0.869$, we depict the combined cluster as red. This representation is not arbitrary, the red color of the combined cluster is inherited from the most important of the two component cluster. The fact that the red cluster is more important than the green one can be seen from the edge from the former to the latter in the quasi-partition at resolution $\tilde{\delta}_3^*$. In this sense, the edge component E_C^* of the quasi-dendrogram formally provides a hierarchical structure between clusters at a fixed resolution apart from the hierarchical structure across resolutions given by the dendrogram component D_C^* of the quasi-dendrogram. E.g., if we focus only on the dendrogram D_C^* in Fig. 6.10(a), the nodes MIN and MAN seem to play the same role. However, when looking at the quasi-partitions at resolutions $\tilde{\delta}_1^*$ and $\tilde{\delta}_2^*$, it follows that MAN has influence over a larger set of nodes than MIN and hence plays a more important role in the clustering for increasing resolutions. Indeed, if we delete the three nodes with the strongest influence structure, namely PRO, FIR, and MAN, and apply the quasi-clustering method $\tilde{\mathcal{H}}^*$ on the remaining 11 nodes, the first merging occurs between the mining MIN and utilities UTI sectors at $\delta = 0.960$. At this same resolution, in the original dendrogram component in Fig. 6.10(a), a main cluster composed of 12 nodes only excluding 'Other services, except government' (OSE) and EHS is formed. This indicates that by removing influential sectors of the economy, the tendency to co-cluster of the remaining sectors is substantially decreased.

Chapter 7

Taxonomy of hierarchical clustering in asymmetric networks

We have developed a theory for hierarchically clustering asymmetric – weighted and directed – networks. Starting from the realization that generalizing methods used to cluster metric data to asymmetric networks is not always intuitive, we defined simple reasonable properties and proceeded to characterize the space of methods that are admissible with respect to them. The properties that we have considered are the following:

(A1) Axiom of Value. In a network with two nodes, the output dendrogram consists of two singleton clusters for resolutions smaller than the maximum of the two dissimilarities and a single two-node cluster for larger resolutions.

(A1') Extended Axiom of Value. Define a canonical asymmetric network of n nodes in which the two directed dissimilarities – which might be different from each other – are the same for any pair of nodes. The output dendrogram consists of n singleton clusters for resolutions smaller than the maximum of the two intervening dissimilarities and, consists of a single n-node cluster for larger resolutions.

(A1") Alternative Axiom of Value. In a network with two nodes, the output dendrogram consists of two singleton clusters for resolutions smaller than the minimum of the two intervening dissimilarities, and consists of a single two-node cluster for larger resolutions.

(A1"') Agnostic Axiom of Value. In a network with two nodes, the output dendrogram consists of two singleton clusters for resolutions smaller than the minimum of the two intervening dissimilarities, and consists of a single two-node cluster for resolutions larger than their maximum.

(A2) Axiom of Transformation. Consider two given networks N and M and a dissimilarity reducing map from the nodes of N to the nodes of M, i.e. a map such that dissimilarities between the image nodes in M are smaller than or equal to the corre-

sponding dissimilarities of the pre-image nodes in N. Then, the resolution at which any two nodes merge into a common cluster in the network M is smaller than or equal to the resolution at which their pre-images merge in the network N.

(P1) Property of Influence. For any network with n nodes, the output dendrogram consists of n singleton clusters for resolutions smaller than the minimum loop cost of the network – the loop cost is the maximum directed dissimilarity when traversing the loop in a given direction, and the minimum loop cost is the cost of the loop of smallest cost.

(P1') Alternative Property of Influence. For any network with n nodes, the output dendrogram consists of n singleton clusters for resolutions smaller than the separation of the network – defined as the smallest positive dissimilarity across all pairs of nodes.

(P2) Scale Preservation. Consider two given networks N and M where the latter is constructed by transforming the dissimilarities in N by a nondecreasing function. Then, the resolution at which any two nodes cluster in network M can be obtained by applying the same nondecreasing function to the merging resolution of those two nodes in network N.

(P2') Linear Scale Preservation. A particular case of property (P2) where the dissimilarity transformations are restricted to nondecreasing linear functions.

(P3) Representability. There exists a collection of representers Ω representing an equivalent clustering method. Given Ω, the represented method links two nodes x and x' by mapping into the network linearly scaled versions of the representers with overlapping images such that both nodes belong to the image of some representer. The resolution at which x and x' merge in the output dendrogram equals the smallest linear scaling needed for the mapping from the scaled representers to the network to be dissimilarity reducing.

(P4) Excisiveness. The dendrogram obtained when clustering a cluster of a given network N coincides with the corresponding branch of the dendrogram obtained when the whole network is clustered.

(P5) Stability. For any two networks N and M, the generalized Gromov-Hausdorff distance between the corresponding output dendrograms is uniformly bounded by the generalized Gromov-Hausdorff distance between the networks.

Throughout this first part of the thesis we identified and described clustering methods satisfying different subsets of the above properties. Several methods were based on finding directed paths of minimum cost, where the path cost was defined as the maximum dissimilarity encountered when traversing the given path. The set of clustering methods that we have considered is comprised by the following:

Reciprocal. Nodes x and x' are clustered together at a given resolution δ if there exists a path linking x to x' such that the directed path costs are not larger than δ in either direction.

Nonreciprocal. Nodes x and x' are clustered together at a given resolution δ if there exist two paths, one linking x to x' and the other linking x' to x, such that both directed path costs are not larger than δ in either direction. In contrast to the reciprocal method, the paths linking x to x' and x' to x may be different.

Grafting. Grafting methods are defined by exchanging branches between the reciprocal and nonreciprocal dendrograms as dictated by an exogenous parameter β. Two grafting methods were studied. In both methods, the reciprocal dendrogram is sliced at resolution β. In the first method, the branches of resolution smaller than β are replaced by the corresponding branches of the nonreciprocal dendrogram. In the second method, the branches of resolution smaller than β are preserved and these branches merge either at resolution β or at the resolution given by the nonreciprocal dendrogram, whichever is larger.

Convex combinations. Given a network N and two clustering methods $\mathcal{H}_1$ and $\mathcal{H}_2$, denote by D_1 and D_2 the corresponding output dendrograms. Construct a symmetric network M so that the dissimilarities between any pair (x, x') is given by the convex combination of the minimum resolutions at which x and x' are clustered together in D_1 and D_2. Cluster the network M with the single linkage method to define a valid dendrogram.

Semi-reciprocal. A semi-reciprocal path of index $t \geq 2$ between two nodes x and x' is formed by concatenating directed paths of length at most t, called secondary paths, from x to x' and back. The nodes at which secondary paths in both directions concatenate must coincide, although the paths themselves might differ. Nodes x and x' are clustered together at a given resolution δ if they can be linked by a semi-reciprocal path of cost not larger than δ.

Algorithmic intermediate. Generalizes the semi-reciprocal clustering methods by allowing the maximum length t of secondary paths to be different in both directions.

Structure representable. A structure representer ω is a network with at least two nodes in which all its positive dissimilarities are equal to 1. A method is structure representable if it can be represented by a collection Ω of strongly connected structure representers ω.

Cyclic. The nth method in this family is a structure representable clustering method where the representer is a directed cycle network with $n > 2$ nodes and unit dissimilarities.

Unilateral. Consider the cost of an undirected path as one where the edge cost between two consecutive nodes is given by the minimum directed cost in both directions. Nodes x and x' are clustered together at a given resolution δ if there exists an undirected path linking x and x' of cost not larger than δ.

We can build a taxonomy of hierarchical clustering from the perspective of axioms and properties and an intertwined taxonomy from the perspective of clustering methods as we

Table 7.1: Hierarchical clustering methods and properties

		Reciprocal	**Non-reciprocal**	Grafting	Convex combs.	Semi-reciprocal	Algorithmic intermediate	**Structure represent.**	Cyclic	**Unilateral**
(A1)	**Axiom of Value**	x	x	x	x	x	x	x	x	
(A1')	Extended Axiom of Value	x	x	x	x	x	x	x	x	
(A1")	**Alt. Axiom of Value**									x
(A1"')	Agnostic Axiom of Value	x	x	x	x	x	x	x	x	x
(A2)	**Axiom of Transformation**	x	x	x	x	x	x	x	x	x
(P1)	Property of Influence	x	x	x	x	x	x	x	x	
(P1')	Alt. Property of Influence	x	x	x	x	x	x	x	x	x
(P2)	**Scale Preservation**	x	x			x	x	x	x	x
(P2')	Linear Scale Preservation	x	x		x	x	x	x	x	x
(P3)	Representability	x	x					x	x	x
(P4)	**Excisiveness**	x	x					x	x	x
(P5)	Stability	x	x			x	x	x	x	x

summarize in Table 7.1 and elaborate in the following sections.

7.1 Taxonomy of axioms and properties

The taxonomy from the perspective of axioms and properties is encoded in the rows in Table 7.1. For most of this first part of the thesis, the Axioms of value (A1) and transformation (A2) were requirement for admissibility. All of the methods enumerated above satisfy the Axiom of Transformation whereas all methods, except for unilateral clustering, satisfy the Axiom of Value. Although seemingly weak, (A1) and (A2) are a stringent source of structure. E.g., we showed that admissibility with respect to (A1) and (A2) is equivalent to admissibility with respect to the apparently stricter conditions given by the Extended Axiom of Value (A1') combined with (A2). Likewise, we showed that the Property of Influence (P1) is implied by (A1) and (A2). This latter fact can be interpreted as stating that the requirement of bidirectional influence in two-node networks combined with the Axiom of Transformation implies a requirement for loops of influence in all networks. Given that (A1') and (P1) are implied by (A1) and (A2) and that all methods except for unilateral clustering satisfy (A1) and (A2) it follows that all methods other than unilateral clustering satisfy (A1') and (P1) as well.

The Alternative Axiom of Value (A1") is satisfied by unilateral clustering only, which is also the only method listed above that satisfies the Alternative Property of Influence (P1') but does not satisfy the (regular) Property of Influence. We have also proved that (P1') is implied by (A1") and (A2) in the same manner that (P1) is implied by (A1) and (A2). Since the Agnostic Axiom of Value (A1"') encompasses (A1) and (A1") all of the methods listed above satisfy (A1"').

Scale Preservation (P2), Linear Scale Preservation (P2'), Representability (P3), and Excisiveness (P4) were introduced to winnow down the space of methods admissible with respect to the Axiom of Value (A1) and the Axiom of Transformation (A2). E.g., Scale Preservation (P2) is satisfied by reciprocal, nonreciprocal, semi-reciprocal, algorithmic intermediates, structure representable, cyclic and unilateral clustering but is not satisfied by the grafting and convex combination families. Linear Scale Preservation (P2') is satisfied by the same subset of methods plus the convex combination method whenever the methods being combined satisfy (P2') separately. This is not true for (regular) Scale Preservation. Grafting methods still violate the weaker notion of Linear Scale Preservation (P2').

Representability (P3), i.e. the ability to specify a given clustering method by its action on a set of networks, is satisfied by reciprocal, nonreciprocal, structure representable, cyclic, and unilateral methods but it is not satisfied by the rest of the methods that we studied. Moreover, we showed that methods satisfying (P3) can be decomposed into a symmetrizing operation followed by the application of single linkage clustering. Excisiveness (P4), a property encoding local consistency of the methods, is satisfied by the same subset of methods that satisfy (P3) and is violated by the rest. This is consistent with the result shown stating that, for admissible methods, Representability is equivalent to Excisiveness and Linear Scale Preservation.

To study stability we adopted the Gromov-Hausdorff distance, which was shown to be properly defined, therefore allowing the quantification of differences between networks. Since output dendrograms are equivalent to finite ultrametric spaces which in turn are particular cases of networks, this distance can be used to compare both the given networks and their corresponding output ultrametrics. The notion of stability of a given method that we adopted is that the distance between two outputs produced by the given hierarchical clustering method is bounded by the distance between the original networks. This means that clustering methods are non-expansive maps in the space of networks, i.e. they do not increase the distance between the given networks. An intuitive interpretation of the stability property is that similar networks yield similar dendrograms. The Stability Property (P5) is satisfied by reciprocal, nonreciprocal, semi-reciprocal, algorithmic intermediates, cyclic, structure representable, and unilateral clustering methods. The grafting and convex combination families are not stable in this sense.

7.2 Taxonomy of methods

A classification from the perspective of methods follows from reading the columns in Table 7.1. This taxonomy is more interesting than the one in Section 7.1 because the reciprocal, nonreciprocal, structure representable, and unilateral methods not only satisfy desirable properties but have also been proved to be either extreme or unique among those methods that are admissible with respect to some subset of properties.

Indeed, reciprocal $\mathcal{H}^{\mathrm{R}}$ and nonreciprocal $\mathcal{H}^{\mathrm{NR}}$ clustering were shown to be extremes of the range of methods that satisfy (A1)-(A2) in that the clustering outputs of these two methods provide uniform upper and lower bounds, respectively, for the output of every other method under this axiomatic framework. These two methods also satisfy all the other desirable properties that are compatible with (A1). I.e., they satisfy the extended and Agnostic Axioms of value, the Property of Influence, and, implied by it, the Alternative Property of Influence. They are also respectively represented by the two node unit cycle and the countable family of all unit cycles, scale preserving, excisive, and stable in terms of the generalized Gromov-Hausdorff distance.

Unilateral clustering $\mathcal{H}^{\mathrm{U}}$ is the unique method that abides by the alternative set of axioms (A1")-(A2). In that sense it plays the dual role of reciprocal and nonreciprocal clustering when we replace the Axiom of Value (A1) with the Alternative Axiom of Value (A1"). Unilateral clustering also satisfies all the desirable properties that are compatible with (A1"). It satisfies the Agnostic Axiom of Value, the Alternative Property of Influence, Scale Preservation, Excisiveness, and Stability. Unilateral clustering is also represented by a network of two nodes with unit dissimilarity in one direction and undefined dissimilarity in the opposite direction.

Unilateral $\mathcal{H}^{\mathrm{U}}$ and reciprocal $\mathcal{H}^{\mathrm{R}}$ clustering were shown to be extreme among methods that are admissible with respect to (A1"')-(A2). Unilateral clustering yields uniformly minimal ultrametric distances, while reciprocal clustering yields uniformly maximal ultrametric distances.

Structure representable clustering methods constitute the only family of methods that satisfy the regular set of axioms (A1)-(A2) plus the properties of Scale Preservation and Representability. Alternatively, structure representable methods were shown to be the only methods satisfying axioms (A1)-(A2), Scale Preservation, and Excisiveness. Since $\mathcal{H}^{\mathrm{R}}$ and $\mathcal{H}^{\mathrm{NR}}$ are scale preserving and excisive, the fact that they belong to the family of structure representable methods is consistent with these uniqueness results. Furthermore, structure representable methods were shown to be stable as defined in terms of the Gromov-Hausdorff distance.

The cyclic clustering methods $\mathcal{H}^{\circlearrowleft_t}$ limit the extent to which influence can propagate by allowing propagation through loops of maximum length controlled by the parameter t. The flexibility given by this parameter makes the cyclic clustering methods useful in practice when clustering networks where the bidirectional assumption of reciprocal clustering is too restrictive and the cycles of arbitrary length in nonreciprocal clustering are too permissive. Being a subset of the structure representable family, the cyclic methods satisfy every property compatible with axiom (A1).

We also considered families of methods that are admissible with respect to (A1)-(A2), but that fail to satisfy Scale Preservation and/or Excisiveness. These methods are generically regarded as intermediate methods since, given that they satisfy (A1)-(A2), they yield ultrametrics that lie between the outputs of reciprocal and nonreciprocal clustering. The first such family considered is that of grafting methods $\mathcal{H}^{\mathrm{R/NR}}(\beta)$ and $\mathcal{H}^{\mathrm{R/R_{max}}}(\beta)$. They satisfy the axioms and properties that can be derived from (A1)-(A2), i.e. the Extended Axiom of Value, the Agnostic Axiom of Value, the Property of Influence and the Alternative Property of Influence. Their dependance on a cutting parameter β is the reason why they both fail to fulfill the Scale Preservation, Linear Scale Preservation and Representability properties. For the case of $\mathcal{H}^{\mathrm{R/NR}}(\beta)$, although not shown here, it is possible to prove that it does not satisfy Excisiveness or Stability, hence, impairing practicality of these methods.

Convex combination clustering methods $\mathcal{H}^{12}_{\theta}$ constitute another family of intermediate methods considered. Their admissibility is based on the result that the convex combination of two admissible methods is itself an admissible clustering method. However, although

Table 7.2: Hierarchical clustering algorithms

Method	Observations	Notation	Formula
Reciprocal		u_X^{R}	$\Big(\max\big(A_X, A_X^T\big)\Big)^{n-1}$
Nonreciprocal		u_X^{NR}	$\max\Big(A_X^{n-1}, (A_X^T)^{n-1}\Big)$
Grafting	Reciprocal/nonreciprocal	$u_X^{\mathrm{R/NR}}(\beta)$	$u_X^{\mathrm{NR}} \circ \mathbb{I}\left\{u_X^{\mathrm{R}} \leq \beta\right\} + u_X^{\mathrm{R}} \circ \mathbb{I}\left\{u_X^{\mathrm{R}} > \beta\right\}$
Convex Combinations	Given $\mathcal{H}^1$ and $\mathcal{H}^2$	$u_X^{12}(\theta)$	$\Big(\theta\, u_X^1 + (1-\theta)\, u_X^2\Big)^{n-1}$
Semi-reciprocal (t)	Secondary paths of length t	$u_X^{\mathrm{SR}(t)}$	$\Big(\max\Big(A_X^{t-1}, (A_X^T)^{t-1}\Big)\Big)^{n-1}$
Algorithmic intermediate	Given parameters t and t'	$u_X^{t,t'}$	$\Big(\max\Big(A_X^t, (A_X^T)^{t'}\Big)\Big)^{n-1}$
Structure representable	Depends on representer	u_X^{Ω}	
Cyclic	Loops of length at most t	$u_X^{\circlearrowleft t}$	$(\max\big(A_X, (A_X^T)^{t-1}\big))^{n-1}$
Unilateral		u_X^{U}	$\Big(\min\big(A_X, A_X^T\big)\Big)^{n-1}$
Directed single linkage	Quasi-clustering	$\tilde{u}_X^*$	A_X^{n-1}
Single linkage	Symmetric networks	u_X^{SL}	A_X^{n-1}

not proved here, it is not hard to see that the convex combination operation does not preserve relevant properties such as Scale Preservation, Representability, Excisiveness and Stability. This means that if two methods satisfy one of the mentioned properties, their convex combination is not guaranteed to satisfy it. Linear Scale Preservation is the only property which cannot be derived from axioms (A1)-(A2) and is preserved when performing the convex combination of two methods.

Semi-reciprocal clustering methods $\mathcal{H}^{\mathrm{SR}(t)}$ allow the formation of cyclic influences in a more restrictive way than nonreciprocal clustering but more permissive than reciprocal clustering, controlled by the integer parameter t. Their interpretation is similar to cyclic clustering methods $\mathcal{H}^{\circlearrowleft t}$ but their practicality might be limited due to violation of some relevant properties. Although, semi-reciprocal clustering methods were shown to be scale preserving and stable, they fail to be representable and excisive. Algorithmic intermediate clustering methods $\mathcal{H}^{t,t'}$ are a generalization of semi-reciprocal methods and share their same properties.

7.3 Algorithms and applications

Algorithms for the application of the methods described were developed using a min-max dioid algebra on the nonnegative reals. In this algebra, the regular sum is replaced by the minimization operator and the regular product by maximization. The k-th power of the dissimilarity matrix was shown to contain in position i, j the minimum path cost corresponding to going from node i to node j in at most k hops. Since path costs played a major role in the definition of clustering methods, dioid matrix powers were presented as a natural framework for algorithmic development.

The reciprocal ultrametric was computed by first symmetrizing directed dissimilarities to their maximum and then computing increasing powers of the symmetrized dissimilarity matrix until stabilization. For the nonreciprocal case, the opposite was shown to be true, i.e., we first take successive powers of the asymmetric dissimilarity matrix until stabilization and then symmetrize the result via a maximum operation. In this way, the extreme nature of these methods was also illustrated in the algorithmic domain. In a similar fashion, algorithms for the remaining clustering methods presented were developed in terms of finite matrix powers, thus exhibiting computational tractability of our clustering constructions. A summary of all the algorithms presented in this first part of the thesis is available in Table 7.2.

Clustering algorithms were applied to two real-world networks. We gained insight about migrational preferences of individuals within United States by clustering a network of internal migration. In addition, we applied the developed clustering theory to a network containing information about how sectors of the U.S. economy interact to generate gross domestic product. In this way, we learned about economic sectors exhibiting pronounced interdependence and reasoned their relation with the rest of the economy.

The migration network example illustrates the different clustering outputs obtained when we consider the Axiom of Value (A1) or the Alternative Axiom of Value (A1") as conditions for admissibility. Unilateral clustering, the only method compatible with (A1"), forms clusters around influential states like California and Texas by merging each of these states with other smaller ones around them. On the other hand, methods compatible with (A1) like reciprocal clustering, tend to first merge states with balanced bidirectional influence such as two different populous states or states sharing urban areas. In this way, reciprocal clustering sees California first merging with Texas for being two very influential states and Washington merging with Oregon for sharing the urban area of Portland. Moreover, the similarity between the reciprocal and nonreciprocal outcomes indicates that no other clustering method satisfying axiom (A1) would reveal new information, thus, intermediate clustering methods were not applied.

For the network of economic sectors, reciprocal and nonreciprocal clustering output essentially different dendrograms, indicating the ubiquity of influential cycles between sectors. Reciprocal clustering first merges sectors of bidirectional influence such as professional services with administrative services and the farming sector with the food and beverage sector. Nonreciprocal clustering, on the other hand, captures cycles of influence such as the one between oil and gas extraction, petroleum and coal products, and the construction sector. However, nonreciprocal clustering propagates influence through arbitrarily large cycles, which might be undesirable in practice. The observed difference between the reciprocal and the nonreciprocal dendrograms motivated the application of a clustering method with intermediate behavior such as the cyclic clustering method of length 3. Its cyclic propagation of influence is closer to the real behavior of sectors within the economy and, thus, we obtained a more reasonable clustering output.

7.4 Symmetric networks and asymmetric quasi-ultrametrics

In hierarchical clustering of asymmetric networks we output a symmetric ultrametric to summarize information about the original asymmetric structure. As a particular case, we considered the construction of symmetric ultrametrics when the original network is symmetric. As a generalization, we studied the problem of defining and constructing asymmetric ultrametrics associated with asymmetric networks.

By restricting our theory to the particular case of symmetric networks, we strengthened an existing uniqueness result. Previous results showed that single linkage is the only admissible clustering method for finite metric spaces under a framework determined by three axioms. In this work, we showed that single linkage is the only admissible method for symmetric networks – a superset of metric spaces – in a framework determined only by two axioms, i.e. the Symmetric Axiom of Value (B1) and the Axiom of Transformation (A2), out of the three axioms considered in previous literature.

Hierarchical clustering methods output dendrograms, which are symmetric data structures. When clustering asymmetric networks, requiring the output to be symmetric might be undesirable. In this context we defined quasi-dendrograms, a generalization of dendrograms that admits asymmetric relations, and developed a theory for quasi-clustering methods, i.e. methods that output quasi-dendrograms when applied to asymmetric networks. In this context, we revised the notion of admissibility by introducing the Directed Axiom of Value ($\tilde{A}1$) and the Directed Axiom of Transformation ($\tilde{A}2$). Under this framework, we showed that directed single linkage – an asymmetric version of the single linkage clustering method – is the only admissible method. Furthermore, we proved an equivalence between quasi-dendrograms and quasi-ultrametrics that generalizes the known equivalence between dendrograms and ultrametrics. Algorithmically, the quasi-ultrametric produced by directed single linkage can be computed by applying iterated min-max matrix power operations to the dissimilarity matrix of the network until stabilization.

Directed single linkage can be used to understand relationships that cannot be understood when performing (regular) hierarchical clustering. In particular, the directed influences between clusters of a given resolution define a partial order between clusters which permits making observations about the relative importances of different clusters. This was corroborated through the application of directed single linkage to the United Stated internal migration network. Regular hierarchical clustering uncovers the grouping of California with other West Coast states and the grouping of Massachusetts with other New England States. Directed single linkage shows that California is the dominant state in the West Coast whereas Massachusetts appears as the dominant state in New England. When applied to the network of interactions between sectors of the United States economy, directed single linkage revealed the prominent influence of manufacturing, finance and professional services over the rest of the economy.

Part II

Network Projections onto Metric Spaces

Chapter 8

Canonical projections for symmetric networks

In the first part of this thesis we presented an in-depth analysis of hierarchical clustering for asymmetric networks. Equivalently, we studied how to map networks onto ultrametric (and quasi-ultrametric) spaces. Since ultrametrics $\mathcal{U}$ constitute a subset of all networks $\tilde{\mathcal{N}}$, one may reinterpret hierarchical clustering as a *projection* operation from all the networks onto the subset of more structured ones.

In this second part of the thesis we extend and elaborate this concept of projection onto structured spaces. We consider image spaces more general than $\mathcal{U}$ but that still preserve some notion of structure encoded by a triadic relation akin to the strong triangle inequality in ultrametrics. We denominate these structures as q-metric spaces and denote by $\mathcal{M}_q$ the set of all such spaces. In the current chapter we study projections $\mathcal{P}_q$ from symmetric networks $\mathcal{N}$ onto q-metric spaces $\mathcal{M}_q$. In Chapter 9, we extend our domain space to consider possibly asymmetric networks $\tilde{\mathcal{N}}$ but constrain our analysis to projections onto the space of (regular) metric spaces $\mathcal{M}_1$. Furthermore, in Chapter 10, we generalize our image set to study projections from $\tilde{\mathcal{N}}$ onto both quasi-metric spaces $\tilde{\mathcal{M}}$ (Section 10.1) and q-metric spaces $\mathcal{M}_q$ (Section 10.2). Finally, in Chapter 11 we leverage the relation between hierarchical clustering and dioid algebras presented in Section 3.5 to study metric projections at a higher level of algebraic abstraction. This allows us to handle, e.g., networks in which the edge weights are not given by numbers but rather by elements of a pre-specified power set.

Before delving into the development of an axiomatic framework for the study of metric projections, in Section 8.1 we introduce a series of mathematical preliminaries that complement those presented in Section 2.1.

8.1 Metric projections and q-metric spaces

In this second part of the thesis we consider weighted and directed graphs or networks. Throughout Part I, we assumed that dissimilarities were defined for all pairs of nodes (with the exception of representer networks). Here, we relax this requirement and formally define a network $G = (V, E, W)$ as a triplet formed by a finite set of n nodes or vertices V, a set of edges $E \subset V \times V$ where $(x, y) \in E$ represents an edge from $x \in V$ to $y \in V$, and a map $W : E \to \mathbb{R}_{++}$ from the set of edges to the strictly positive reals, representing weights $W(x, y) > 0$ associated with each edge (x, y). The weights represent dissimilarities, i.e. the smaller the weight the more similar the nodes are. We assume that the graphs of interest do not contain self-loops, i.e., $(x, x) \notin E$ for all $x \in V$. Recall that we denote by $\tilde{\mathcal{N}}$ the set of all (possibly directed) networks whereas we reserve the notation $\mathcal{N}$ to denote the set of all undirected networks, i.e, networks where $(x, y) \in E$ implies $(y, x) \in E$ and $W(x, y) = W(y, x)$. This latter set of undirected networks will be of specific interest in the current chapter.

Metrics and their generalizations play in Part II a central role akin to that played by ultrametrics in the preceding part.

Definition 4 *Given a set X, a metric $d : X \times X \to \mathbb{R}_+$ is a function from pairs of elements to the non-negative reals satisfying the following properties for every $x, y, z \in X$:*

(i) Identity: $d(x, y) = 0$ if and only if $x = y$.

(ii) Symmetry: $d(x, y) = d(y, x)$.

(iii) Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

The ordered pair $M = (X, d)$ is said to be a *metric space* [9] and the set of all finite metric spaces is denoted by $\mathcal{M}$. By contrast, whenever a function $\tilde{d} : X \times X \to \mathbb{R}_+$ satisfies the identity and triangle inequality properties but not necessarily the symmetry property, we say that $\tilde{d}$ is a quasi-metric and that $\tilde{M} = (X, \tilde{d})$ is a quasi-metric space, and we denote the set of all such spaces by $\tilde{\mathcal{M}}$.

We also consider a more general class of spaces termed *q-metric spaces* that are parametrized by $q \in [1, \infty]$. For finite $q < \infty$, a q-metric space is a pair $M = (X, d)$ where the function d satisfies the symmetry and identity properties but a q-triangle inequality in lieu of the regular triangle inequality. This *q-triangle inequality* is such that for all $x, y, z \in X$, it holds,

$$d(x, y)^q \leq d(x, z)^q + d(z, y)^q. \tag{8.1}$$

When $q = \infty$ we define an ∞-metric space as a pair $M = (X, d)$ that satisfies the symmetry and identity properties as well as the ∞-*triangle inequality*, for all $x, y, z \in X$,

$$d(x, y) \leq \max(d(x, z), d(z, y)). \tag{8.2}$$

Notice that the regular definition of metric space is recovered when $q = 1$. When $q = \infty$, the ∞-triangle inequality is equivalent to the strong triangle inequality that characterizes ultrametric spaces [cf. (2.12) in Definition 1]. Another instance of interest is that of 2-metric spaces, in which case all of the triangles in the space are acute angled triangles; see Section 8.3.3.

Throughout the thesis, we interpret q-metrics as particular cases of networks which we can do if we associate every q-metric space (V, d) to the complete network (V, E, W_d) where the edge set contains every possible edge except self-loops, i.e. $(x, y) \in E$ for all $x \neq y \in V$. Furthermore, the edge weights W_d are given by the q-metric d, i.e., $W_d(x, y) = d(x, y)$. We represent q-metric spaces as (V, d) or its network equivalent (V, E, W_d) interchangeably. The set of all q-metric spaces is denoted by $\mathcal{M}_q$ where $\mathcal{M} \equiv \mathcal{M}_1$. Notice that for all $q > q' > 1$ we must have

$$\mathcal{M}_q \subset \mathcal{M}_{q'} \subset \mathcal{M}_1 \equiv \mathcal{M} \subset \mathcal{N}. \tag{8.3}$$

A closely related definition is that of a norm [9].

Definition 5 *given vector space Y, a norm $\|\cdot\|$ is a function $\|\cdot\| : Y \to \mathbb{R}_+$ from Y to the non-negative reals such that, for all vectors $v, w \in Y$ and scalar β, it satisfies:*

(i) Positiveness: $\|v\| \geq 0$ with equality if and only if $v = \vec{0}$.

(ii) Positive homogeneity: $\|\beta\, w\| = |\beta|\, \|w\|$.

(iii) Subadditivity: $\|v + w\| \leq \|v\| + \|w\|$.

A commonly used family of norms for vectors in $\mathbb{R}^l$ are the p-norms $\|\cdot\|_p$ for real $p \in [1, \infty]$, where the norm of a vector $v = [v_1, \ldots, v_l]$ is given by

$$\|v\|_p := \begin{cases} \left(\sum_{i=1}^{l} |v_i|^p\right)^{1/p} & \text{for } p < \infty, \\ \max_i |v_i| & \text{for } p = \infty. \end{cases} \tag{8.4}$$

Recall from Section 2.1 that a path $P_{xx'}$ is an ordered sequence of nodes, $P_{xx'} = [x = x_0, x_1, \ldots, x_{l-1}, x_l = x']$, which starts at x and finishes at x' and, when dealing with non-complete networks, we require that $e_i = (x_i, x_{i+1}) \in E$ for $i = 0, \ldots, l-1$. For a given norm $\|\cdot\|$, we define the *length* of a given path $P_{xx'} = [x = x_0, \ldots, x_l = x']$ as

$$\|P_{xx'}\| := \|[W(x_0, x_1), \ldots, W(x_{l-1}, x_l)]\|, \tag{8.5}$$

i.e., the norm of the vector that consists of the weights associated to the links in the path. Note that we purposely abuse notation by applying the norm $\|\cdot\|$ directly on the path $P_{xx'}$. From (8.4) and (8.5) it follows that the path cost, ubiquitous in Part I [cf. (2.6)], is here denoted by $\|P_{xx'}\|_\infty$.

Of central importance in this second part is the concept of shortest path from x to x' which we denote by $\tilde{d}(x, x')$, i.e.,

$$\tilde{d}(x, x') := \min_{P_{xx'}} \|P_{xx'}\|_1. \tag{8.6}$$

It should be noted that $\tilde{d}$ is a valid quasi-metric since the shortest path between two nodes x and y is at most as long as the shortest path going from x to an intermediate node z and then from z to y.

We study the design of metric projections $\mathcal{P}$ with the objective of representing networks as metric spaces, or more generally, projections $\mathcal{P}_q$ onto q-metric spaces. Formally, for all node sets V we define a metric projection $\mathcal{P} : \tilde{\mathcal{N}} \to \mathcal{M}$ as a map that projects every (possibly directed) network onto a metric space while preserving V. We say that two metric projections $\mathcal{P}$ and $\mathcal{P}'$ are equivalent, and we write $\mathcal{P} \equiv \mathcal{P}'$, if $\mathcal{P}(G) = \mathcal{P}'(G)$, for all $G \in \tilde{\mathcal{N}}$. Finally, we denote by $\tilde{\mathcal{P}}$ the quasi-metric projection that maps any network G onto the space $(V, \tilde{d})$ with the quasi-metric given by the shortest paths in G [cf. (8.6)].

8.2 Axioms of Projection and Injective Transformation

The broad definition of $\mathcal{P}_q$ as a node-preserving and q-metric inducing map presented in Section 8.1 admits maps of undesirable behavior. E.g., we may define the map $\mathcal{P}_q^1$ such that for any $G = (V, E, W)$ it outputs the space $(V, d) = \mathcal{P}_q^1(G)$ where d is defined as $d(x, x') = 1$ and $d(x, x) = 0$ for all $x \neq x' \in V$. It is immediate to see that the output space is q-metric for all q, since it is ∞-metric or ultrametric. However, the defined q-metric d completely ignores the edge structure in E and the edge weights in W, which is undesirable. In order to discard unreasonable projections like $\mathcal{P}_q^1$, we follow and axiomatic approach as done for hierarchical clustering in the first part of this thesis and encode in the form of the Axioms of Projection and Injective Transformation, desirable properties that a map $\mathcal{P}_q$ should satisfy.

Recalling that the set of q-metric spaces $\mathcal{M}_q$ is a subset of the set of all networks $\mathcal{N}$ [cf. (8.3)], we define the following axiom:

> *(AA1) Axiom of Projection.* Every q-metric space $M \in \mathcal{M}_q$ is a fixed point of the projection map $\mathcal{P}_q$, i.e. $\mathcal{P}_q(M) = M$.

Given that our goal is the design of maps that transform general networks into more structured q-metric spaces, if we already have a q-metric space there is no justification to change

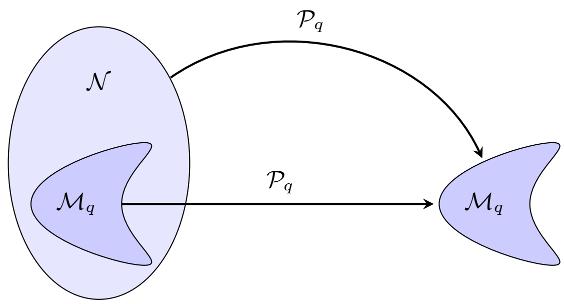


Figure 8.1: Axiom of Projection for q-metric spaces. The q-metric space $\mathcal{M}_q$ is an invariant set of the projection map $\mathcal{P}_q$.

it; see Fig. 8.1. This concept is captured in axiom (AA1) where we define $\mathcal{M}_q$ as the fixed set of $\mathcal{P}_q$. Equivalently, we say that the map $\mathcal{P}_q$ restricted to $\mathcal{M}_q$ is the identity map. It is immediate that axiom (AA1) implies idempotency of $\mathcal{P}_q$, which is a requirement of projection maps, hence, its denomination as Axiom of Projection.

The second restriction on the set of allowable maps $\mathcal{P}_q$ formalizes our expectations for the behavior of $\mathcal{P}_q$ when confronted with a transformation of the underlying node set V, edge set E, and edge weights W, and it is a minor variation on the Axiom of Transformation introduced in Section 2.2. Since we now consider non-complete networks, we redefine our notion of dissimilarity reducing map as follows: the *injective* map $\phi : V \to V'$ is called a *dissimilarity reducing* map if it holds that $(\phi(x), \phi(x')) \in E'$ and $W(x, x') \geq W'(\phi(x), \phi(x'))$ for all $(x, x') \in E$; see Fig. 8.2. A dissimilarity reducing map matches every edge in G with an edge in G' of less or equal weight. Notice that, since ϕ is defined to be injective, the codomain set V' must have at least as many nodes as the domain set V.

Intuitively, if we look at any path $P_{xx'}$ between nodes x and x' in G, the existence of the dissimilarity reducing map ϕ ensures that there is an associated path $P_{\phi(x)\phi(x')}$ in G' between the nodes $\phi(x)$ and $\phi(x')$ such that the weight of every link in this second path is not greater than the corresponding links in the first one. On top of this, there might exist additional paths between $\phi(x)$ and $\phi(x')$ in G' that are not the image under ϕ of any path between x and x' in G. Thus, it is expected for nodes $\phi(x)$ and $\phi(x')$ to be closer to each other in the output q-metric spaces. E.g., in Fig. 8.2 there is only one path between x_1 and x_2 with weights 2 and 1 whereas between $\phi(x_1) = y_1$ and $\phi(x_2) = y_2$ there are two paths: the transformed path under ϕ with smaller weights both of 1 plus a direct path of weight 2. Hence, we expect nodes y_1 and y_2 to be closer to each other than x_1 and x_2 in the output q-metric spaces.

The Axiom of Injective Transformation that we introduce next is a formal statement of the intuition described above:

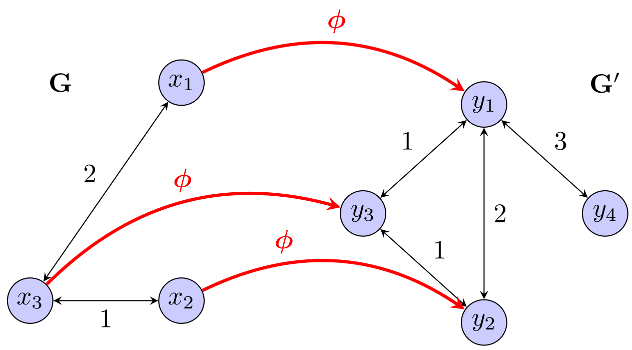


Figure 8.2: Dissimilarity reducing map. The injective map ϕ takes every edge in network G to an edge in network G' of less or equal weight.

(AA2) Axiom of Injective Transformation. Consider any two networks $G = (V, E, W)$ and $G' = (V', E', W')$ and any (injective) dissimilarity reducing map $\phi : V \to V'$. Then, for all $x, x' \in V$, the output q-metric spaces $(V, d) = \mathcal{P}_q(G)$ and $(V', d') = \mathcal{P}_q(G')$ satisfy

$$d(x, x') \geq d'(\phi(x), \phi(x')). \tag{8.7}$$

As done in the first part of the thesis, we say that a projection $\mathcal{P}_q$ is *admissible* if it satisfies axioms (AA1) and (AA2). In the next section, we study the set of admissible projections $\mathcal{P}_q$ for every q.

Remark 18 (Related axiomatic constructions) In [10, 11] the authors propose three axioms to study hierarchical clustering of metric spaces. Given the relation between hierarchical clustering and ultrametrics [10], such a problem can be reformulated in terms of our current notation as the definition of maps from $\mathcal{M}_1$ to $\mathcal{M}_\infty$. In this chapter, the domain set is $\mathcal{N}$ (which is a superset of $\mathcal{M}_1$) and we use a unified framework to study projections onto all $\mathcal{M}_q$, of which a particular case is $q = \infty$. The Axioms of Projection and Injective Transformation are related to two of the axioms in [10, 11]. In Section 8.3.2 we show that, using our more general framework, we recover the unicity result in [10, 11] with less stringent axioms. Regarding the study of hierarchical clustering in Part I of this thesis, we may interpret the clustering methods as projections from $\tilde{\mathcal{N}}$ to $\mathcal{M}_\infty$. In this chapter, we consider a more restricted class of symmetric networks $\mathcal{N}$ but, in contrast to Part I, the networks need not be complete. Extensions of the axiomatic framework here presented for the projection of *directed* networks $\mathcal{N}$ onto general q-metric spaces $\mathcal{M}_q$ are presented in Section 10.2.

8.3 Uniqueness of metric projections

After posing the Axioms of Projection and Injective Transformation, we first seek to answer if *any* map $\mathcal{P}_q$ satisfies them. In this direction, given a graph $G = (V, E, W)$, for each q we define the *canonical* q-metric projection $\mathcal{P}_q^*$ with output $(V, d_q^*) = \mathcal{P}_q^*(G)$ where the q-metric d_q^* between points x and x' is given by

$$d_q^*(x, x') := \min_{P_{xx'}} \|P_{xx'}\|_q. \tag{8.8}$$

In (8.8), to find the distance between two points, we look for the path that links these nodes while minimizing its q-norm, as defined in (8.4). We restrict our attention to connected networks G to ensure that $d_q^*(x, x')$ is well-defined for every pair of nodes $x, x' \in V$.

For the method $\mathcal{P}_q^*$ to be a properly defined q-metric projection for all q, we need to establish that (V, d_q^*) is a valid q-metric space. Furthermore, it can also be shown that $\mathcal{P}_q^*$ satisfies axioms (AA1)-(AA2). We prove both assertions in the following proposition.

Proposition 21 *The canonical q-metric projection map $\mathcal{P}_q^*$ is valid and admissible. I.e., d_q^* defined by* (8.8) *is a q-metric for all undirected networks G and $\mathcal{P}_q^*$ satisfies the Axioms of Projection (AA1) and Injective Transformation (AA2).*

Proof: We first prove that d_q^* is indeed a q-metric on the node set V. That $d_q^*(x, x') = d_q^*(x', x)$ follows from combining the facts that the original graph G is undirected and that the q-norm of a path $\|P_{xx'}\|_q$ does not depend on the order in which its links are traversed. Moreover, that $d_q^*(x, x') = 0$ if and only if $x = x'$ is a consequence of the positiveness property of the q-norms (cf. Definition 5). To verify that the q-triangle inequality holds, let $P_{xx'}$ and $P_{x'x''}$ be paths that achieve the minimum in (8.8) for $d_q^*(x, x')$ and $d_q^*(x', x'')$, respectively. Then, for finite q it follows that

$$\begin{aligned} d_q^*(x, x'')^q &= \min_{P_{xx''}} \|P_{xx''}\|_q^q \leq \|P_{xx'} \uplus P_{x'x''}\|_q^q \\ &= \|P_{xx'}\|_q^q + \|P_{x'x''}\|_q^q = d_q^*(x, x')^q + d_q^*(x', x'')^q, \end{aligned} \tag{8.9}$$

where the inequality follows from the fact that the concatenated path $P_{xx'} \uplus P_{x'x''}$ is *a particular* path between x and x'' while the definition of $d_q^*(x, x'')$ minimizes the norm across all such paths.

To see that the Axiom of Projection (AA1) is satisfied, pick an arbitrary q-metric space $M = (V, d) \in \mathcal{M}_q$ and denote by $(V, d_q^*) = \mathcal{P}_q^*(M)$ the output of applying the canonical q-metric projection to M. For an arbitrary pair of nodes $x, x' \in V$, we have that

$$d_q^*(x, x') = \min_{P_{xx'}} \|P_{xx'}\|_q \leq \|[x, x']\|_q = d(x, x'), \tag{8.10}$$

for all q, where the inequality comes from specializing the path $P_{xx'}$ to the path $[x, x']$ with just one link from x to x'. Moreover, if we denote by $P^*_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ the path achieving the minimum in (8.10), then we may leverage the fact that d satisfies the q-triangle inequality to write

$$d(x, x') \leq \left(\sum_{i=0}^{l-1} d(x_i, x_{i+1})^q \right)^{1/q} = \|P^*_{xx'}\|_q = d^*_q(x, x'). \tag{8.11}$$

Upon substituting (8.11) into (8.10), we obtain that all the inequalities are, in fact, equalities, implying that $d^*_q(x, x') = d(x, x')$. Since nodes x, x' were chosen arbitrarily, it must be that $d \equiv d^*_q$ which implies that $\mathcal{P}^*_q(M) = M$, as wanted.

To show fulfillment of axiom (AA2), consider two networks $G = (V, E, W)$ and $G' = (V', E', W')$ and a dissimilarity reducing map $\phi : V \to V'$. Let $(V, d_q) = \mathcal{P}^*_q(G)$ and $(V', d'_q) = \mathcal{P}^*_q(G')$ be the outputs of applying the canonical projection to networks G and G', respectively. For an arbitrary pair of nodes $x, x' \in V$, denote by $P^*_{xx'} = [x = x_0, \ldots, x_l = x']$ a path that achieves the minimum in (8.8) so as to write

$$d_q(x, x') = \|P^*_{xx'}\|_q. \tag{8.12}$$

Consider the transformed path $P^*_{\phi(x)\phi(x')} = [\phi(x) = \phi(x_0), \ldots, \phi(x_l) = \phi(x')]$ in the set V'. Since the transformation ϕ does not increase dissimilarities, we have that for all links in this path $W'(\phi(x_i), \phi(x_{i+1})) \leq W(x_i, x_{i+1})$. Combining this observation with (8.12) we obtain,

$$\|P^*_{\phi(x)\phi(x')}\|_q \leq d_q(x, x'). \tag{8.13}$$

Further note that $P_{\phi(x)\phi(x')}$ is a particular path joining $\phi(x)$ and $\phi(x')$ whereas the metric d'_q is given by the minimum across all such paths. Therefore,

$$d'_q(\phi(x), \phi(x')) \leq \|P^*_{\phi(x)\phi(x')}\|_q. \tag{8.14}$$

Upon replacing (8.13) into (8.14), it follows that $d'_q(\phi(x), \phi(x')) \leq d_q(x, x')$, as required by the Axiom of Injective Transformation. The same proof can be replicated to show the validity of the result for the case $q = \infty$. ■

Given that we have shown that the canonical q-metric projection $\mathcal{P}^*_q$ satisfies axioms (AA1)-(AA2), two questions arise: i) Are there other projections satisfying (AA1)-(AA2)?; and ii) Is the projection $\mathcal{P}^*_q$ special in any sense? Both questions are answered by the following uniqueness theorem.

Theorem 17 *Let $\mathcal{P}_q : \mathcal{N} \to \mathcal{M}_q$ be a q-metric projection, and $\mathcal{P}^*_q$ be the canonical pro-*

jection with output q-metric as defined in (8.8). *If $\mathcal{P}_q$ satisfies the Axioms of Projection (AA1) and Injective Transformation (AA2) then $\mathcal{P}_q \equiv \mathcal{P}_q^*$, for all q.*

Proof: Given an arbitrary network $G = (V, E, W)$ and a fixed q, denote by $(V, d) = \mathcal{P}_q(G)$ and $(V, d^*) = \mathcal{P}_q^*(G)$ the output q-metric spaces when applying a generic admissible q-metric projection and the canonical q-metric projection, respectively. We will show that

$$d^*(x, x') \le d(x, x') \le d^*(x, x'), \tag{8.15}$$

for all $x, x' \in V$. Given that G was chosen arbitrarily, this implies that $\mathcal{P}_q \equiv \mathcal{P}_q^*$, as wanted.

We begin by showing that $d(x, x') \le d^*(x, x')$ for all $x, x' \in V$. Consider an arbitrary pair of points x and x' and let $P_{xx'} = [x = x_0, \ldots, x_l = x']$ be a path achieving the minimum in (8.8) so that, for finite q, we can write

$$d^*(x, x') = \left(\sum_{i=0}^{l-1} W(x_i, x_{i+1})^q \right)^{1/q}. \tag{8.16}$$

Focus now on a series of undirected two-node networks $G_i = (V_i, E_i, W_i)$ for $i = 0, \ldots, l-1$, such that $V_i = \{z, z'\}$ and $E_i = \{(z, z'), (z', z)\}$ for all i but with different weights given by $W_i(z, z') = W_i(z', z) = W(x_i, x_{i+1})$. Since every network G_i is already a q-metric – in fact, any undirected two-node network is a valid q-metric for all q – and the method $\mathcal{P}_q$ satisfies the Axiom of Projection (AA1), if we define $(\{z, z'\}, d_i) := \mathcal{P}_q(G_i)$ we must have that $d_i(z, z') = W(x_i, x_{i+1})$, i.e., every graph G_i is a fixed point of the map $\mathcal{P}_q$.

Consider transformations $\phi_i : \{z, z'\} \to V$ given by $\phi_i(z) = x_i$, $\phi_i(z') = x_{i+1}$ so as to map z and z' in G_i to subsequent points in the path $P_{xx'}$ used in (8.16). This implies that maps ϕ_i are dissimilarity reducing since they are injective and the only edge in G_i is mapped to an edge of the exact same weight in G for all i. Thus, it follows from the Axiom of Injective Transformation (AA2) that

$$d(\phi_i(z), \phi_i(z')) = d(x_i, x_{i+1}) \le d_i(z, z') = W(x_i, x_{i+1}). \tag{8.17}$$

To complete the proof we use the fact that since d is a q-metric and $P_{xx'} = [x = x_0, \ldots, x_l = x']$ is a path joining x and x', the q-triangle inequality dictates that

$$d(x, x') \le \left(\sum_{i=0}^{l-1} d(x_i, x_{i+1})^q \right)^{1/q} \le \left(\sum_{i=0}^{l-1} W(x_i, x_{i+1})^q \right)^{1/q}, \tag{8.18}$$

where we used (8.17) for the second inequality. The proof that $d(x, x') \le d^*(x, x')$ follows from substituting (8.16) into (8.18).

$$
\begin{tikzcd}
\mathbf{G} \arrow[r, "\mathcal{P}_q^*"] \arrow[d, "\mathcal{P}_q"'] \arrow[r, bend right, "\phi = \mathrm{Id}"'] & (\mathbf{V}, \mathbf{d}^*) \arrow[d, "\mathcal{P}_q"] \\
(\mathbf{V}, \mathbf{d}) & (\mathbf{V}, \mathbf{d}^*)
\end{tikzcd}
$$

Figure 8.3: Diagram of maps between spaces for the proof of Theorem 17. Since $\mathcal{P}_q$ satisfies axiom (AA2), the existence of the dissimilarity reducing map ϕ allows us to relate d and d^*.

We now show that $d(x, x') \geq d^*(x, x')$ for all $x, x' \in V$. To do this, first notice that for an arbitrary pair of points x and x', if the edge $(x, x') \in E$ then we have that

$$d^*(x, x') = \min_{P_{xx'}} \|P_{xx'}\|_q \leq W(x, x'), \tag{8.19}$$

where the inequality comes from considering the particular path $P_{xx'}$ with only two points $[x, x']$. Hence, the identity map $\phi = \mathrm{Id} : V \to V$ such that $\phi(x) = x$ for all $x \in V$ is a dissimilarity reducing map from G to (V, d^*), since it is injective and every existing edge in G is mapped to an edge with smaller or equal weight. Consequently, we can build the diagram of relations between spaces depicted in Fig. 8.3. The top (blue) and left (red) maps in the figure are given by the definitions at the beginning of this proof while the relation on the right (green) is a consequence of the axiom (AA1). Since the aforementioned identity map ϕ is dissimilarity reducing, we can use the fact that $\mathcal{P}_q$ satisfies axiom (AA2) to state that

$$d(x, x') \geq d^*(\phi(x), \phi(x')) = d^*(x, x'), \tag{8.20}$$

for all $x, x' \in V$, concluding the proof. An analogous proof can be sued to show the result for $q = \infty$. ■

According to Theorem 17, for each q there is one and only one projection map from the set $\mathcal{N}$ of symmetric networks onto $\mathcal{M}_q$ that satisfies the axioms of Projection and Injective Transformation, and this map is $\mathcal{P}_q^*$. Any other conceivable map into $\mathcal{M}_q$ must violate at least one of the axioms. E.g., the map $\mathcal{P}_q^1$ introduced at the beginning of Section 8.2 clearly violates axiom (AA1) since the only fixed point is the metric space with all its distances equal to 1. More interestingly, if we have $q > q'$ then $\mathcal{P}_q^*$ can be viewed as a map into $\mathcal{M}_{q'}$ since $\mathcal{M}_q \subset \mathcal{M}_{q'}$. However, map $\mathcal{P}_q^*$ also violates axiom (AA1) when viewed as a map into $\mathcal{M}_{q'}$, since any q'-metric space $M \in \mathcal{M}_{q'}$ that is *not* in $\mathcal{M}_q$ would not be a fixed point of $\mathcal{P}_q^*$.

8.3.1 Metric spaces

For an arbitrary network $G = (V, E, W)$, we may particularize our analysis to the case of (regular) metric spaces, i.e. $q = 1$. In this case, the canonical projection $\mathcal{P}_1^*$ outputs the metric space $(V, d_1^*) = \mathcal{P}_1^*(G)$ where d_1^* is given by [cf. (8.8)]

$$d_1^*(x, x') = \min_{P_{xx'}} \sum_{i=0}^{l-1} W(x_i, x_{i+1}), \tag{8.21}$$

for all $x, x' \in V$. Equivalently, $\mathcal{P}_1^*$ sets the distance between two nodes to the length of the shortest path between them.

By specializing Theorem 17 to the case $q = 1$, we obtain the following corollary.

Corollary 7 *The shortest path between every pair of nodes is the only admissible metric in networks, where admissibility is given by axioms (AA1)-(AA2).*

From the comparison of networks [67] to the determination of node importance [29, 71], shortest paths constitute a basic feature of network analysis and their application is ubiquitous. Moreover, efficient algorithms for the computation of every shortest path in a network exist [27, 90]. Corollary 7 can also be interpreted as a theoretical justification for utilizing shortest path distances as an intermediate step in nonlinear dimensionality reduction schemes such as Isomap [81]. Furthermore, in Section 8.4.1 we discuss the utility of projecting graphs onto metrics for the approximation of otherwise NP-hard graph theoretical problems that have guaranteed error bounds in polynomial time for metric data [26, 41].

8.3.2 Ultrametric spaces

As discussed in the first part of this thesis, hierarchical clustering of networks can be posed as a projection problem of networks onto ultrametric spaces. When setting $q = \infty$, the canonical projection induces the ultrametric d_∞^* given by [cf. (8.8)]

$$d_\infty^*(x, x') = \min_{P_{xx'}} \max_i W(x_i, x_{i+1}), \tag{8.22}$$

for all $x, x' \in V$. Expression (8.22) implies that the canonical projection $\mathcal{P}_\infty^*$ is equivalent to single linkage hierarchical clustering (2.15). Moreover, for the case $q = \infty$, the ensuing corollary follows from Theorem 17.

Corollary 8 *Single linkage hierarchical clustering is the only admissible hierarchical clustering method for networks, where admissibility is given by axioms (AA1)-(AA2).*

Single linkage has been previously shown to posses desirable theoretical features. In [10], single linkage was shown to be the only admissible method from metric spaces $\mathcal{M}_1$ to ultrametric spaces $\mathcal{M}_\infty$ satisfying three axioms, two of which can be derived from the axioms of Projection and Injective Transformation here presented. This implies that, when specializing our framework for $q = \infty$, there are two main advantages between the axiomatic framework derived here and that in [10]: i) single linkage is the only admissible method from $\mathcal{N}$ to $\mathcal{M}_\infty$, which subsumes unicity from $\mathcal{M}_1$ to $\mathcal{M}_\infty$; and ii) the third axiom considered in [10] is redundant, since the uniqueness result can be derived based solely on the first two.

In practice, single linkage has shown to have some undesirable features like the so-called chaining effect [75]. Nevertheless, our construction is of utility for the practitioner who prefers other hierarchical clustering methods. More specifically, by clearly stating our desired properties as axioms, it is made clear that at least one of the axioms (AA1)-(AA2) must be violated when picking a method different from single linkage.

Remark 19 Both Corollary 2 in Section 3.2.1 and Corollary 8 in the current section determine single linkage as the only admissible hierarchical clustering method for symmetric networks. Nonetheless, in Corollary 2 admissibility is defined in terms of the Symmetric Axiom of Value (B1) and the Axiom of Transformation (A2) whereas in Corollary 8 admissibility is studied with respect to (AA1) and (AA2). Notice that axiom (AA1) is *more stringent* than its counterpart (B1) since the latter requires only two-node networks to be fixed points of the clustering method whereas the former requires this property for metric spaces of all sizes. On the other hand, (AA2) is *less stringent* than (A2) since in (AA2) only injective maps are considered as dissimilarity reducing. It is interesting to notice that, when jointly considered, axioms (AA1)-(AA2) impose the same restrictions on admissibility as (B1)-(A2).

8.3.3 2-metric spaces

Metric and ultrametric spaces are the two most common examples of q-metric spaces. However, for intermediate values of q, i.e., between 1 and ∞, spaces with other special characteristics arise. In particular, when $q = 2$, we obtain a space in which every triangle is acute. More specifically, when the distances between points satisfy the 2-triangle inequality, it can be shown that every angle in every triangle is not greater than 90 degrees. Mimicking the reasoning in Sections 8.3.1 and 8.3.2, it follows that the canonical projection $\mathcal{P}_2^*$ with an associated induced distance between points x and x' given by

$$d_2^*(x, x') = \min_{P_{xx'}} \sqrt{\sum_{i=0}^{l-1} W(x_i, x_{i+1})^2}, \tag{8.23}$$

is the only admissible way of inducing a structured space in a network where every resulting triangle is acute.

8.4 Properties of the canonical projection

The axioms of Projection and Injective Transformation uniquely determine the family of canonical projections $\mathcal{P}_q^*$ for different q. Moreover, additional practical properties can be extracted from the aforementioned axioms. In this section we discuss the properties of optimality (Section 8.4.1), stability (Section 8.4.2), and nestedness (Section 8.4.3).

8.4.1 Optimality

A myriad of combinatorial optimization problems exist, where the goal is to find subsets of nodes or edges of a network that are optimal in some sense. Examples of the former are graph coloring – finding a partition of non-adjacent nodes with smallest cardinality [45] – and the maximum independent set problem – finding a set of non-adjacent nodes with maximal cardinality [80]. Examples of the latter include the traveling salesman problem – finding a path that visits each node exactly once with smallest length [52] – and the minimum bisection problem [8] – separating the network into two pieces with the same number of nodes so that the sum of the weights in the edges that connect the pieces is minimal. In this section we focus on this second category, where the problems are characterized by an objective function that depends on the weights of the edges of the network.

Define the function $f : \mathcal{N} \to \mathbb{R}$ that maps every network G to the minimum cost $f(G)$ of an optimization problem that depends on the structure of G. For the traveling salesman problem, $f(G)$ is the length of the optimal salesman's trajectory in G. For minimum bisection, $f(G)$ is the sum of the weights in the optimal bisection of G. Traveling salesman and minimum bisection are known to be NP-hard and also hard to approximate in general. This means that not only is the problem of finding optimal solutions computationally intractable, but the problem of finding approximate solutions in polynomial time is impossible as well – impossibility unless P = NP is known for the traveling salesman problem [69] and undetermined for the minimum bisection problem [25]. However, when the network under consideration is metric, both problems are approximable in polynomial time [20, 26]. These two examples are not isolated, there are many other combinatorial problems that are approximable when we restrict our attention to metric spaces [41].

We can leverage the fact that combinatorial problems are simpler to solve in metric spaces to efficiently obtain lower bounds for $f(G)$. More specifically, we restrict our attention to cost functions f that do not decrease with increasing edge weights, i.e. for networks G and G' with the same number of nodes, if the identity map is dissimilarity reducing from

$$
\begin{tikzcd}
\mathbf{G} \arrow[r, "\mathcal{P}_q"] \arrow[r, bend right, "\phi = \mathrm{Id}"'] \arrow[d, "\mathcal{P}_q^*"'] & (\mathbf{V}, \mathbf{d}) \arrow[d, "\mathcal{P}_q^*"] \\
(\mathbf{V}, \mathbf{d}^*) & (\mathbf{V}, \mathbf{d})
\end{tikzcd}
$$

Figure 8.4: Diagram of maps between spaces for the proof of Proposition 22. Since $\mathcal{P}_q^*$ satisfies axiom (AA2), the existence of the dissimilarity reducing map ϕ allows us to relate d^* and d.

G' to G then $f(G') \geq f(G)$. Thus, if we project a network G onto a metric space M where no dissimilarity is increased, we may compute the lower bound $f(M)$ efficiently. The optimal choice for this projection is the canonical map $\mathcal{P}_1^*$ as we show next in a more general proposition for q-metric spaces.

Proposition 22 *Given an arbitrary network $G = (V, E, W)$, let $\mathcal{P}_q : \mathcal{N} \to \mathcal{M}_q$ be a generic q-metric projection with output $(V, d) = \mathcal{P}_q(G)$. Then, for any cost function f non-decreasing in the edge weights of G, the canonical projection $\mathcal{P}_q^*$ satisfies*

$$
\begin{aligned}
&\mathcal{P}_q^* = \underset{\mathcal{P}_q}{\operatorname{argmin}}\, f(G) - f(\mathcal{P}_q(G)) \\
&\text{s.to} \qquad d(x, x') \leq W(x, x') \quad \text{for all}\, (x, x') \in E.
\end{aligned}
\tag{8.24}
$$

Proof: That $\mathcal{P}_q^*$ is feasible, meaning that its output q-metric $(V, d^*) = \mathcal{P}_q^*(G)$ satisfies the constraint in problem (8.24), can be shown using the same argument used to write expression (8.19). To see that $\mathcal{P}_q^*$ is optimal, denote by $\mathcal{P}_q$ a feasible q-metric projection with output $(V, d) = \mathcal{P}_q(G)$. The diagram in Fig. 8.4 summarizes the relations between G, (V, d), and (V, d^*). The top (blue) and left (red) maps represent the definitions of the metric projections. The right (green) map is justified by the fact that (V, d) is, by definition, a q-metric and that $\mathcal{P}_q^*$ satisfies the Axiom of Projection (cf. Proposition 21). Moreover, notice that d satisfying the constraint in (8.24) guarantees that the identity map $\phi : V \to V$ from G to (V, d) is dissimilarity reducing. Consequently, we combine the fact that $\mathcal{P}_q^*$ fulfills the Axiom of Injective Transformation (cf. Proposition 21) with the relations between spaces in Fig. 8.4 to write

$$
d^*(x, x') \geq d(\phi(x), \phi(x')) = d(x, x'), \tag{8.25}
$$

for all $x, x' \in V$. Combining (8.25) with the constraint in problem (8.24), we can write that

$$
d(x, x') \leq d^*(x, x') \leq W(x, x'), \tag{8.26}
$$

for all $(x, x') \in E$, showing that $\mathcal{P}_q^*$ provides the tightest lower bound among all feasible

q-metric projections. Optimality of $\mathcal{P}_q^*$ follows from the non-decreasing nature of the cost function f. ■

The optimality result in Proposition 22 provides an efficient way to bound the minimum cost of a combinatorial optimization problem. First, an *upper* bound can be achieved by finding a series of feasible solutions to the problem, e.g., particular circuits for the traveling salesman or cuts for the bisection problem, possibly aided by heuristics designed for the particular problem of interest. Second, we apply the canonical projection $\mathcal{P}_1^*$ to the network under analysis and solve efficiently the combinatorial problem in the obtained metric space [41], with the guarantee that the obtained *lower* bound is the tightest among those achievable via a metric projection that does not increase the dissimilarities.

8.4.2 Stability

For a given projection method $\mathcal{P}_q$, we can ask the question of whether networks that are close to each other result in q-metric spaces that are also close to each other. The answer to this question is affirmative for canonical projection methods $\mathcal{P}_q^*$, as we show in this section.

To formalize our concept of stability we need to quantify how close two given networks are, thus, we require a distance between networks. More specifically, given a node set V and an edge set E, we denote by $\mathcal{N}_{(V,E)}$ the set of all networks defined on nodes V and edges E. Notice that two networks in $\mathcal{N}_{(V,E)}$ might differ on the weights assigned to the common edges E. Hence, given two networks $G, G' \in \mathcal{N}_{(V,E)}$ such that $G = (V, E, W)$ and $G' = (V, E, W')$, we define the q-distance $f_q(G, G')$ between them as

$$f_q(G, G') := \left(\sum_{(x,x') \in E} |W(x, x') - W'(x, x')|^q \right)^{1/q}, \tag{8.27}$$

for $q \geq 1$. That (8.27) defines a valid distance in $\mathcal{N}_{(V,E)}$ follows immediately from noting that we are computing the q-norm of the difference of two vectors containing all the weights in each network [cf. (8.4)]. Moreover, recalling that every q-metric space can be seen as a (complete) network, f_q is also a valid distance between any pair of q-metric spaces sharing the same node set.

The following proposition bounds the distance between two canonically projected networks.

Proposition 23 *The canonical projection method $\mathcal{P}_q^*$ is stable in the sense that, for $G, G' \in \mathcal{N}_{(V,E)}$,*

$$f_q(\mathcal{P}_q^*(G), \mathcal{P}_q^*(G')) \leq K_n f_q(G, G'), \quad \text{with } K_n = \binom{n}{2}^{1/q}, \tag{8.28}$$

for all V and E.

Proof: Given $G = (V, E, W)$ and $G' = (V, E, W')$, denote by $M = (V, d) = \mathcal{P}_q^*(G)$ and $M' = (V, d') = \mathcal{P}_q^*(G')$ the corresponding canonically projected q-metric spaces. Also, denote by η the q-distance between the original networks G and G', i.e. $\eta = f_q(G, G')$. Further, pick arbitrary points $x, x' \in V$ and focus on a path $P_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ achieving the minimum for $d(x, x')$ in (8.8). Thus, if we consider this same path in network G', we have that

$$d'(x, x') \leq \left(\sum_{i=0}^{l} W'(x_i, x_{i+1})^q \right)^{1/q}. \tag{8.29}$$

Given that the set of edges E is a superset of the set of links in $P_{xx'}$, it follows that

$$\begin{aligned} f_q(G, G') = \eta &\geq \left(\sum_{i=0}^{l} \left| W(x_i, x_{i+1}) - W'(x_i, x_{i+1}) \right|^q \right)^{1/q} \\ &\geq \left| \left(\sum_{i=0}^{l} W(x_i, x_{i+1})^q \right)^{1/q} - \left(\sum_{i=0}^{l} W'(x_i, x_{i+1})^q \right)^{1/q} \right|, \end{aligned} \tag{8.30}$$

where the second inequality follows from the fact that, given two generic vectors $\mathbf{a}$ and $\mathbf{b}$, we have that $\|\mathbf{a} - \mathbf{b}\|_q \geq |\|\mathbf{a}\|_q - \|\mathbf{b}\|_q|$. We may rewrite (8.30) to obtain

$$\left(\sum_{i=0}^{l} W'(x_i, x_{i+1})^q \right)^{1/q} \leq \left(\sum_{i=0}^{l} W(x_i, x_{i+1})^q \right)^{1/q} + \eta. \tag{8.31}$$

Upon substitution of (8.29) into (8.31) and noting that the first term on the right-hand side of (8.31) is exactly $d(x, x')$ due to the particular choice of the path $P_{xx'}$, it follows that

$$d'(x, x') \leq d(x, x') + \eta. \tag{8.32}$$

Similarly, by starting with a different path $P'_{xx'}$ achieving the minimum for $d'(x, x')$ in (8.8) we may conclude that $d(x, x') \leq d'(x, x') + \eta$ and, by combining this with (8.32) we obtain that

$$|d(x, x') - d'(x, x')| \leq \eta. \tag{8.33}$$

By definition, the q-distance between M and M' is given by

$$f_q(M, M') = \left(\sum_{x, x' \in V} |d(x, x') - d'(x, x')|^q \right)^{1/q}, \tag{8.34}$$

which we may combine with (8.33) and the fact that there are n choose 2 possible pairs of nodes to write

$$f_q(M, M') \leq \left(\sum_{x,x' \in V} \eta^q \right)^{1/q} \leq \binom{n}{2}^{1/q} \eta. \tag{8.35}$$

To conclude the proof, recall that $\eta = f_q(G, G')$. ■

Intuitively, a projection method $\mathcal{P}_q$ is stable if its application to networks that have small distance between each other results in q-metric spaces that are close to each other. Formally, we require the distance between output q-metric spaces to be bounded by the distance between the original networks times a constant that depends on the size of the networks. This is important when we consider noisy dissimilarity data in the original networks. Proposition 23 ensures that noise has limited effect on output q-metric spaces.

Notice that the constant K_n in (8.28) decays rapidly with increasing q. In particular, the above analysis can be extended to $q = \infty$ to obtain the following corollary.

Corollary 9 *Given two networks $G = (V, E, W)$ and $G' = (V, E, W')$ and their canonical projections onto ultrametric spaces $\mathcal{P}^*_\infty(G) = (V, d)$ and $\mathcal{P}^*_\infty(G') = (V, d')$, we have that*

$$\max_{x,x' \in V} |d(x, x') - d'(x, x')| \leq \max_{(x,x') \in E} |W(x, x') - W'(x, x')|. \tag{8.36}$$

The above corollary ensures that single linkage hierarchical clustering (cf. Section 8.3.2) is stable in the sense that its application to a pair of networks does not increase the maximum difference among all their dissimilarities. In particular, if G' is a perturbed version of G where the maximum perturbation is denoted by η then the application of single linkage on G' generates an ultrametric which differs in less than η from the one that we would have obtained by clustering G.

Remark 20 At first sight, Corollary 9 might resemble a restatement of Theorem 14 for symmetric networks. Notice however that the Gromov-Hausdorff distance considered in Section 5.4 differs from the one introduced here. In the current section we are restricting the stability analysis to perturbations in the edge weights and, thus, we utilize the node labels in the definition of the distance. Formally, for $q = \infty$, our notion of distance here coincides with that of distortion in (5.79) where the bijection is given by the identity map.

8.4.3 Nestedness

From the increasing structure imposed by the q-triangle inequality as q increases, it follows that $\mathcal{M}_{q'} \subseteq \mathcal{M}_q$ for $q' \geq q$. This implies that the canonical projection of any network

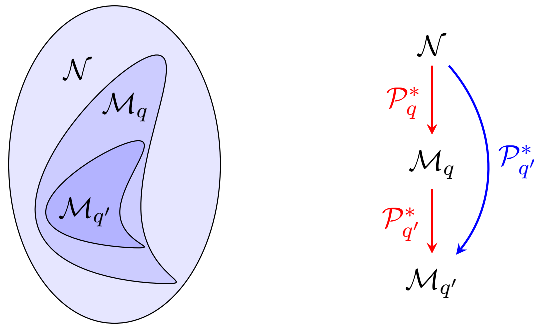


Figure 8.5: Nestedness of canonical projections. The canonical projection of a network in $\mathcal{N}$ onto $\mathcal{M}_{q'}$ (blue) is invariant to intermediate canonical projections onto spaces $\mathcal{M}_q$ (red) for $q \leq q'$.

onto a q'-metric space can be alternatively achieved by a direct application of $\mathcal{P}_{q'}^*$ or by first applying $\mathcal{P}_q^*$ to the network and then applying $\mathcal{P}_{q'}^*$ to the resulting q-metric space; see Fig. 8.6. Both approaches are equivalent, as we formally state next.

Proposition 24 *Given an arbitrary network $G = (V, E, W)$, we have that, for $q' \geq q$,*

$$\mathcal{P}_{q'}^*(G) = \mathcal{P}_{q'}^*(\mathcal{P}_q^*(G)). \tag{8.37}$$

Proof : Define the q'-metric spaces $M_{q'}^1 = (V, d_1) := \mathcal{P}_{q'}^*(G)$ and $M_{q'}^2 = (V, d_2) := \mathcal{P}_{q'}^*(\mathcal{P}_q^*(G))$, as well as the q-metric space $M_q^0 = (V, d_0) := \mathcal{P}_q^*(G)$. To prove the proposition, we must show that $d_1(x, x') = d_2(x, x')$ for all $x, x' \in V$. Consider the diagram in Fig. 8.6, where the blue map and the two left-most green maps represent the aforementioned definitions and the remaining green map is justified by the Axiom of Projection. Assume that the identity map $\phi : V \to V$ where $\phi(x) = x$ for all $x \in V$ is dissimilarity reducing from G to M_q^0 and that the same identity map is also dissimilarity reducing from M_q^0 to $M_{q'}^1$ (top row of Fig. 8.6). Then, from the fact that $\mathcal{P}_{q'}^*$ satisfies the Axiom of Injective Transformation, it would follow that $d_1(x, x') \geq d_2(x, x') \geq d_1(x, x')$ for all $x, x' \in V$ (bottom row of Fig. 8.6), showing the desired equality. Thus, to complete the proof we need to show that the identity maps ϕ are effectively dissimilarity reducing. That ϕ is dissimilarity reducing from G to M_q^0 was shown in the proof of Theorem 17 [cf. (8.19)]. Finally, to see that ϕ is dissimilarity reducing from M_q^0 to $M_{q'}^1$, pick an arbitrary pair of nodes $x, x' \in V$ and let $P_{xx'} = [x = x_0, \ldots, x_l = x']$ be a path achieving the minimum in (8.8) for d_0 so that we can write

$$d_0(x, x') = \left(\sum_{i=0}^{l-1} W(x_i, x_{i+1})^q \right)^{1/q} \geq \left(\sum_{i=0}^{l-1} W(x_i, x_{i+1})^{q'} \right)^{1/q'} \geq d_1(x, x'), \tag{8.38}$$

$$
\begin{tikzcd}
\mathbf{G} \arrow[r, "\mathcal{P}_q^*"] \arrow[d, "\mathcal{P}_{q'}^*"'] \arrow[dr, phantom, "\phi = \mathrm{Id}"] & \mathbf{M_q^0} \arrow[r, dashed] \arrow[d, "\mathcal{P}_{q'}^*"] \arrow[dr, phantom, "\phi = \mathrm{Id}"] & \mathbf{M_{q'}^1} \arrow[d, "\mathcal{P}_{q'}^*"] \\
\mathbf{M_{q'}^1} & \mathbf{M_{q'}^2} & \mathbf{M_{q'}^1}
\end{tikzcd}
$$

Figure 8.6: Diagram of maps between spaces for the proof of Proposition 24. Since $\mathcal{P}_{q'}^*$ satisfies axiom (AA2), ϕ being dissimilarity reducing allows us to show that spaces $M_{q'}^1$ and $M_{q'}^2$ are equivalent.

where the first inequality is obtained by combining the facts that $q' \geq q$ and the weights W are positive, and the second inequality comes from the fact that $d_1(x, x')$ is obtained by minimizing over all paths from x to x' and here we are considering a particular path $P_{xx'}$. From (8.38) it follows that the identity map is dissimilarity reducing from M_q^0 to $M_{q'}^1$, concluding the proof. ■

Proposition 24 shows that the q'-metric space associated with a given network G is independent of any intermediate canonical projections to q-metric spaces for $q \leq q'$. Intuitively, this result implies that the intermediate map $\mathcal{P}_q^*$ induces part of the structure imposed by $\mathcal{P}_{q'}^*$. Thus, it is equivalent to induce the whole structure in one step by applying $\mathcal{P}_{q'}^*$ or doing it gradually by applying maps $\mathcal{P}_q^*$ for $q \leq q'$.

A direct consequence of (8.37) is that if one is interested in, e.g., computing the single linkage hierarchical clustering output of a given network G (cf. Section 8.3.2) then there is no gain (or loss) in first projecting the network G onto a metric space and then computing the clustering output of the resulting metric space.

Chapter 9

Admissible projections for asymmetric networks

In Chapter 8 we studied how to project symmetric networks onto q-metric spaces. In the current chapter we extend the domain of our projections to include all (possibly asymmetric) networks but restrain the image to 1-metric spaces. Extensions for projections onto q-metric spaces for $q \neq 1$ and quasi-metric spaces are discussed in Chapter 10. Given that we focus on asymmetric networks, some results in the current chapter have a strong resemblance with the main results found in Part I.

After extending the axiomatic framework to accommodate asymmetric networks (Section 9.1), in Section 9.2 we introduce two admissible projection methods SymPro and ProSym. These methods, closely related to reciprocal and nonreciprocal clustering, can be shown to bound all other admissible metric projections (Section 9.3).

9.1 Axiom of Symmetrization

The ideal projection map $\mathcal{P} : \tilde{\mathcal{N}} \to \mathcal{M}$ is one that enforces a metric structure on the projected network but, at the same time, preserves distinctive features of the original network. In order to design such projection operators, we follow the same strategy as the rest of the thesis and adopt an axiomatic approach.

Since networks might be directed while metric spaces are symmetric, we introduce the concept of a symmetrizing function.

Definition 6 A symmetrizing function $s : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is one that satisfies for all $a, b, c, d, a' \geq a, b' \geq b$:

(i) Identity: $s(a, a) = a$ and $s(a, b) = 0 \iff a = b = 0$.

(ii) Symmetry: $s(a, b) = s(b, a)$.

(iii) Monotonicity: $s(a, b) \leq s(a', b')$.

(iv) Subadditivity: $s(a + b, c + d) \leq s(a, c) + s(b, d)$.

Common examples of valid symmetrizing functions include $s(a, b) = \max(a, b)$ and $s(a, b) = (a + b)/2$, although more convolved examples can be constructed. Nevertheless, we are agnostic to the particular choice of symmetrizing function and, in turn, the validity of our results hold for any function abiding by Definition 6. Given a symmetrizing function and a possibly asymmetric network $G = (V, E, W)$, we define the symmetrizing map $\mathcal{S}$ as one that generates $\mathcal{S}(G) := (V, \mathcal{S}(E), \mathcal{S}(W))$ such that $(x, x') \in \mathcal{S}(E)$ if both $(x, x') \in E$ and $(x', x) \in E$, and

$$\mathcal{S}(W)(x, x') := s(W(x, x'), W(x', x)), \tag{9.1}$$

for all $(x, x') \in \mathcal{S}(E)$. Moreover, given a path $P_{xx'}$ in G, we denote by $s(P_{xx'})$ a path containing the same nodes as $P_{xx'}$ but where the weights are given by those in $\mathcal{S}(G)$. From Definition 6 the following instrumental result follows.

Proposition 25 *The symmetrized network $\mathcal{S}(\tilde{M})$ is a metric space for all quasi-metric spaces $\tilde{M} \in \tilde{\mathcal{M}}$.*

Proof: Denoting by $(V, d) = \mathcal{S}(\tilde{M})$ the symmetrized version of $\tilde{M} = (V, \tilde{d})$, that d satisfies the symmetry property of metrics (cf. Definition 4) follows immediately from the symmetry condition in Definition 6. Moreover, the identity of d follows by combining the identity property of $\tilde{d}$ and the identity condition on s in Definition 6. Lastly, to see that d satisfies the triangle inequality, pick arbitrary nodes $x, y, z \in V$ and, leveraging the fact that $\tilde{d}$ is a quasi-metric we have that

$$\tilde{d}(x, z) \leq \tilde{d}(x, y) + \tilde{d}(y, z), \qquad \tilde{d}(z, x) \leq \tilde{d}(z, y) + \tilde{d}(y, x). \tag{9.2}$$

From the monotonicity property of s, it then follows that

$$d(x, z) = s\big(\tilde{d}(x, z), \tilde{d}(z, x)\big) \leq s\big(\tilde{d}(x, y) + \tilde{d}(y, z), \tilde{d}(y, x) + \tilde{d}(z, y)\big). \tag{9.3}$$

Finally, by recalling the subadditivity property of s we may further bound (9.3) as

$$d(x, z) \leq s\big(\tilde{d}(x, y), \tilde{d}(y, x)\big) + s\big(\tilde{d}(y, z), \tilde{d}(z, y)\big) = d(x, y) + d(y, z), \tag{9.4}$$

concluding the proof. ∎

Proposition 25 states that any valid symmetrizing function can induce symmetry in quasi-metric spaces – thus, transforming them into metric spaces – without disrupting the

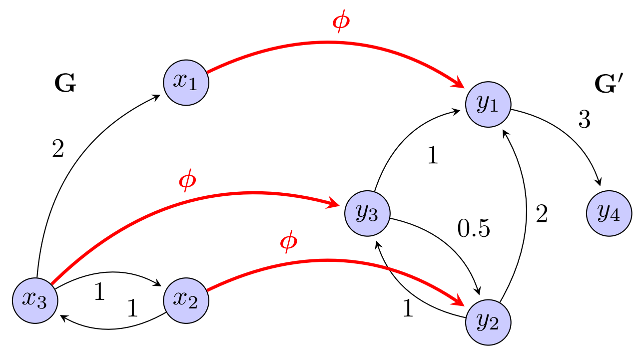


Figure 9.1: Dissimilarity reducing map for asymmetric networks. Map ϕ is injective and every existing edge in G is mapped to an edge in G' of equal or smaller weight.

identity or triangle inequality properties. Based on this observation, we state the following axiom to be satisfied by the metric projection $\mathcal{P}$.

> *(AS1) Axiom of Symmetrization.* For every quasi-metric space $\tilde{M} \in \tilde{\mathcal{M}}$, we must have that $\mathcal{P}(\tilde{M}) = \mathcal{S}(\tilde{M})$.

Under the premise that a reasonable projection $\mathcal{P}$ should not introduce unnecessary modifications to the input network, axiom (AS1) states that if a symmetrization is enough to obtain a metric space then $\mathcal{P}$ should implement this symmetrization and nothing more. Notice that by combining the fact that $\mathcal{M} \subset \tilde{\mathcal{M}}$ and the identity property of s, axiom (AS1) implies that $\mathcal{M}$ is a fixed set of $\mathcal{P}$. Equivalently, we have that $\mathcal{P}$ is equal to the identity map when restricted to the metric spaces in $\mathcal{M}$. Additionally, axiom (AS1) entails that $\mathcal{P}$ is idempotent meaning that $\mathcal{P}(\mathcal{P}(G)) = \mathcal{P}(G)$ since $\mathcal{P}(G) \in \mathcal{M}$ and $\mathcal{M}$ is a fixed set of $\mathcal{P}$.

As a second axiom, we adopt the Axiom of Injective Transformation (AA2) introduced in Section 8.2. Notice that the fact that we now include directed networks in our analysis does not invalidate the definition of a dissimilarity reducing map introduced in Section 8.2. In Fig. 9.1 we exemplify a dissimilarity reducing map for directed graphs.

Since it does not lead to confusion, we still say that a metric projection $\mathcal{P}$ is *admissible* if it satisfies axioms (AS1) and (AA2). In the next two section, we study the set of admissible projections.

9.2 SymPro and ProSym projections

We define the *symmetrize-then-project* (SymPro) map $\overline{\mathcal{P}} : \tilde{\mathcal{N}} \to \mathcal{M}$ that takes every possibly directed network $G = (V, E, W)$ into the metric space $\overline{M} = \overline{\mathcal{P}}(G) := (V, \overline{d})$ where $\overline{d}$ is defined as

$$\overline{d}(x,x') := \min_{P_{xx'}} \sum_{i=0}^{l-1} s\left(W(x_i,x_{i+1}), W(x_{i+1},x_i)\right) = \min_{P_{xx'}} \|s(P_{xx'})\|_1, \tag{9.5}$$

for all $x, x' \in V$. Notice that in (9.5) we first symmetrize the network using s and then we search for the shortest paths in the symmetrized network $\mathcal{S}(G)$. We restrict our analysis to networks $\mathcal{S}(G)$ that are connected, ensuring the validity of definition (9.5). Equivalently, the distance $\overline{d}(x,x')$ is given by the bidirectional path joining x and x' that attains the minimum symmetrized length; see Fig. 9.2.

For $\overline{\mathcal{P}}$ to be a properly defined metric projection method, we need to establish that $\overline{d}$ is indeed a valid metric. This is stated in the ensuing proposition, where we also show admissibility of $\overline{\mathcal{P}}$.

Proposition 26 *The SymPro metric projection map $\overline{\mathcal{P}}$ is valid and admissible. I.e., $\overline{d}$ defined in* (9.5) *is a metric for all networks and $\overline{\mathcal{P}}$ satisfies axioms (AS1) and (AA2).*

Proof: We first show that $\overline{d}$ is a valid metric. That $\overline{d}(x,x') = \overline{d}(x',x)$ follows from combining (9.5) with the symmetry property of s (cf. Definition 6), while the identity property of $\overline{d}$ follows from the identity condition on s. To verify that the triangle inequality holds, let $P_{xx'}$ and $P_{x'x''}$ be paths that achieve the minimum in (9.5) for $\overline{d}(x,x')$ and $\overline{d}(x',x'')$, respectively. Then, it holds that

$$\overline{d}(x,x'') \leq \|s(P_{xx'} \uplus P_{x'x''})\|_1 = \|s(P_{xx'})\|_1 + \|s(P_{x'x''})\|_1 = \overline{d}(x,x') + \overline{d}(x',x''), \tag{9.6}$$

where the inequality follows from the fact that the concatenated path $P_{xx'} \uplus P_{x'x''}$ is *a particular* path between x and x'' while (9.5) is a minimization across all such paths.

To see that the Axiom of Symmetrization (AS1) is fulfilled, pick an arbitrary quasi-metric space $\tilde{M} = (V,W)$ and denote by $(V,\overline{d})$ and $(V,\mathcal{S}(W))$ the outputs of applying $\overline{\mathcal{P}}$ and $\mathcal{S}$ to $\tilde{M}$, respectively. For an arbitrary pair of nodes $x,x' \in V$, we have that

$$\overline{d}(x,x') = \min_{P_{xx'}} \|s(P_{xx'})\|_1 \leq s\big(W(x,x'), W(x',x)\big) = \mathcal{S}(W)(x,x'), \tag{9.7}$$

where the inequality comes from specializing $P_{xx'}$ to the path with just one link from x to x'. Furthermore, if we denote by $P^*_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ a path achieving the minimum in (9.7), then we can leverage that $\mathcal{S}(W)$ satisfies the triangle inequality (cf. Proposition 25) to write

$$\mathcal{S}(W)(x,x') \leq \sum_{i=0}^{l-1} \mathcal{S}(W)(x_i,x_{i+1}) = \|s(P^*_{xx'})\|_1 = \overline{d}(x,x'). \tag{9.8}$$

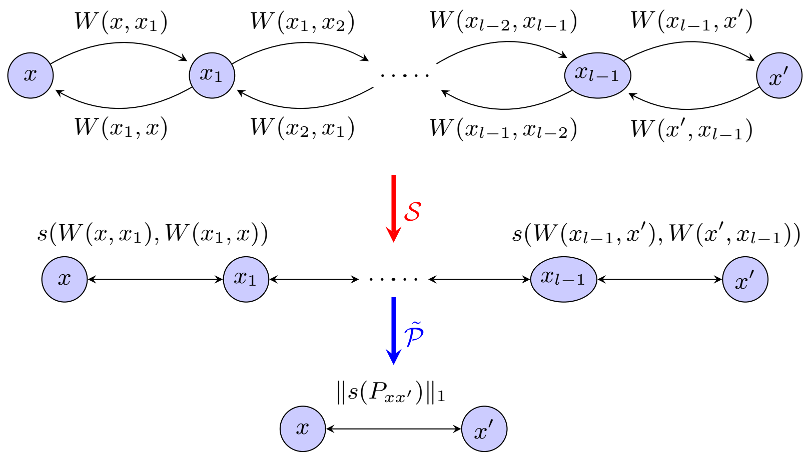


Figure 9.2: SymPro projection. After symmetrizing a bidirectional path $P_{xx'}$ between x and x' we compute its length $\|s(P_{xx'})\|_1$. The metric $\overline{d}(x, x')$ is given by the minimum length among all such paths.

Upon combining (9.7) and (9.8), it follows that $\overline{d}(x, x') = \mathcal{S}(W)(x, x')$. Since nodes x, x' were chosen arbitrarily it must be that $\overline{d} \equiv \mathcal{S}(W)$, implying that $\overline{\mathcal{P}}(\tilde{M}) = \mathcal{S}(\tilde{M})$.

In order to show that axiom (AA2) is satisfied, consider two networks $G = (V, E, W)$ and $G' = (V', E', W')$, a dissimilarity reducing map $\phi : V \to V'$, and denote by $(V, \overline{d}) = \overline{\mathcal{P}}(G)$ and $(V', \overline{d}') = \overline{\mathcal{P}}(G')$ the corresponding outputs of the SymPro projection. For an arbitrary pair of nodes $x, x' \in V$, denote by $P^*_{xx'} = [x = x_0, \ldots, x_l = x']$ a path that achieves the minimum in (9.5) so as to write $\overline{d}(x, x') = \|s(P^*_{xx'})\|_1$. Consider the transformed path $P^*_{\phi(x)\phi(x')} = [\phi(x) = \phi(x_0), \ldots, \phi(x_l) = \phi(x')]$ in the set V'. Combining the fact that the injective map ϕ does not increase dissimilarities and the monotonicity property of s (cf. Definition 6), it follows that

$$\|s(P^*_{\phi(x)\phi(x')})\|_1 \leq \overline{d}(x, x'). \tag{9.9}$$

Further note that $P^*_{\phi(x)\phi(x')}$ is a particular path joining $\phi(x)$ and $\phi(x')$ whereas the metric $\overline{d}'$ is given by the minimum across all such paths. Therefore,

$$\overline{d}'(\phi(x), \phi(x')) \leq \|s(P^*_{\phi(x)\phi(x')})\|_1. \tag{9.10}$$

Upon replacing (9.9) into (9.10), the required result follows. ■

The SymPro projection computes the distance between x and x' based on the length of symmetrized bidirectional paths. By contrast, the *project-then-symmetrize* (ProSym) projection relaxes the restriction of the paths being bidirectional and considers paths, possibly different, linking x to x' and x' to x. Formally, we define the ProSym map $\underline{\mathcal{P}} : \tilde{\mathcal{N}} \to \mathcal{M}$

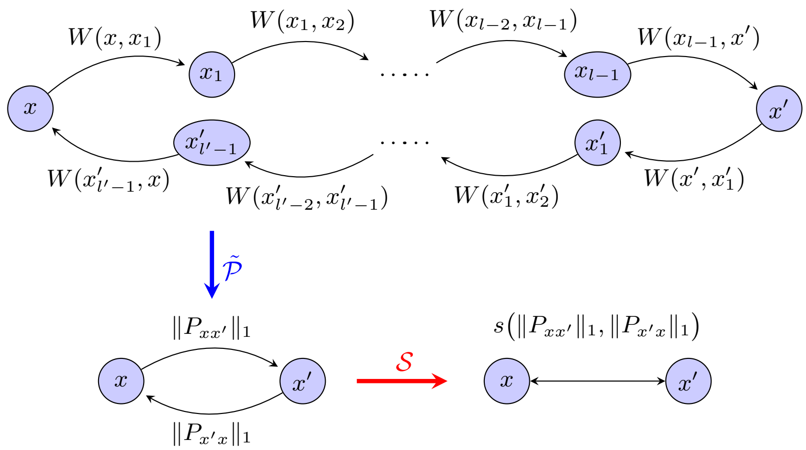


Figure 9.3: ProSym projection. We first find the length of the shortest paths from x to x' and from x' to x separately. The metric $\underline{d}(x, x')$ is given by the symmetrized value between these two.

such that the obtained metric space $\underline{M} = \underline{\mathcal{P}}(G) := (V, \underline{d})$ has a distance $\underline{d}$ defined as

$$\underline{d}(x, x') := s\Big(\min_{P_{xx'}} \sum_{i=0}^{l-1} W(x_i, x_{i+1}), \min_{P_{x'x}} \sum_{i=0}^{l'-1} W(x_i, x_{i+1})\Big) = s\Big(\min_{P_{xx'}} \|P_{xx'}\|_1\, , \min_{P_{x'x}} \|P_{x'x}\|_1\Big), \tag{9.11}$$

for all $x, x' \in V$. An illustration of the definition in (9.11) is shown in Fig. 9.3. We consider forward paths $P_{xx'}$ going from x to x' and backward paths $P_{x'x}$ going from x' to x, and we determine the length of each of these paths. We then search independently for the best forward and backward paths that minimize the respective lengths among all possible paths. The projected distance between x and x' is given by the symmetrized version of these two values.

As it is the case with SymPro, we can verify that $\underline{d}$ is a properly defined metric and that, as a consequence, the ProSym projection method is valid. The method $\underline{\mathcal{P}}$ also satisfies axioms (AS1)-(AA2) as the following proposition states.

Proposition 27 *The ProSym metric projection map $\underline{\mathcal{P}}$ is valid and admissible. I.e., $\underline{d}$ defined in* (9.11) *is a metric for all networks and $\underline{\mathcal{P}}$ satisfies axioms (AS1) and (AA2).*

Proof: The proof that $\underline{d}$ is indeed a metric is similar to that in Proposition 26 except for the part concerning the triangle inequality that we specify next. Let $P_{xx'}$ and $P_{x'x}$ be paths achieving the minimum in (9.11) for $\underline{d}(x, x')$ and similarly, let $P_{x'x''}$ and $P_{x''x'}$ be minimizing paths for $\underline{d}(x', x'')$. Then, it holds that

$$\begin{aligned} \underline{d}(x, x'') &\leq s(\|P_{xx'} \uplus P_{x'x''}\|_1, \|P_{x''x'} \uplus P_{x'x}\|_1) = s(\|P_{xx'}\|_1 + \|P_{x'x''}\|_1, \|P_{x''x'}\|_1 + \|P_{x'x}\|_1) \\ &\leq s(\|P_{xx'}\|_1, \|P_{x'x}\|_1) + s(\|P_{x'x''}\|_1, \|P_{x''x'}\|_1) = \underline{d}(x, x') + \underline{d}(x', x''), \end{aligned} \tag{9.12}$$

where the first inequality follows from the fact that the concatenated paths form a particular choice of paths between x and x'', and the second inequality follows from the subadditivity property of s (cf. Definition 6).

To see that the Axiom of Symmetrization (AS1) is fulfilled, pick an arbitrary quasi-metric space $\tilde{M} = (V, W)$ and denote by $(V, \underline{d})$ and $(V, \mathcal{S}(W))$ the outputs of applying $\underline{\mathcal{P}}$ and $\mathcal{S}$ to $\tilde{M}$, respectively. For an arbitrary pair of nodes $x, x' \in V$, we have that

$$\underline{d}(x, x') = s\big(W(x, x'), W(x', x)\big) = \mathcal{S}(W)(x, x'), \tag{9.13}$$

where the first equality follows from combining the monotonicity of s with the fact that $\min_{P_{xx'}} \|P_{xx'}\|_1 = W(x, x')$ since W satisfies a directed version of the triangle inequality. Since nodes x, x' were chosen arbitrarily, it must be that $\underline{d} \equiv \mathcal{S}(W)$ which implies that $\underline{\mathcal{P}}(\tilde{M}) = \mathcal{S}(\tilde{M})$, as wanted. Finally, the proof that $\underline{\mathcal{P}}$ satisfies axiom (AA2) is similar to that in Proposition 26 and, thus, is omitted here. ■

The metric spaces obtained by applying $\overline{\mathcal{P}}$ and $\underline{\mathcal{P}}$ to an example network G where the symmetrizing function is $s(a, b) = \max(a, b)$ are shown in Fig. 9.4. In the SymPro projection, we first symmetrize G to the maximum of the weights in both directions, whenever both edges exist. In this particular example, the symmetrized edge between a and c has a weight of $s(1, 3) = 3$, the one between c and b has a weight of 2 and there is no edge between a and b in $\mathcal{S}(G)$. We then compute the shortest paths between every pair of points in this symmetrized network to obtain $\overline{M}$ as in Fig. 9.4. By contrast, for $\underline{M}$ we first compute the unidirectional shortest paths between every pair of points and then symmetrize. For example, the unidirectional shortest path from a to c is given by the path through b with a total length of 2, and when symmetrized gives rise to $\underline{d}(a, c) = 2$. In a similar manner, the rest of the metric space $\underline{M}$ can be derived.

9.3 Extreme metric projections

Given that we have established the existence of two projection methods satisfying axioms (AS1)-(AA2), the question whether these two constructions are the only possible ones arises and, if not, whether they are special in some sense. We prove in this section that SymPro and ProSym bound all possible admissible projections in a well-defined sense, a result akin to that in Section 3.2 for hierarchical clustering. To explain this more clearly, first notice

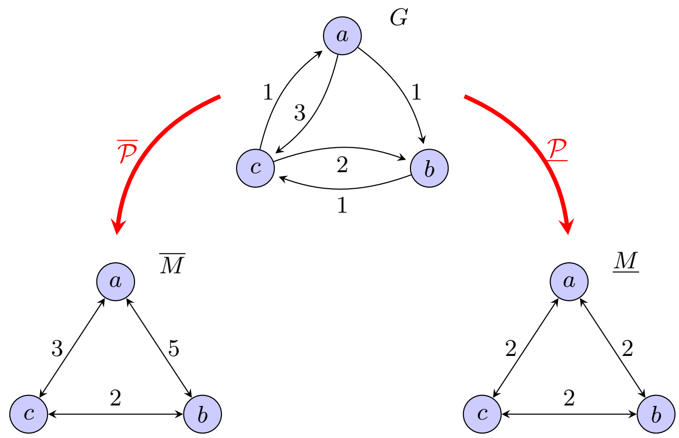


Figure 9.4: The metric spaces $\overline{M}$ and $\underline{M}$ obtained by applying the SymPro and ProSym projection methods to the network G, respectively.

that in Fig. 9.4 every distance in $\underline{d}$ is upper bounded by the corresponding distance in $\overline{d}$. As we show next, this is not an artifact of this example but rather a feature of the projection methods. More importantly, the distances induced by any other admissible projection method are contained between the distances obtained SymPro and ProSym, as stated next.

Theorem 18 *Consider an admissible projection method $\mathcal{P}$ satisfying axioms (AS1)-(AA2). For an arbitrary network $G = (V, E, W)$ denote by $(V, d) = \mathcal{P}(G)$ the output of $\mathcal{P}$ applied to G. Then, for all pairs of nodes $x, x' \in V$*

$$\underline{d}(x, x') \leq d(x, x') \leq \overline{d}(x, x'), \tag{9.14}$$

where $\underline{d}$ and $\overline{d}$ denote the metrics obtained from the ProSym and SymPro projections as respectively defined in (9.11) *and* (9.5)*.*

Proof: We begin by showing that $d(x, x') \leq \overline{d}(x, x')$ for all $x, x' \in V$. For this we use a procedure similar to the one used in the proofs of Theorems 4 and 17. Consider an arbitrary pair of points x and x' and let $P_{xx'} = [x = x_0, \ldots, x_l = x']$ be a path achieving the minimum in (9.5) so that $\overline{d}(x, x') = \|s(P_{xx'})\|_1$. Focus now on a series of two-node networks $G_i = (V_i, E_i, W_i)$ for $i = 0, \ldots, l-1$, such that $V_i = \{z, z'\}$ and $E_i = \{(z, z'), (z', z)\}$ for all i but with different weights given by $W_i(z, z') = W(x_i, x_{i+1})$ and $W_i(z', z) = W(x_{i+1}, x_i)$. Since every network G_i is already a quasi-metric space and the method $\mathcal{P}$ satisfies the Axiom of Symmetrization (AS1), if we define $(\{z, z'\}, d_i) := \mathcal{P}(G_i)$ we must have that $d_i(z, z') = s(W(x_i, x_{i+1}), W(x_{i+1}, x_i))$.

Consider injective maps $\phi_i : \{z, z'\} \to V$ given by $\phi_i(z) = x_i$, $\phi_i(z') = x_{i+1}$ so as to map z and z' in G_i to subsequent points in the path $P_{xx'}$. This implies that maps ϕ_i are dissimilarity reducing since they are injective and the edges in G_i are mapped to edges of the

$$
\begin{tikzcd}
\mathbf{G} \arrow[r, "\tilde{\mathcal{P}}"] \arrow[r, bend right, "\phi = \mathrm{Id}"'] \arrow[d, "\mathcal{P}"'] & (\mathbf{V}, \tilde{\mathbf{d}}) \arrow[d, "\mathcal{P} \equiv \mathcal{S}"] \\
(\mathbf{V}, \mathbf{d}) & (\mathbf{V}, \underline{\mathbf{d}})
\end{tikzcd}
$$

Figure 9.5: Diagram of maps between spaces for the proof of Theorem 18. Since $\mathcal{P}$ satisfies axiom (AA2), the existence of the dissimilarity reducing map ϕ allows us to relate d and $\underline{d}$.

exact same weight in G for all i. Thus, it follows from the Axiom of Injective Transformation (AA2) that

$$d(\phi_i(z), \phi_i(z')) = d(x_i, x_{i+1}) \leq d_i(z, z') = s(W(x_i, x_{i+1}), W(x_{i+1}, x_i)). \tag{9.15}$$

To complete the proof we use the fact that since d is a metric and $P_{xx'} = [x = x_0, \ldots, x_l = x']$ is a path joining x and x', the triangle inequality dictates that

$$d(x, x') \leq \sum_{i=0}^{l-1} d(x_i, x_{i+1}) \leq \|s(P_{xx'})\|_1 = \overline{d}(x, x'), \tag{9.16}$$

as wanted, where we used (9.15) for the second inequality and the last equality follows from the original choice of the path $P_{xx'}$.

We now show that $d(x, x') \geq \underline{d}(x, x')$ for all $x, x' \in V$. To do this, recall the definition of $\tilde{d}$ from (8.6) and notice that

$$\tilde{d}(x, x') \leq W(x, x'), \tag{9.17}$$

since the shortest path between two nodes cannot be larger than the direct path between them. Hence, the identity map $\phi = \mathrm{Id} : V \to V$ such that $\phi(x) = x$ for all $x \in V$ is a dissimilarity reducing map from G to $(V, \tilde{d})$, since it is injective and every existing edge in G is mapped to an edge with smaller or equal weight. Also, by comparing (8.6) and (9.11) it follows that $\underline{d}(x, x') = s(\tilde{d}(x, x'), \tilde{d}(x', x))$. Consequently, we can build the diagram of relations between spaces depicted in Fig. 9.5. The top (blue) and left (red) maps represent the definitions of the canonical quasi-metric projection $\tilde{\mathcal{P}}$ and the generic metric projection $\mathcal{P}$, respectively. The fact that for the right (green) map we have $\mathcal{P} \equiv \mathcal{S}$ is a consequence of axiom (AS1). Since the aforementioned identity map ϕ is dissimilarity reducing, we can use the fact that $\mathcal{P}$ satisfies axiom (AA2) to state that

$$d(x, x') \geq \underline{d}(\phi(x), \phi(x')) = \underline{d}(x, x'), \tag{9.18}$$

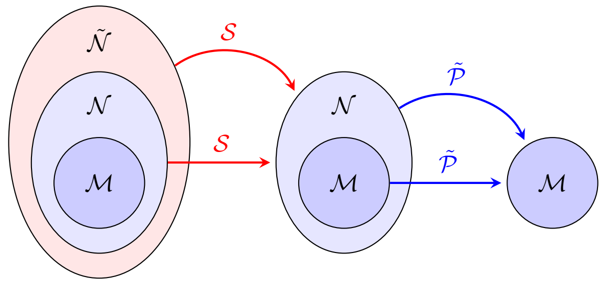


Figure 9.6: The SymPro $\overline{\mathcal{P}}$ projection as the composition of the symmetrizing map $\mathcal{S}$ followed by the projection $\tilde{\mathcal{P}}$.

for all $x, x' \in V$, concluding the proof. ■

According to Theorem 18, ProSym applied to a given network $G = (V, E, W)$ yields a uniformly minimal metric among those output by all admissible metric projections whereas SymPro yields a uniformly maximal metric. Equivalently, any projection methods that outputs a pairwise distance smaller than that obtained by ProSym or larger than that achieved by SymPro must violate at least one of the axioms (AS1)-(AA2).

The extreme nature of ProSym and SymPro can be understood from an intuitive perspective. In order to transform a general dissimilarity into a metric, we need to impose the symmetry and triangle inequality properties. Two conceivable ways of enforcing the aforementioned properties can be derived by imposing them in a sequential order, i.e., first symmetrizing and then enforcing the triangle inequality – giving rise to the SymPro –, or vice versa – entailing the ProSym method. A schematic representation of both methods is depicted in Figs. 9.6 and 9.7. Both methods $\overline{\mathcal{P}}$ and $\underline{\mathcal{P}}$ map the set of all possibly directed networks $\tilde{\mathcal{N}}$ into the set of metric spaces $\mathcal{M}$, however, from (9.5) it follows that $\overline{\mathcal{P}} = \tilde{\mathcal{P}} \circ \mathcal{S}$ as illustrated in Fig. 9.6. Notice that we first apply $\mathcal{S}$ to map $\tilde{\mathcal{N}}$ into the set of symmetric networks $\mathcal{N}$ and then apply the canonical projection. The opposite is true for $\underline{\mathcal{P}} = \mathcal{S} \circ \tilde{\mathcal{P}}$ where we first map $\tilde{\mathcal{N}}$ onto $\tilde{\mathcal{M}}$ via the canonical quasi-metric projection and then symmetrize the quasi-metric spaces using $\mathcal{S}$. Notice that $\mathcal{M}$ is a fixed set of the constituent maps $\mathcal{S}$ and $\tilde{\mathcal{P}}$. Consequently, $\mathcal{M}$ is a fixed set of both $\overline{\mathcal{P}}$ and $\underline{\mathcal{P}}$. This is as it should be for all admissible projections, as discussed after introducing axiom (AS1).

9.3.1 Metric projections for symmetric networks

If we restrict attention to the subset $\mathcal{N} \subset \tilde{\mathcal{N}}$ of symmetric networks, SymPro and ProSym are equivalent methods since both reduce to $\tilde{\mathcal{P}}$; see Figs. 9.6 and 9.7. To see this formally, consider a symmetric network $G = (V, E, W)$ and observe that as a consequence of the identity property of s, the symmetrization in (9.5) becomes unnecessary. Thus, the definition of $\overline{d}$ is reduced to $\overline{d}(x, x') = \min_{P_{xx'}} \|P_{xx'}\|_1 = \tilde{d}(x, x')$. Further note

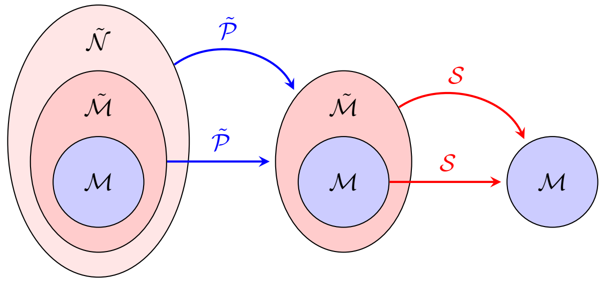


Figure 9.7: The ProSym $\underline{\mathcal{P}}$ projection as the composition of the projection $\tilde{\mathcal{P}}$ followed by the symmetrizing map $\mathcal{S}$.

that the weights of any given path $P_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ and its reciprocal $P_{x'x} = [x' = x_l, x_{l-1}, \ldots, x_0 = x]$ are the same. It follows then that $\underline{d}$ reduces to

$$\underline{d}(x, x') = s\Big(\min_{P_{xx'}} \|P_{xx'}\|_1 \, , \, \min_{P_{x'x}} \|P_{x'x}\|_1 \Big) = \min_{P_{xx'}} \|P_{xx'}\|_1 = \overline{d}(x, x'), \tag{9.19}$$

where again we used the identity property of s to write the second equality. The equivalence between SymPro and ProSym in (9.19) along with Theorem 18 demonstrates that when projecting symmetric networks onto metric spaces there exist a unique method satisfying (AS1)-(AA2) which coincides with the metric given by the shortest path between each pair of points. This is consistent with the uniqueness result in Section 8.3 for $q = 1$. More specifically, Theorem 18 describes a bounded family of potential admissible methods for projecting possibly asymmetric networks. However, when particularizing this result to symmetric networks, the uniqueness result in Section 8.3 is recovered. Finally, it should be noted that in Section 8.3 projections to more general q-metric spaces are analyzed. Extensions of the framework here described to q-metric spaces can also be developed, as described in Section 10.2.

9.3.2 Intermediate metric projections

SymPro and ProSym bound the range of all admissible metric projection methods. As was seen in Fig. 9.4, $\overline{\mathcal{P}}$ and $\underline{\mathcal{P}}$ are different in general, thus, following a reasoning similar to that in Section 3.3 we may ask: Are there intermediate methods contained between SymPro and ProSym? The answer to this question is positive as we demonstrate in this section by effectively constructing an intermediate projection method.

In SymPro, we first symmetrize the dissimilarities between every pair of nodes x and x'; see Fig. 9.2. We can reinterpret this operation as being a symmetrization between the shortest paths from x to x' and vice versa containing *at most two nodes* in each direction. Project-then-symmetrize-then-project (PSP) is a generalization of this idea where paths consisting of at most t nodes in each direction are allowed. Recall from Section 3.3.3 that, given an integer $t \geq 2$, we denote by $P^t_{xx'}$ a path starting at x and finishing at x' with at most t nodes. We reserve the notation $P_{xx'}$ to denote paths from x to x' where no maximum is imposed on the number of nodes. Given an arbitrary network $G = (V, E, W)$, we denote by $W^{(t)}(x, x')$ the length of the shortest path from x to x' of at most t nodes

$$W^{(t)}(x, x') := \min_{P^t_{xx'}} \|P^t_{xx'}\|_1. \tag{9.20}$$

We define the family of PSP methods $\hat{\mathcal{P}}^{(t)}$ with output $(V, \hat{d}^{(t)}) = \hat{\mathcal{P}}^{(t)}(G)$ as the one for which the metric $\hat{d}^{(t)}$ between x and x' is given by

$$\hat{d}^{(t)}(x, x') = \min_{P_{xx'}} \sum_{i=0}^{l-1} s\big(W^{(t)}(x_i, x_{i+1}), W^{(t)}(x_{i+1}, x_i)\big). \tag{9.21}$$

We can interpret (9.21) as the application of SymPro [cf. (9.5)] to a network with dissimilarities $W^{(t)}$ given by the length of shortest paths of at most t nodes. Consequently, it follows from Proposition 26 that $\hat{d}^{(t)}$ is a validly defined metric. Moreover, PSP satisfies axioms (AS1)-(AA2) as stated next.

Proposition 28 *The PSP metric projection map $\hat{\mathcal{P}}^{(t)}$ is valid and admissible. I.e., $\hat{d}^{(t)}$ defined in* (9.21) *is a metric for all networks and $\hat{\mathcal{P}}^{(t)}$ satisfies axioms (AS1) and (AA2) for all integers $t \geq 2$.*

Proof: That $\hat{\mathcal{P}}^{(t)}$ outputs a valid metric was already established in the paragraph preceding the proposition, thus, we restrict our proof to showing fulfillment of axioms (AS1) and (AA2). That $\hat{\mathcal{P}}^{(t)}$ satisfies axiom (AS1) follows from observing that in quasi-metric spaces $W^{(t)}(x_i, x_{i+1}) = W(x_i, x_{i+1})$ for all $t \geq 2$ since the shortest paths are given by direct connections between nodes. Thus, when $\hat{\mathcal{P}}^{(t)}$ is applied to quasi-metric spaces, (9.21) reduces to (9.5), making $\hat{\mathcal{P}}^{(t)}$ and $\overline{\mathcal{P}}$ equivalent, and we have already shown fulfillment of (AS1) by $\overline{\mathcal{P}}$ in Proposition 26.

To see that $\hat{\mathcal{P}}^{(t)}$ satisfies (AA2), consider two networks $G = (V, E, W)$ and $G' = (V', E', W')$ and a dissimilarity reducing map $\phi : V \to V'$ between them. Further, denote by $P^*_{xx'} = [x = x_0, \ldots, x_l = x']$ a path that achieves the minimum in (9.21), so that we can write

$$\hat{d}^{(t)}(x, x') = \sum_{i=0}^{l-1} s\big(W^{(t)}(x_i, x_{i+1}), W^{(t)}(x_{i+1}, x_i)\big). \tag{9.22}$$

Consider the shortest paths $P^t_{x_i x_{i+1}}$ and $P^t_{x_{i+1} x_i}$ in G containing at most t nodes and of length $W^{(t)}(x_i, x_{i+1})$ and $W^{(t)}(x_i, x_{i+1})$, respectively. Further, focus on their images $P^t_{\phi(x_i)\phi(x_{i+1})}$ and $P^t_{\phi(x_{i+1})\phi(x_i)}$ in G' under ϕ. Since ϕ is dissimilarity reducing, $W'^{(t)}(\phi(x_i), \phi(x_{i+1})) \leq W^{(t)}(x_i, x_{i+1})$ and $W'^{(t)}(\phi(x_{i+1}), \phi(x_i)) \leq W^{(t)}(x_{i+1}, x_i)$ for all i. Thus, from the monotonicity property of s (cf. Definition 6) we can state that

$$s\big(W'^{(t)}(\phi(x_i),\phi(x_{i+1})),W'^{(t)}(\phi(x_{i+1}),\phi(x_i))\big)\leq s\big(W^{(t)}(x_i,x_{i+1}),W^{(t)}(x_{i+1},x_i)\big). \quad (9.23)$$

By combining (9.23) and the definition of $\hat{d}^{(t)}$ in (9.21), it follows that $\hat{d}^{(t)}(\phi(x),\phi(x'))\leq\hat{d}^{(t)}(x,x')$, concluding the proof. ■

PSP is a countable family of metric projections parameterized by integer $t\geq 2$ representing the allowed maximum node-length of shortest paths considered prior to symmetrization. SymPro and ProSym are equivalent to PSP for specific values of t. Since $W(x,x')=W^{(2)}(x,x')$ for all networks, a direct comparison of (9.5) and (9.21) suffices to see that $\hat{\mathcal{P}}^{(2)}\equiv\overline{\mathcal{P}}$, i.e., that PSP recovers SymPro when $t=2$. ProSym can be obtained as $\hat{\mathcal{P}}^{(t)}$ for any t larger than or equal to the number of nodes in the network analyzed. To see this, notice that minimizing over $P_{xx'}$ is equivalent to minimizing over $P_{xx'}^{(t)}$ for all $t\geq n$, since we are looking for minimizing paths in a network with non-negative dissimilarities. Therefore, visiting the same node twice is non-optimal. This implies that $P_{xx'}^{(n)}$ contains all possible minimizing paths between x and x'. Hence, $W^{(t)}(x,x')=\min_{P_{xx'}}\|P_{xx'}\|_1$ for all $t\geq n$. Comparing (9.19) and (9.21), it then follows that $\hat{d}^{(t)}\leq\underline{d}$ for all $t\geq n$ and, since $\hat{d}^{(t)}$ was shown to be admissible (cf. Proposition 28), Theorem 18 forces the previous inequality to be an equality. This indeed implies that $\hat{\mathcal{P}}^{(t)}\equiv\underline{\mathcal{P}}$ for all $t\geq n$.

SymPro and ProSym enforce the symmetry and the triangle inequality in the projected space in a sequential manner. By contrast, PSP can be intuitively understood as an intermediate method because it first forces a partial version of the triangle inequality, it then imposes symmetry and finally completes the enforcement of the triangle inequality. To make this intuition clearer, first notice that if a set of weights W satisfies the triangle inequality then the direct path between every pair x and x' is the shortest path between them, having length $W(x,x')$. For the intermediate weights $W^{(t)}$ employed in (9.21) we have that the direct path with weight $W^{(t)}(x,x')$ is guaranteed to be the shortest path from x to x' *among* those paths of node-length at most t nodes. In this sense, $W^{(t)}$ satisfies a partial version of the triangle inequality that, after symmetrization, is completed to ensure the validity of the PSP family of metric projections.

A complete characterization of intermediate projection methods as well as the determination of additional desirable properties that can be used in combination with axioms (AS1)-(AA2) to further winnow the space of admissible projections are appealing future research avenues as further discussed in Chapter 12.

Chapter 10

Extensions and applications of metric projections

Having laid the foundations for metric projection of asymmetric networks in Chapter 9, we now present natural extensions of the framework as well as applications of it. Regarding the former, in Section 10.1 we study projections onto quasi-metric spaces. The results in this section can be considered as the metric counterparts of those found in Chapter 4 for clustering. Moreover in Section 10.2 we extend our framework to include projections onto the more general set of q-metric spaces. In terms of applications, we first explore how metric projections can be leveraged to perform efficient search in networks (Section 10.3). More precisely, we analyze the procedure of first projecting networks onto q-metric spaces and then efficiently searching these spaces via natural generalizations of the so-called metric trees. Lastly, in Section 10.4 we examine how metric projections can facilitate the visualization of asymmetric data. It is empirically observed that the visualizations associated with different metric projections unveil diverse aspects of the data.

10.1 Quasi-metric projections

Symmetry and the fulfillment of the triangle inequality are two defining features of metric spaces. In some settings, such as metric embeddings (Section 10.4), both properties play a central role. However, in other cases, the appeal of metric spaces stems almost exclusively from the triangle inequality. E.g., nearest neighbor search in metric spaces can be speeded up by exploiting this latter property (Section 10.3). Whenever the triangle inequality but not the symmetry is an essential property for the application at hand, projecting the network onto a (symmetric) metric space might introduce unnecessary distortions. Hence, we study the projection of networks onto (asymmetric) quasi-metric spaces following an axiomatic approach mimicking the developments in previous sections.

Although the Axiom of Injective Transformation (AA2) can be kept unchanged, axiom (AS1) must be modified to accommodate the asymmetry in the output, giving rise to the directed axiom ($\tilde{\text{A}}$S1).

($\tilde{A}$S1) Directed Axiom of Symmetrization. For every quasi-metric space $\tilde{M} \in \tilde{\mathcal{M}}$, we must have that $\mathcal{P}(\tilde{M}) = \tilde{M}$.

Axiom ($\tilde{\text{A}}$S1) imposes the natural condition that when the network that we are trying to project onto a quasi-metric space is already a quasi-metric space, an admissible projection method must leave it unaltered. There is a unique projection onto quasi-metrics that satisfies ($\tilde{\text{A}}$S1) and (AA2), as stated next.

Proposition 29 *Let $\mathcal{P} : \tilde{\mathcal{N}} \to \tilde{\mathcal{M}}$ be a quasi-metric projection for asymmetric networks. If $\mathcal{P}$ satisfies axioms ($\tilde{A}$S1) and (AA2) then $\mathcal{P} \equiv \tilde{\mathcal{P}}$ with output quasi-metric as defined in* (8.6).

Proof: That $\tilde{d}$ as defined in (8.6) is a valid quasi-metric was already established in Section 8.1. Additionally, the fulfillment of ($\tilde{\text{A}}$S1) and (AA2) by $\tilde{\mathcal{P}}$ can be shown following a procedure similar to that used in the proof of Proposition 26 and, thus, we omit it here. To show the equivalence $\mathcal{P} \equiv \tilde{\mathcal{P}}$, we denote by $(V, d) = \mathcal{P}(G)$ and $(V, \tilde{d}) = \tilde{\mathcal{P}}(G)$ the output quasi-metric spaces when applying a generic admissible quasi-metric projection and the canonical quasi-metric projection, respectively, to network $G = (V, E, W)$. If we show that

$$\tilde{d}(x, x') \leq d(x, x') \leq \tilde{d}(x, x'), \tag{10.1}$$

for all $x, x' \in V$, this would imply that that $\mathcal{P} \equiv \tilde{\mathcal{P}}$ since G was selected as an arbitrary network.

In order to show that $d(x, x') \leq \tilde{d}(x, x')$ we may employ a reasoning based on two-node networks analogous to that used in proving that $d(x, x') \leq \overline{d}(x, x')$ in Theorem 18. In this case, we need to leverage the fact that every two-node network is a quasi-metric space and these must remain unchanged when the admissible projection $\mathcal{P}$ is applied to them [cf. axiom ($\tilde{\text{A}}$S1)].

Lastly, in showing that $\tilde{d}(x, x') \leq d(x, x')$, we may use a slight modification on the argument illustrated in Fig. 9.5, here depicted in Fig. 10.1. The top (blue) and left (red) maps correspond to the definitions of the corresponding projections whereas the right (green) map is equivalent to an identity due to axiom ($\tilde{\text{A}}$S1). After establishing that the identity map $\phi = \text{Id}$ is a dissimilarity reducing map from G to $(V, \tilde{d})$, axiom (AA2) forces $\tilde{d}(x, x') \leq d(x, x')$, as wanted. ■

Multiple admissible methods exist for the task of projecting asymmetric networks onto (symmetric) metric spaces. Firstly, in handling the asymmetry, we have the freedom to

$$
\begin{tikzcd}
\mathbf{G} \arrow[r, "\tilde{\mathcal{P}}"] \arrow[r, bend right, "\phi = \mathrm{Id}"'] \arrow[d, "\mathcal{P}"'] & (\mathbf{V}, \tilde{\mathbf{d}}) \arrow[d, "\mathcal{P}"] \\
(\mathbf{V}, \mathbf{d}) & (\mathbf{V}, \tilde{\mathbf{d}})
\end{tikzcd}
$$

Figure 10.1: Diagram of maps between spaces for the proof of Proposition 29.

choose a symmetrizing function s as long as it satisfies the conditions in Definition 6. In second place, even when s is fixed, Theorem 18 shows the existence of a potentially infinite set of admissible methods bounded between SymPro and ProSym of which the family of PSP methods is a concrete example. By contrast, when the interest is in projecting symmetric networks onto metric spaces, Theorem 17 and the discussion in Section 9.3.1 show that the landscape of admissible methods is much simpler, namely there is a unique admissible method for completing the task. This points at the fact that the symmetry mismatch between the input and the output of the projection maps is responsible for the added complexity in the set of admissible projections. Proposition 29 indicates that this is indeed the case. More precisely, when fixing the symmetry mismatch by allowing for asymmetric outputs, we recover a unique admissible method. Moreover, this unique way $\tilde{\mathcal{P}}$ of imposing quasi-metric structure on asymmetric networks coincides with the unique admissible projection for inducing metric structure on symmetric networks.

10.2 Projections onto q-metric spaces

A generalization of metric spaces are the so-called *q-metric spaces*, a larger class of structured spaces parametrized by $q \in [1, \infty]$ introduced in Section 8.1. The methodology outlined in Sections 9.1 to 10.1 can be extended to encompass the study of projections of weighted graphs onto q-metrics spaces. As will be seen, the additional degree of freedom granted by parameter q has both theoretical and practical implications. Among the former, it allows us to relate our results to seemingly unrelated fields of knowledge such as hierarchical clustering, and regarding the latter, the parameter q enables an accuracy/speed tradeoff when performing approximate nearest neighbor searches in directed graphs (Section 10.3).

We extend the shortest path notation $\tilde{d}$ to account for the length of the shortest path from x to x' as measured by its q-norm, i.e.,

$$
\tilde{d}_q(x, x') := \min_{P_{xx'}} \|P_{xx'}\|_q. \tag{10.2}
$$

It should be noted that $\tilde{d}_q$ is a valid quasi-q-metric, i.e., a q-metric that need not be symmetric. Moreover, we denote by $\tilde{\mathcal{P}}_q$ the canonical quasi-q-metric projection that maps any network G onto the space $(V, \tilde{d}_q)$ with the quasi-metric given by the shortest paths in G as in (10.2).

Our goal now is to devise node-preserving q-metric projections $\mathcal{P}_q : \tilde{\mathcal{N}} \to \mathcal{M}_q$ to map asymmetric networks onto q-metric spaces. As done before, we pursue this objective via an axiomatic construction.

A q-symmetrizing function $s : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is one that satisfies all the requirements in Definition 6 plus a fifth one (trivially satisfied when $q = 1$) for $a, b \geq 0$:

v) q-invariance: $s(a, b)^q = s(a^q, b^q)$.

We say that a symmetrizing function is ∞-invariant if it is q-invariant for q tending to ∞. Notice that $s(a, b) = \max(a, b)$ is a q-symmetrizing function for all q while $s(a, b) = (a+b)/2$ is not valid for $q > 1$. Denoting by $\mathcal{S}_q$ the symmetrizing map associated with an arbitrary q-symmetrizing function, the following generalization of Proposition 25 holds.

Proposition 30 *The symmetrized network $\mathcal{S}_q(\tilde{M}_q)$ is a q-metric space for all quasi-q-metric spaces $\tilde{M}_q \in \tilde{\mathcal{M}}_q$.*

Proof: The main difference with the proof of Proposition 25 is in showing the fulfillment of the q-triangle inequality. Denoting by $(V, d) = \mathcal{S}(\tilde{M}_q)$ the symmetrized version of $\tilde{M}_q = (V, \tilde{d})$, pick arbitrary nodes $x, y, z \in V$ and, leveraging the q-invariance property of s we have that

$$d(x, z)^q = s(\tilde{d}(x, z), \tilde{d}(z, x))^q = s(\tilde{d}(x, z)^q, \tilde{d}(z, x)^q). \tag{10.3}$$

Further, from the fact that $\tilde{d}$ is a quasi-q-metric it follows that

$$\begin{aligned} d(x, z)^q &\leq s(\tilde{d}(x, y)^q + \tilde{d}(y, z)^q, \tilde{d}(z, y)^q + \tilde{d}(y, x)^q) \\ &\leq s(\tilde{d}(x, y)^q, \tilde{d}(y, x)^q) + s(\tilde{d}(y, z)^q, \tilde{d}(z, y)^q), \end{aligned} \tag{10.4}$$

where the second inequality follows from the subadditivity of s. Finally, using again the q-invariance of s we have that

$$d(x, z)^q \leq s(\tilde{d}(x, y), \tilde{d}(y, x))^q + s(\tilde{d}(y, z), \tilde{d}(z, y))^q = d(x, y)^q + d(y, z)^q, \tag{10.5}$$

concluding the proof. ■

Based on Proposition 30, we propose the following modification of axiom (AS1).

(q-AS1) Axiom of q-Symmetrization. For every quasi-q-metric space $\tilde{M}_q \in \tilde{\mathcal{M}}_q$, we must have that $\mathcal{P}_q(\tilde{M}_q) = \mathcal{S}_q(\tilde{M}_q)$.

As was the case for (AS1), the generalized axiom (q-AS1) states that if a q-symmetrization is enough to obtain a metric space then $\mathcal{P}_q$ should implement this symmetrization and nothing more. The combination of axioms (q-AS1) and (AA2) forms the notion of q-admissibility, a characterization of desirable q-metric projections.

The generalized version $\overline{\mathcal{P}}_q : \tilde{\mathcal{N}} \to \mathcal{M}_q$ of SymPro (cf. Section 9.2) projects a network G onto the q-metric space $\overline{M}_q = \overline{\mathcal{P}}_q(G) := (V, \overline{d}_q)$ where $\overline{d}_q$ is defined as

$$\overline{d}_q(x, x') := \min_{P_{xx'}} \|s(P_{xx'})\|_q, \tag{10.6}$$

for all $x, x' \in V$. As was the case in (9.5), in (10.6) we first symmetrize the network using s and then we search for the path whose length is minimum as measured by the q-norm in the symmetrized network $\mathcal{S}(G)$. Analogously, we may define $\underline{\mathcal{P}}_q : \tilde{\mathcal{N}} \to \mathcal{M}_q$, the generalization of the ProSym projection with associated q-metric $\underline{d}_q$ defined as

$$\underline{d}_q(x, x') := s\Big(\min_{P_{xx'}} \|P_{xx'}\|_q \,,\, \min_{P_{x'x}} \|P_{x'x}\|_q\Big), \tag{10.7}$$

for all $x, x' \in V$. In accordance with Propositions 26 and 27 both methods $\overline{\mathcal{P}}_q$ and $\underline{\mathcal{P}}_q$ are valid and q-admissible, as stated next.

Proposition 31 *$\overline{d}_q$ defined in (10.6) and $\underline{d}_q$ defined in (10.7) are valid q-metrics and $\overline{\mathcal{P}}_q$ and $\underline{\mathcal{P}}_q$ satisfy axioms (q-AS1) and (AA2).*

Proof: The only significant different with the proofs in Propositions 26 and 27 is in showing that $\underline{d}_q$ is indeed a q-metric. More specifically, in showing the fulfillment of the q-triangle inequality. Let $P_{xx'}$ and $P_{x'x}$ be paths achieving the minimum in (10.7) for $\underline{d}_q(x, x')$ and similarly, let $P_{x'x''}$ and $P_{x''x'}$ be minimizing paths for $\underline{d}_q(x', x'')$. Then, it holds that

$$\begin{aligned} \underline{d}_q(x, x'')^q &\leq s(\|P_{xx'} \uplus P_{x'x''}\|_q, \|P_{x''x'} \uplus P_{x'x}\|_q)^q = s(\|P_{xx'} \uplus P_{x'x''}\|_q^q, \|P_{x''x'} \uplus P_{x'x}\|_q^q) \\ &= s(\|P_{xx'}\|_q^q + \|P_{x'x''}\|_q^q, \|P_{x''x'}\|_q^q + \|P_{x'x}\|_q^q), \end{aligned} \tag{10.8}$$

where the first inequality follows from the fact that the concatenated paths form a particular choice of paths between x and x'', and the first equality is due to the q-invariance of s. We can then use the subadditivity property of s followed by a second application of the q-

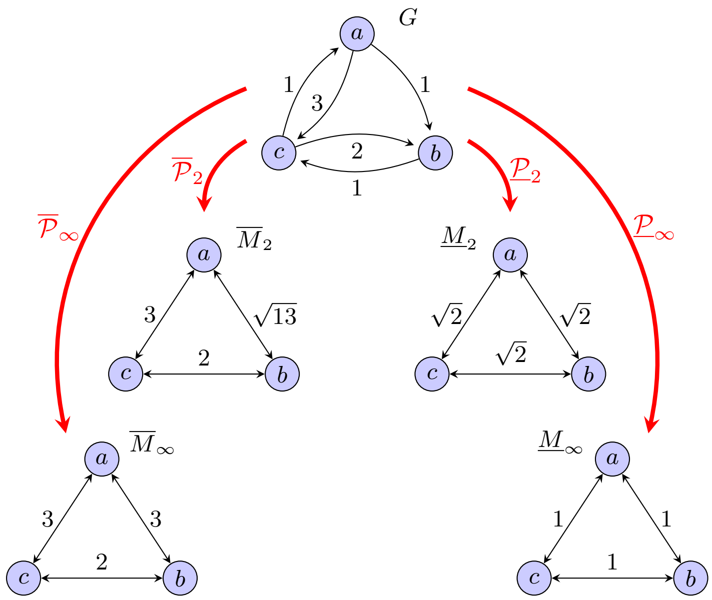


Figure 10.2: The q-metric spaces obtained by applying $\overline{\mathcal{P}}_q$ and $\underline{\mathcal{P}}_q$ to network G for $q = 2$ and $q = \infty$.

invariance to write

$$\begin{aligned} \underline{d}_q(x, x'')^q &\leq s(\|P_{xx'}\|_q^q, \|P_{x'x}\|_q^q) + s(\|P_{x'x''}\|_q^q, \|P_{x''x'}\|_q^q) \qquad (10.9) \\ &= s(\|P_{xx'}\|_q, \|P_{x'x}\|_q)^q + s(\|P_{x'x''}\|_q, \|P_{x''x'}\|_q)^q = \underline{d}_q(x, x')^q + \underline{d}_q(x', x'')^q, \end{aligned}$$

as wanted. ■

Continuing the example in Fig. 9.4, we compute the output q-metric spaces when applying $\overline{\mathcal{P}}_q$ and $\underline{\mathcal{P}}_q$ for $q \in \{2, \infty\}$ to the same G and using the same symmetrizing function $s(a, b) = \max(a, b)$; see Fig. 10.2. Comparing the outputs of $\overline{\mathcal{P}}_q$ for different q (including $q = 1$ in Fig. 9.4) it can be seen that as q increases, the edge weight between a given pair of points is not increased. E.g., the weight between nodes a and b decreases form 5 when $q = 1$ to 3 when $q = \infty$. The same phenomenon is observed for $\underline{\mathcal{P}}_q$ and, as can be shown, this is an immediate consequence of the fact that the norm $\|\mathbf{v}\|_p$ of any vector $\mathbf{v}$ decreases with increasing p. Another observable phenomenon in Fig. 10.2 is that the edge weights output by $\overline{\mathcal{P}}_q$ seem to dominate those output by $\underline{\mathcal{P}}_q$ for every value of q. Recall that Theorem 18 showed this property for $q = 1$, however, it is true for projections onto general q-metric spaces. Moreover, every other admissible q-metric is contained between these two, as stated next.

Theorem 19 *Consider an admissible projection method $\mathcal{P}_q$ satisfying axioms (q-AS1)-(AA2). For an arbitrary network $G = (V, E, W)$ denote by $(V, d_q) = \mathcal{P}_q(G)$, then for* *all pairs of nodes* $x, x' \in V$

$$\underline{d}_q(x, x') \leq d_q(x, x') \leq \overline{d}_q(x, x'), \tag{10.10}$$

where $\underline{d}_q$ *and* $\overline{d}_q$ *are defined as in* (10.7) *and* (10.6).

Proof: The proof that $d_q(x, x') \leq \overline{d}_q(x, x')$ follows the same steps as those in the proof of Theorem 18 by leveraging the fact that the two-node networks G_i there defined are valid quasi-q-metric spaces for all q. In showing that $\underline{d}_q(x, x') \leq d_q(x, x')$ we may also mimic the procedure in Theorem 18 but based on a slightly modified version of the diagram in Fig. 9.5 where the top (blue) map is replaced by the more general projection onto quasi-q-metric spaces $\tilde{\mathcal{P}}_q$. ■

Theorem 19 extends the extreme result in Theorem 18 from metric to general q-metric projections. This result is not surprising since $\overline{\mathcal{P}}_q$ and $\underline{\mathcal{P}}_q$ can be described as the composition of a symmetrizing map and a map inducing the q-triangle inequality, i.e., $\overline{\mathcal{P}}_q = \tilde{\mathcal{P}}_q \circ \mathcal{S}$ and $\underline{\mathcal{P}}_q = \mathcal{S} \circ \tilde{\mathcal{P}}_q$. Consequently, the same intuitive description of extremity introduced after Theorem 18 holds in this case as well. Finally, given inequality (10.10) it is reasonable to search for q-metric projections that lie between $\underline{\mathcal{P}}_q$ and $\overline{\mathcal{P}}_q$. Indeed, natural extensions of the PSP projections in Section 9.3 but based on the q-norm of the involved paths can be shown to satisfy axioms (q-AS1) and (AA2).

As explained in Section 10.1, it is sometimes the case that the asymmetry of the input data is a feature that we want to preserve in the output space. For these situations, we extend the analysis of projections onto quasi-metric spaces in Section 10.1 to accommodate the more general quasi-q-metric spaces as outputs.

In order to adjust the notion of admissibility to the setting at hand, we introduce the following modification on the Axiom of q-Symmetrization.

(q-$\tilde{A}$S1) Directed Axiom of q-Symmetrization. For every quasi-q-metric space $\tilde{M}_q \in \tilde{\mathcal{M}}_q$, we must have that $\mathcal{P}_q(\tilde{M}_q) = \tilde{M}_q$.

Axiom (q-$\tilde{\text{A}}$S1) encodes the requirement that whenever the input network is already a quasi-q-metric space then this space must remain unchanged by the projection map. There is a unique projection onto quasi-q-metric spaces that satisfies (q-$\tilde{\text{A}}$S1) and (AA2), as stated in the following proposition. Being similar to the proof of Proposition 29, the proof is omitted to avoid repetition.

Proposition 32 *Let* $\mathcal{P}_q : \tilde{\mathcal{N}} \to \tilde{\mathcal{M}}_q$ *be a quasi-q-metric projection for asymmetric networks. If* $\mathcal{P}_q$ *satisfies axioms (q-$\tilde{A}$S1) and (AA2) then* $\mathcal{P}_q \equiv \tilde{\mathcal{P}}_q$ *with output quasi-q-metric as defined in* (10.2).

Proposition 32 extends the uniqueness result in Proposition 29 from $q = 1$ to a general q, thus completing the generalization of the results presented in this second part of the thesis. In particular, this extends the validity of our study for projections onto ∞-metric or ultrametric spaces, as stated in the following remark.

Remark 21 (Hierarchical clustering methods) As discussed before, our study of projections onto ∞-metric spaces is equivalent to the study of hierarchical clustering of asymmetric networks. In Part I of this thesis, we derived two admissible methods – denominated reciprocal and nonreciprocal clustering – and showed that all other hierarchical clustering methods are contained between these two. Reciprocal and nonreciprocal clustering can be shown to coincide with our projection methods $\overline{\mathcal{P}}_\infty$ in (10.6) and $\underline{\mathcal{P}}_\infty$ in (10.7), respectively. Hence, the main result in Section 3.2 is a particular case of Theorem 19 for $q = \infty$. Moreover, in Section 3.3.3 the family of semi-reciprocal clustering methods was introduced as admissible methods contained between reciprocal and nonreciprocal clustering. It can be established that these semi-reciprocal methods coincide with the SPS methods here presented based on path lengths as measured by the ℓ_∞ norm.

10.3 Efficient search in networks

Given a network $G = (V, E, W)$, assume that we have access only to a subset of the network $G' = (V', E', W')$ where $V' \subset V$ and E' and W' are the restrictions of E and W to V', respectively. We are then revealed an additional point $z \in V \backslash V'$ and are interested in finding the node $x \in V'$ closest to z, i.e., the node x for which $W(z, x)$ is minimized. The described setting occurs frequently in practice, e.g., in the implementation of k nearest neighbor (k-NN) methods [4, Chapter 2] where G is the dataset of interest and G' is the training set. The complexity of the mentioned task depends on how structured network G is. When no structure is present, an exhaustive search is the only option and z must be compared with every node in V'. By contrast, when G is a metric space, then the NN of z can be found efficiently by using metric trees [83, 84, 91]. In this section, we propose an efficient search strategy in networks by first projecting a general network onto either a q-metric space or a quasi-q-metric space and then leveraging this structure for search via the construction of a generalized metric tree.

Intuitively, if z is far away from a node x in a metric space, i.e. $W(z, x)$ is large, then the triangle inequality implies that z will also be far away from any node x' close to x, thus, there is no need to consider node x' as a potential candidate for the NN of z. Metric trees formally leverage this intuition by constructing hierarchical structures of V' in order to accelerate search. Metric trees arise as an efficient alternative to other multidimensional space-partitioning structures, such as k-d trees [3], which do not leverage directly the metric

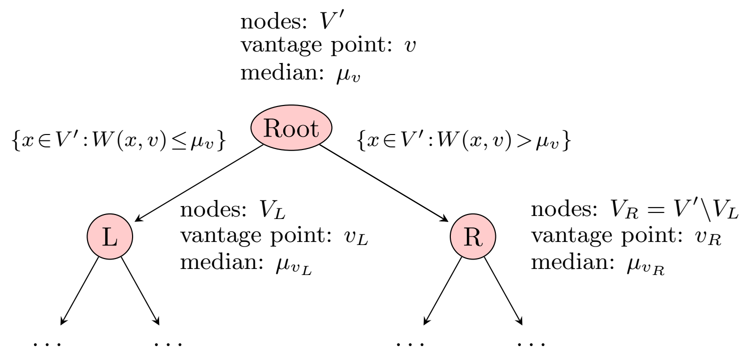


Figure 10.3: Vantage point tree. The whole point set V' is associated with the root of the tree. A vantage point v is chosen at random and V' is partitioned into the left and right children of the root depending on the distance of each point to the vantage point. The process is repeated iteratively to construct the whole tree.

structure of the data but require the data to be embedded in a coordinate space. In this section we focus on the vantage point (vp) tree [91], one of the most popular types of metric tree.

10.3.1 Search in metric spaces

The implementation of a metric tree is a two-step process: we first construct the tree and then utilize it for (possibly multiple) queries. To construct a vp tree given G', we begin by associating the whole node set V' to the root of the tree and we pick a node (the vantage point) at random, say $v \in V'$. We then compute the median μ_v of the distances $W(v, x)$ from the vantage point to every other node $x \in V'$ and partition V' into two blocks: one containing the nodes whose distance to v is smaller than or equal to μ_v and the other one containing the rest of V'. The nodes in the first block are assigned to the left child of the root of the vp tree while the right child consists of the nodes in the second block. We iteratively repeat this procedure within each of the children until every leaf in the vp tree is associated to a single point in V'; see Fig. 10.3. For more details, see [91].

To efficiently search a vp tree for the NN of a query point z, we traverse the nodes of the tree and compare z *only with the vantage point* of the current node of the vp tree. Moreover, we leverage the triangle inequality to discard branches of the vp tree without even traversing them, reducing the number of measurements needed to find the NN of z. More specifically, assume that we are searching at an intermediate node in the vp tree, say node L in Fig. 10.3 and the current best estimate of the NN is at distance τ from z, which can be initialized as $\tau = \infty$ for the root of the vp tree. We then compute the distance $W(z, v_L)$ between z and the vantage point v_L associated to the current node in the vp tree. If $W(z, v_L) < \tau$, we then update our estimate of τ. In order to continue traversing the vp

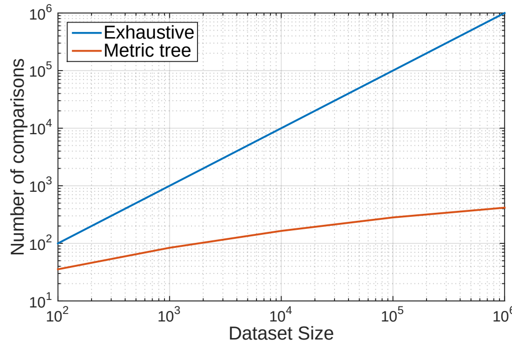


Figure 10.4: Number of comparisons needed to find the nearest neighbor of a point in a metric space as a function of the space size for exhaustive search and search aided using metric trees.

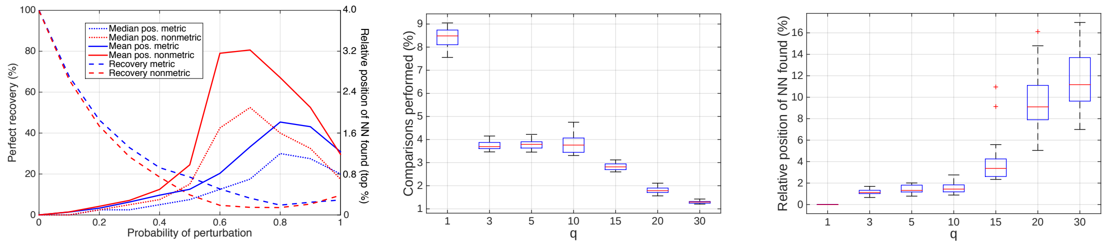
tree, we follow the ensuing rules where v is the vantage point of the current node in the vp tree

$$\begin{cases} \text{a) } W(z,v) \leq \mu_v - \tau & \Rightarrow \text{ visit only the left child,} \\ \text{b) } \mu_v - \tau < W(z,v) \leq \mu_v + \tau & \Rightarrow \text{ visit left \& right child,} \\ \text{c) } \mu_v + \tau < W(z,v) & \Rightarrow \text{ visit only right child.} \end{cases} \tag{10.11}$$

Even though statements a) and c) entail that we discard part of the nodes in V' during our search, the way the metric tree is constructed guarantees that the NN of z is not contained among the discarded nodes.

The construction of the vp tree, a one-time computational effort, can be shown to have complexity $\mathcal{O}(n \log n)$ where n is the cardinality of V'. However, once it is built it can be used to reduce the complexity of a brute force linear search from $\mathcal{O}(n)$ to an expected cost of $\mathcal{O}(\log n)$ [91]. To corroborate this, we construct metric spaces of varying sizes by embedding points in a square area of $\mathbb{R}^2$ and consider their Euclidean distance as the dissimilarity values W. In Fig. 10.4 we plot the average number of comparisons – values of W – needed to find the nearest neighbor of a query point in this metric space as a function of n for exhaustive and metric-tree search. This average is computed across 1,000 queries. As expected, exhaustive search complexity grows linearly with n whereas vp tree's complexity grows logarithmically. Notice that there is a marked difference in the number of measurements required, e.g., for $n = 10^6$ the metric tree search can be performed with an expected cost of around 500 measurements.

Motivated by the computational gain depicted in Fig. 10.4, a possible way to search a non-metric symmetric network G is to first project it onto a metric space M via the canonical projection $M = \mathcal{P}_1^*(G)$ and then construct a vp tree on M. Notice that this construction guarantees an efficient search of the NN in M of a given query. However, we



(a) (b) (c)

Figure 10.5: (a) Percentage of perfect recovery (dashed lines, left y-axis), and mean and median relative positions of search result (solid and pointed lines, right y-axis) as a function of the probability of perturbation in a metric network when the tree search is performed in the resulting non-metric space (red) and when the space is previously projected using $\mathcal{P}_1^*$ (blue). (b) Number of comparisons needed to search a metric space when projected first onto a q-metric space via $\mathcal{P}_q^*$ as a function of q. Larger q correspond to more efficient searches. (c) Search performance as indicated by the relative position of the NN found when performing the searches in (b). Smaller q correspond to more accurate searches.

are interested in finding the NN in G, thus, potentially committing an error. Intuitively, the furthest away the structure of G is from being metric, the larger the error in the NN found. In order to illustrate this effect, we generate metric spaces obtained by randomly embedding 1,000 points in $\mathbb{R}^{100}$ and considering their Euclidean distances as dissimilarities between them. We then obtain (non-metric) perturbed versions of each metric space by multiplying a subset of the dissimilarities by $1 + \delta$ where δ is a random variable uniformly distributed in $[0, 10]$. The subset of dissimilarities to modify is chosen randomly with probability of perturbation r. In Fig. 10.5(a) we illustrate the average search performance over 1,000 queries as a function of r (blue lines). The dashed line illustrates the percentage of perfect recovery (left y-axis), i.e., the proportion of the 1,000 queries in which the node found coincides with the actual NN of the query point. The solid and the pointed lines represent, respectively, the mean and median relative positions of the actual node found (right y-axis). E.g., a value in 0%–1% indicates that the node found is actually contained among the 10 nearest nodes (1% of 1000) to the query. Finally, to demonstrate the value of the projection method proposed, we also illustrate the search performance when the vp tree is constructed directly on G, i.e. when we apply the aforementioned construction scheme and navigation rules [cf. (10.11)] to G even though it is non-metric. First of all, notice that when $r = 0$, both schemes work perfectly since $G = M$ corresponds to a metric space. For other values of r, the vp tree constructed on M (blue lines) consistently outperforms the one constructed on G (red lines). E.g., for $r = 0.6$ the median and mean relative positions of the nodes found on M are in the top 0.5% and 0.8%, respectively, which contrast with the ones found on G which are in the top 1.7% and 3.2%, respectively. In terms of the rate of perfect recovery, the difference between both approaches is less conspicuous but there is still value in projecting G onto M. E.g., for $r = 0.3$ the rate of perfect recovery without projecting is 28% whereas with the projection it increases to 33%. Furthermore, notice that for large values of r when most of the edges in G are perturbed, the structure becomes more similar to a metric space and, thus, there is an improvement in the search performance both on G and M.

When $q > 1$, q-metric spaces contain a more stringent structure than regular metric spaces and this additional structure can be used to speed up search, as we show next.

Proposition 33 *When using a vp tree for nearest neighbor search in a q-metric space, the following rules ensure optimality of the node found [cf.* (10.11)*]*

$$\begin{cases} \text{a) } W(z,v)^q \leq \mu_v^q - \tau^q & \Rightarrow \text{ visit only the left child,} \\ \text{b) } \mu_v^q - \tau^q < W(z,v)^q \leq \mu_v^q + \tau^q & \Rightarrow \text{ visit left \& right child,} \\ \text{c) } \mu_v^q + \tau^q < W(z,v)^q & \Rightarrow \text{ visit only right child,} \end{cases} \tag{10.12}$$

where z is the query point, v is the vantage point of the current node of the vp tree, μ_v is the median distance between v and the other points in the current node of the vp tree, and τ is the distance from z to the current best NN estimate.

Proof: We need to show that by following the rules in (10.12) we are not discarding any node that could be the NN of our query point z. Assume that a) is true, then the q-triangle inequality implies that, for all $t \in V'$,

$$W(v,t)^q \leq W(v,z)^q + W(z,t)^q \leq \mu_v^q - \tau^q + W(z,t)^q, \tag{10.13}$$

where the second inequality follows from a). Furthermore, for every t belonging to the right child of the current node of the vp tree we have (by construction) that $W(v,t) > \mu_v$. Combining this fact with (10.13) it follows that $W(z,t) > \tau$ for every t in the right child. This implies that every node t in the right child is at a distance from z larger than the current best estimate τ and, thus, can be discarded. Similarly, if we assume that c) is true, we may leverage the q-triangle inequality to write that, for all $t \in V'$,

$$\mu_v^q + \tau^q < W(z,v)^q \leq W(z,t)^q + W(t,v)^q. \tag{10.14}$$

If we combine (10.14) with the fact that for every t in the left child $W(t,v) \leq \mu_v$, it follows that $W(z,t) > \tau$ for every t in the left child, concluding the proof. ■

Notice that when $q = 1$, the conditions in (10.12) boil down to those in (10.11), as expected. Further, the range encompassed by (10.12)-b is smaller than that in (10.11)-b, and decreases in size with larger q. To see this, compute the q-th power of (10.11)-b and, from the fact that μ_v and τ are positive, it follows that $\mu_v^q + \tau^q \leq (\mu_v + \tau)^q$ showing that the upper bound in (10.12)-b is tighter. Similarly, whenever $0 \leq \tau \leq \mu_v$ – otherwise, both lower bounds are effectively zero –, we have that $(\mu_v - \tau)^q \leq \mu_v^q - \tau^q$, showing that the lower bound in (10.12)-b is tighter as well compared to the one in (10.11)-b. This implies that in more structured spaces – larger q – we are more likely to discard parts of the vp tree – satisfying conditions a) or c) – speeding up the search. Based on this observation, one can project a metric space onto q-metric spaces with $q > 1$ in order to increase search speed with the cost of decreasing search performance due to the deformation introduced when projecting the original metric space. Figs. 10.5(b) and 10.5(c) illustrate this tradeoff. We generate 20 metric spaces with 1,000 points each and perform 100 queries in each metric space. In Fig. 10.5(b) we illustrate (via a box plot) the distribution of the 20 averages of the number of comparisons needed (as a percentage of 1000) to perform the 100 searches when first projecting the metric space applying $\mathcal{P}_q^*$ for varying q. In Fig. 10.5(c) we plot the distribution of the relative position of the NN found as a function of q. Notice that when $q = 1$, the node found is always the correct one (0%) since the original space was chosen to be metric. Furthermore, when, e.g., $q = 3$ the number of measurements needed to search the vp tree is reduced to around half of those needed for metric spaces (cf. Fig. 10.5(b)) while the neighbors found are within the top 2% candidates among the 1,000 points (cf. Fig. 10.5(c)). For large values of q, the reduction in computation is noticeable – around 8 times for $q = 30$ – but the detriment in performance is also large. Depending on the application, the value of q can be tuned to find the correct equilibrium between computational efficiency and search performance.

10.3.2 Search in quasi-metric spaces

We begin by specifying the construction of a *vp quasi-metric tree*, a natural extension of the vp trees introduced in Section 10.3.1 for the search in quasi-metric spaces. To construct a vp quasi-metric tree given G', we begin (as in regular vp trees) by associating the whole node set V' to the root of the tree and pick a vantage point at random, say $v \in V'$. We compute the median μ_1 of the distances $W(v, x)$ *from* the vantage point to every other node $x \in V'$ and the median μ_2 of the distances $W(x, v)$ *to* the vantage point. We then determine a three-block covering of V' where the first block contains the points whose distance *from* v is smaller than or equal to μ_1, the second block contains the points whose distance *to* v is larger than μ_2, and the last block contains all the nodes in V' that are not included in the previous two blocks. The nodes in the first block are assigned to the left child of the root of the tree, the right child consists of the nodes in the second block, and the middle child contains the points in the third block. We iteratively repeat this procedure within each of the children until every leaf in the tree is associated to a single point in V'; see Fig. 10.6.

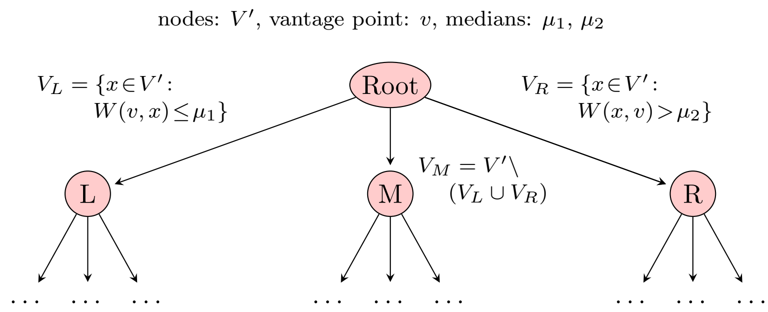


Figure 10.6: Vantage point quasi-metric tree. Unlike vp metric trees, quasi-metric trees are not binary and their hierarchical structure depends on both the distances to and from the chosen vantage point.

Notice that unlike the vp metric tree, the introduced quasi-metric tree is not binary but rather each non-leaf node can have up to three children nodes. Nevertheless, whenever the space of interest is symmetric, we have that $\mu_1 = \mu_2$ and it follows that the third block in the aforementioned partition is empty. Thus, the tree becomes binary and boils down to the vp metric tree in Section 10.3.1.

To efficiently search a vp quasi-metric tree for the NN of a query point z, we leverage the directed triangle inequality to discard branches of the tree without even traversing them, reducing the number of measurements needed. More specifically, assume that we are searching at an intermediate node in the tree, say node R in Fig. 10.6 and the current best estimate of the NN is at distance τ from z. We then compute the distances $W(z, v_R)$ and $W(v_R, z)$ between z and the vantage point v_R associated to the current node in the tree. If $W(z, v_R) < \tau$, we update our estimate of τ. In order to continue traversing the tree, we follow the ensuing rules where v is the current vantage point and the medians μ_1 and μ_2 are computed based on the subset of V' associated with the current node in the tree

$$\begin{cases} \text{a) } W(v,z)^q \le \mu_1^q - \tau^q & \Rightarrow \text{ visit only the left child,} \\ \text{b) } W(z,v)^q > \mu_2^q + \tau^q & \Rightarrow \text{ visit only right child,} \\ \text{c) otherwise} & \Rightarrow \text{ visit all three children.} \end{cases} \tag{10.15}$$

As was the case for metric trees, even though statements a) and b) entail that we discard part of the nodes in V', we can guarantee that the NN of z is not contained among the discarded nodes, as we show next.

Proposition 34 *When using a vantage point quasi-metric tree for nearest-neighbor search in a quasi-q-metric space, the rules in* (10.15) *ensure optimality of the neighbor found.*

Proof: We need to show that by following the rules in (10.15) we are not discarding any node that could be the NN of our query point z. Assume that a) is true, then the directed q-triangle inequality implies that, for all $t \in V'$,

$$W(v,t)^q \leq W(v,z)^q + W(z,t)^q \leq \mu_1^q - \tau^q + W(z,t)^q. \tag{10.16}$$

Furthermore, for every t not belonging to the left child of the current node of the tree, we have (by construction) that $W(v,t) > \mu_1$. Combining this fact with (10.16) it follows that $W(z,t) > \tau$. This implies that every node t not belonging to the left child is at a distance from z larger than the current best estimate τ and, thus, can be discarded. Similarly, if we assume that b) is true, we may leverage the directed q-triangle inequality to write that, for all $t \in V'$,

$$\mu_2^q + \tau^q < W(z,v)^q \leq W(z,t)^q + W(t,v)^q. \tag{10.17}$$

If we combine (10.17) with the fact that for every t not in the right child of the current node we have that $W(t,v) \leq \mu_2$, it then follows that $W(z,t) > \tau$, concluding the proof. ■

Notice that the construction of the tree requires a one-time computational effort and can then be utilized to speed up multiple queries in the same dataset. Moreover, the computational gains during search increase with q. To see this, notice that if either (10.15)-a) or (10.15)-b) are satisfied for a given q then they must be satisfied for all q' where $q' > q$. As for metric trees, this implies that in more structured spaces – larger q – we are more likely to discard parts of the tree speeding up the search. Based on this observation, one can project a quasi-metric space onto quasi-q-metric spaces with $q > 1$ in order to increase search speed with the cost of decreasing search performance.

A possible way to search a non-metric network G is to first project it onto either a q-metric space M or a quasi-q-metric space $\tilde{M}$ and then construct a vp metric tree on M or a vp quasi-metric tree on $\tilde{M}$; see Fig. 10.7.

We consider nearest-neighbor searches under two different settings. In Setting 1, we generate 100 quasi-metric spaces with 1,000 points each and perform 100 queries in each space. In Setting 2, by contrast, the original networks are not quasi-metric spaces. The asymmetric dissimilarities in the second setting are generated by first computing the (symmetric) distances between 1,000 points randomly plotted in a two-dimensional square and then multiplying each of these distances (in each direction) by a random number uniformly chosen between 0.5 and 1.5. The quasi-metric spaces of the first setting are generated by applying $\tilde{\mathcal{P}}_1$ [cf. (10.2)] to the asymmetric networks of the second setting.

In Fig. 10.7(a) we illustrate the average difference in ranking between the nearest neighbor found and the true nearest neighbor. E.g., a value of 2 indicates that among the 1,000 nodes there are on average two better options than the node found. We illustrate this quantity for three search strategies: i) applying $\tilde{\mathcal{P}}_q$ to project the original space onto a quasi-q-metric space and then performing a quasi-metric tree search (blue); ii) applying the ProSym $\underline{\mathcal{P}}_q$ method to project the original space onto a q-metric space and then performing a metric tree search (green); and iii) the counterpart of the previous strategy but based on the SymPro $\overline{\mathcal{P}}_q$ method (yellow).

For Setting 1 and $q = 1$, the application of $\tilde{\mathcal{P}}_1$ in our first search strategy does not modify the space and, thus, we are guaranteed to always find the nearest neighbor (cf. Proposition 34). Even for the symmetric projections $\overline{\mathcal{P}}_1$ and $\underline{\mathcal{P}}_2$ the difference with the real nearest neighbor averaged over all realizations is negligible. When we increase q the search performance slightly decreases, with the quasi-metric projection outperforming the two metric projections analyzed. However, the points found are still in the top 0.2% (2 out of 1,000) closest points to the query. In Setting 2, where the original spaces are not quasi-metric, we see first a decay in the search performance – although still returning results in the top 1% – and it can be observed that the quasi-metric projection is no longer the uniformly best strategy. As it turns out, both $\overline{\mathcal{P}}_q$ or $\tilde{\mathcal{P}}_q$ return the best results depending on the value of q. The decrease in performance for larger values of q is associated with faster searches, where the number of comparisons needed to traverse the (quasi-)metric trees is reduced; see Fig. 10.7(b). Considering that the search spaces contain 1,000 points, the amount of queried distances is greatly reduced compared to a brute force search. For example, for the first setting and $q = 1$, the number of queries is reduced to around 12% of those needed for a brute force search, for $q = 3$ this number is reduced below 5%, and for $q = 10$ this number is further reduced below 3% for all three projections studied. A similar behavior can be observed for Setting 2. For the case of the search in quasi-metric trees, since two nodes in the tree can have associated the same vantage point, we can query the distance to a vantage point more than once when traversing the tree. This is depicted by the light blue portion of the plotted bars. Fig. 10.7 shows that q can be used to tune the tradeoff between search accuracy and speed, where low values of q yield better accuracy while large values of q are associated to faster but less accurate searches.

10.4 Visualization of asymmetric data

Visualization methods facilitate the understanding of high-dimensional data by projecting it into familiar low-dimensional domains such as $\mathbb{R}^2$. A customary way of performing such projection is via multidimensional scaling (MDS) [23], where the data is embedded in Euclidean space while minimizing a measure of distortion with respect to the original dissimilarities between the data points. Even though the classical MDS formulation only admits symmetric dissimilarities as input, extensions for the asymmetric case have been proposed [22, 36, 94]. Most existing approaches either symmetrize the data first or decompose the asymmetric dissimilarities into symmetric and skew-symmetric parts and then focus on

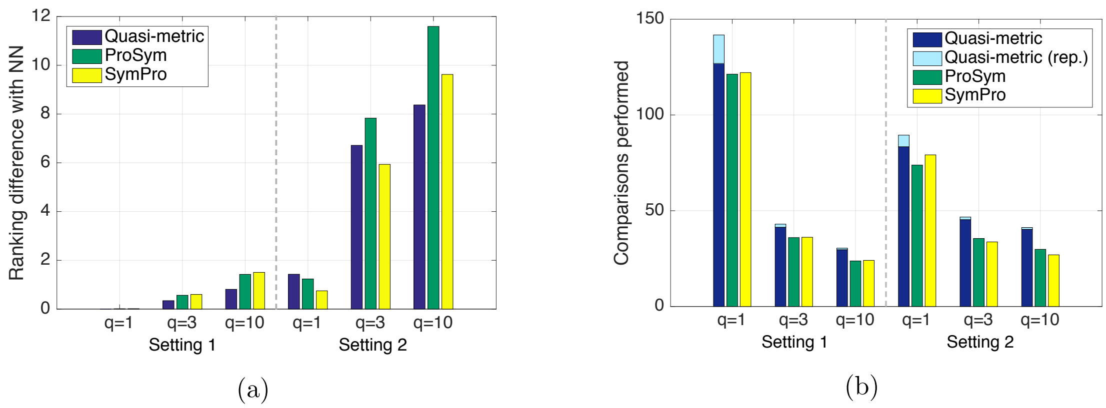
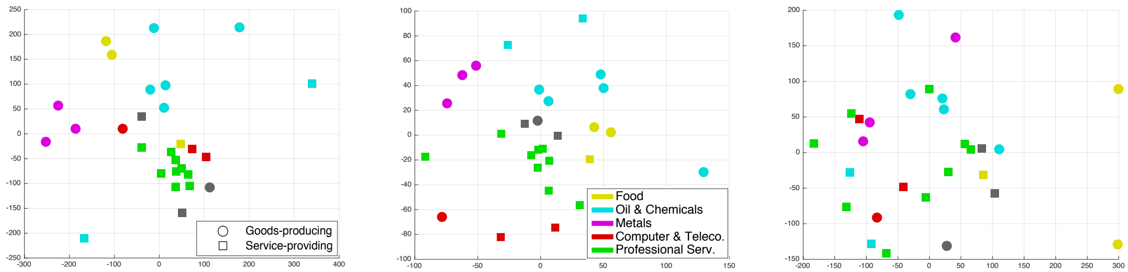


Figure 10.7: (a) Ranking difference between the approximate nearest neighbor found and the real nearest neighbor when the original space is projected to a (quasi-)q-metric space for different values of q under two settings. Results are averaged over 10,000 queries performed in 100 graphs (100 queries per graph) containing 1,000 nodes. (b) Number of comparisons required to find the approximate nearest neighbor.

the former for the low-dimensional representation while incorporating the information of the latter in various ways, see e.g., [94]. By contrast, we use the framework here developed to project the originally asymmetric dissimilarities into (symmetric) metric spaces to which we can then apply MDS. We illustrate this approach in a data set representing the interactions between industrial sectors in the economy of the United States (U.S.).

As already mentioned, the U.S. Department of Commerce publishes a yearly table of inputs and outputs organized by economic sectors; see Section 6.2 for details. For the current experiment, we focus on the set I of 28 industrial sectors with the largest production. Based on the similarity function $U(i,i')$ representing how much of the production of sector i is used as input of sector i', we define the network $G_I = (I, E_I, W_I)$ where the dissimilarity function W_I satisfies $W_I(i,i) = 0$ for all $i \in I$ and, for $i \neq i' \in I$, is given by $W_I(i,i') = 1/U(i,i')$. Notice that a small dissimilarity from sector i to sector i' implies that sector i' highly relies on the output of sector i as input for its own production.

We apply non-metric MDS with Kruskal's stress [50] to three symmetrized versions of G_I: i) $\overline{G}_I = \overline{\mathcal{P}}(G_I)$ obtained from SymPro with the symmetrizing function $s(a,b) = (a+b)/2$ [cf. (9.5)], ii) $\underline{G}_I = \underline{\mathcal{P}}(G_I)$ obtained from ProSym with the same s [cf. (9.11)], and iii) $\mathcal{S}(G)$ obtained by directly symmetrizing the weights W_I with no prior or posterior projection. In Fig. 10.8 we present the MDS representations for the three cases analyzed. In the scatter plots, each point represents an industrial sector while the points' shape and color indicate features of the sectors. More specifically, circular and squared markers refer to goods-producing and service-providing industries, respectively. Moreover, the marker color indicates activities related to a specific subject, e.g., yellow for food or magenta for metals; see legend in Fig. 10.8(b).



(a) SymPro (b) ProSym (c) Symmetrization

Figure 10.8: Multidimensional scaling representation of the input-output relation among the most productive industrial sectors of the U.S. during 2011 after performing (a) a SymPro projection, (b) a ProSym projection, and (c) only a symmetrization. Point shapes indicate the segment (goods vs. services) while point colors specify broad activity types.

The visualizations associated with SymPro and ProSym reveal different aspects of the asymmetric network G_I. Fig. 10.8(a) shows that $\overline{\mathcal{P}}$ produces an almost perfect linear separation between service-providing (squares) and good-producing (circles) industries. By contrast, Fig. 10.8(b) illustrates that $\underline{\mathcal{P}}$ yields a grouping of the sectors by subject (colors). For example, the three yellow sectors related to the food industry – 'Farms', 'Food products', and 'Food services' – are depicted close to each other in Fig. 10.8(b) whereas in Fig. 10.8(a) the first two (good-producing) sectors are plotted close to each other but far away from the last (service-providing) sector. To see why this is true notice that SymPro, by first symmetrizing the dataset, promotes small distances between sectors that are closely related by bidirectional paths. These bidirectional relations tend to occur within the same vertical segment of the economy (extraction of raw material, manufacturing, and services) and are more common among service sectors, thus, the high concentration of squares in the central portion of Fig. 10.8(a). ProSym, in contrast, by first computing unidirectional shortest paths, promotes closeness among sectors that are strongly related in at least one direction. This tends to occur across vertical segments of the economy, e.g., from 'Farms' to 'Food products' and from the latter to 'Food services', and thus a color-coded clustering is observed in Fig. 10.8(b). Lastly, Fig. 10.8(c) reveals that a simple symmetrization of the data prior to the application of MDS does not seem to generate a meaningful low-dimensional representation, since the points do not show a clear separation neither by shape nor by color. An intuitive explanation of why the projection onto a metric space prior to MDS yields a more meaningful representation is given in the following remark.

Remark 22 (Directed Isomap) Isomap is a well-established nonlinear dimensionality reduction scheme [81]. Given a set of data points, Isomap first generates a nearest-neighbor graph, then computes the shortest path between every pair of nodes, and finally uses this geodesic distances as inputs for MDS. The idea of the graph geodesic computation is to capture the distances along the manifold formed by the points and not in the original space of embedding. When given asymmetric dissimilarities, one could conceivably symmetrize them either before or after the computation of the shortest paths. Consequently, both the application of $\overline{\mathcal{P}}$ or $\underline{\mathcal{P}}$ followed by MDS are natural extensions of Isomap to asymmetric domains, thus explaining the satisfactory low-dimensional representations in Figs. 10.8(a) and 10.8(b).

Chapter 11

Dioid metric spaces

A common thread throughout this thesis has been the projection of unstructured spaces – composed of a set of nodes and binary relations between them – onto their structured counterparts in which the binary relations are bound to satisfy some kind of triangle inequality. In the present chapter, we rephrase this projection problem in the domain of dioid spaces. Dioid spaces are natural generalizations of networks where the weights between nodes take values in an algebraic construction called dioid. In this way, by analyzing projections between spaces at a high level of algebraic abstraction we obtain powerful results that can then be particularized to a domain of interest by specifying the underlying dioid. In Section 11.1 we present the algebraic concepts needed to formally introduce the notion of a dioid. In Section 11.2 we further abstract networks as dioid spaces and formally define dioid metric projections, i.e., maps that induce a metric-like structure in dioid spaces. We then propose an axiomatic approach to study these projections – outlined in Section 11.3 – leading to a uniqueness result. Finally, Section 11.4 analyzes how the results obtained at the level of dioid spaces particularize to more familiar and concrete domains.

11.1 The algebra of dioids

A *dioid* is a set endowed with two operations that we can interpret as addition and multiplication. Nevertheless, in order to formally define a dioid, we begin by introducing the simpler notion of a *monoid*.

Definition 7 *A* monoid $(E, \oplus)$ *is a set E endowed with an operation $\oplus$ that satisfies the following properties for all $a, b, c \in E$:*

(i) Associativity: $(a \oplus b) \oplus c = a \oplus (b \oplus c)$.

(ii) Neutral element: There is a neutral element $\epsilon \in E$ such that $a \oplus \epsilon = \epsilon \oplus a = a$.

Whenever a monoid satisfies the property that for all $a, b \in E$ it holds $a \oplus b = b \oplus a$, we say that the monoid is commutative. Two other properties that are important in our framework are the notions of idempotent and selective monoids. A monoid is said to be *idempotent* if the operation $a \oplus b$ satisfies $a \oplus a = a$, for all $a \in E$ whereas a monoid is said to be *selective* if the outcome of the operation $a \oplus b$ satisfies $a \oplus b \in \{a, b\}$, for all $a, b \in E$. This means that a monoid is selective if the operation $a \oplus b$ selects one of the two elements a or b. Observe that if a monoid is selective it must be idempotent.

Also central to the definition of a dioid are the notions of preorder, order, and total order which we formally state next.

Definition 8 *Given a set E and a binary relation $\prec$, consider the following properties for all $a, b, c \in E$:*

(i) Reflexivity: $a \prec a$.

(ii) Transitivity: If $a \prec b$ and $b \prec c$ then $a \prec c$.

(iii) Antisymmetry: If $a \prec b$ and $b \prec a$ then $a = b$.

(iv) Completeness: We either have $a \prec b$ or $b \prec a$ or both.

If $\prec$ satisfies (i)-(iv) it is said to be a total order, if it satisfies (i)-(iii) it is an order, and if it satisfies (i)-(ii) it is a preorder.

A total order is a relationship in which all elements of E can be compared to each other and therefore completely ordered. An order is a relationship in which some elements are comparable but some other elements need not be. There may be elements $a, b \in E$ for which neither $a \prec b$ nor $b \prec a$ are true. A preorder is further relaxed to allow different elements to satisfy $a \prec b$ and $b \prec a$ simultaneously. Elements for which this is true can be thought of as equivalent with respect to the preorder relationship.

If the operation $\oplus$ in the monoid $(E, \oplus)$ is commutative it induces at least a preorder in the set E as we formally state in the following proposition.

Proposition 35 (Ch.1 Sec.3.3 in [32]) *Consider a commutative monoid $(E, \oplus)$ and define the relationship $\preceq$ as*

$$a \preceq b \iff \exists\, c \in E : \quad a \oplus c = b. \tag{11.1}$$

The relationship $\preceq$ is a preorder that we call the canonical preorder of the monoid $(E, \oplus)$.

In general, the canonical preorder need not be an order nor a total order. Whenever the relationship $\preceq$ in (11.1) is an order, we say that the monoid is *canonically ordered*. In particular, this can be shown to be true when the operation $\oplus$ is idempotent [32]. Furthermore, whenever $\oplus$ is selective, the relationship $\preceq$ becomes a total order. The concepts of monoids and orders presented are used to introduce the notion of a dioid in the following formal definition.

Definition 9 *A* dioid $(E, \oplus, \otimes)$ *is a set E endowed with two operations $\oplus$ and $\otimes$ satisfying the following properties:*

(i) Addition monoid: The pair $(E, \oplus)$ is a commutative monoid with neutral element ϵ.

(ii) Multiplication monoid: The pair $(E, \otimes)$ is a monoid with neutral element e.

(iii) Order: The addition monoid $(E, \oplus)$ is canonically ordered.

(iv) Absorption: The element ϵ is absorbing for $\otimes$, i.e. $a \otimes \epsilon = \epsilon \otimes a = \epsilon$ for all $a \in E$.

(v) Distributive property: The multiplication $\otimes$ is right and left distributive with respect to the addition $\oplus$. I.e., for all $a, b, c \in E$ we have

$$\begin{aligned} a \otimes (b \oplus c) &= (a \otimes b) \oplus (a \otimes c), \\ (b \oplus c) \otimes a &= (b \otimes a) \oplus (c \otimes a). \end{aligned} \tag{11.2}$$

As per Definition 9, a dioid contains two operations $\oplus$ and $\otimes$ that are intended to generalize the notions of regular addition and multiplication. A difference with these, however, is that the multiplication $\otimes$ is not required to be commutative. One important feature that distinguishes the dioid from the more familiar notion of ring is that the additive inverse need not exist for the former but must exist for the latter. Putting it differently, the addition induces a group structure in rings whereas in dioids it induces a canonical order. This order is essential for the framework here developed since it allows us to write inequalities and, in particular, it enables the generalization of the triangle inequality to the realm of dioid spaces, as illustrated in the next section.

11.2 Dioid spaces and the triangle inequality

A dioid $(E, \oplus, \otimes)$ is said to be *complete* if every subset $A \subseteq E$ has a supremum with respect to the canonical order induced by $\oplus$ and the distributive property (11.2) is also satisfied for infinite sums. In a complete dioid, we define the top element T as the sum of all the elements in the dioid,

$$T = \bigoplus_{a \in E} a. \tag{11.3}$$

Notice that for all $a \in E$,

$$T \oplus a = T, \tag{11.4}$$

To see this, by (11.1) we have that $T \oplus a \succeq T$. Conversely, since $T \oplus a \in E$, by the definition of the top element in (11.3), it must be that $T \oplus a \preceq T$, and the result in (11.4) follows from the antisymmetry of the order $\preceq$.

A *dioid space* consists of a finite set X and a function f mapping pairs of elements in X to elements of a given dioid.

Definition 10 *Given a complete dioid $\mathfrak{A} = (E, \oplus, \otimes)$ with top element T, a finite* dioid space *$Q = (X, f)$ consists of a finite set X and a function $f : X \times X \to E$ that satisfies for all $x, x' \in X$:*

(i) Identity: $f(x, x') = T \iff x = x'$,

(ii) Symmetry: $f(x, x') = f(x', x)$.

We say that the dioid $\mathfrak{A}$ underlies the dioid space Q. A space that satisfies (ii) and the right-to-left implication in (i) is termed a dioid semi-*space.*

A dioid space generalizes the concept of a symmetric network. It consists of a finite set of nodes X and 'weighted' edges between pairs of nodes, where the weights are given by elements of E in the underlying dioid $\mathfrak{A}$. Even though the node set X is required to be finite, the set E can be infinite, e.g., it can consist of the non-negative reals as in regular networks.

A dioid *metric* space is a dioid space $Q = (X, f)$ where the function f satisfies an additional condition for all $x, x', x'' \in X$

$$f(x, x') \succeq f(x, x'') \otimes f(x'', x'). \tag{11.5}$$

We refer to condition (11.5) as a *dioid triangle inequality.* Analogously, a dioid semi-space that satisfies (11.5) is a dioid semi-metric space.

To see why (11.5) is a reasonable generalization of the familiar triangle inequality to the domain of dioid spaces, consider the set $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \infty$ and the dioid $\mathfrak{A} = (\bar{\mathbb{R}}_+, \min, +)$. Notice that the top element in $\mathfrak{A}$ is $T = \min(\bar{\mathbb{R}}_+) = 0$ so that the definition of dioid space (Definition 10) coincides with that of a weighted network with no self-loops. Moreover, using (11.1) we may rewrite (11.5) and specialize it to dioid $\mathfrak{A}$ to obtain

$$\exists\, c \in \bar{\mathbb{R}}_+ : \quad \min(c, f(x, x'') + f(x'', x')) = f(x, x'). \tag{11.6}$$

It immediately follows that (11.6) can be expressed in a more suitable form as $f(x, x') \leq f(x, x'') + f(x'', x')$, which is the regular triangle inequality. Thus, the dioid triangle inequality in (11.5) recovers the regular triangle inequality when the underlying dioid is $(\bar{\mathbb{R}}_+, \min, +)$. In Section 11.4 we see that other familiar inequalities, such as the strong triangle inequality [cf. (2.12)], can be recovered by specializing (11.5) to a suitable dioid.

Given a complete dioid $\mathfrak{A} = (E, \oplus, \otimes)$, denote by $\mathcal{Q}^{\mathfrak{A}}$ the set of all dioid spaces $Q = (X, f)$ underlain by $\mathfrak{A}$, i.e. where f takes values in E. Similarly, denote by $\mathcal{Q}^{\mathfrak{A}}_{\mathrm{m}} \subset \mathcal{Q}^{\mathfrak{A}}$ the subset containing all dioid metric spaces. We define a dioid metric projection

$$\mathcal{P} : \mathcal{Q}^{\mathfrak{A}} \to \mathcal{Q}^{\mathfrak{A}}_{\mathrm{m}}, \tag{11.7}$$

as a map that assigns to every dioid space (X, f) a dioid metric space $\mathcal{P}(X, f) = (X, f_m)$ defined on the same set X. As was the main theme throughout the thesis, our objective is to establish an axiomatic framework to study how to project dioid spaces into their more structured counterpart of dioid metric spaces.

11.3 Canonical projection for dioid spaces

We begin by recasting the Axiom of Projection introduced in Section 8.2 for a generic dioid metric projection.

> *(AA1) Axiom of Projection.* Given a dioid $\mathfrak{A}$, every dioid metric space $Q \in \mathcal{Q}^{\mathfrak{A}}_{\mathrm{m}}$ is a fixed point of the projection $\mathcal{P}$, i.e. $\mathcal{P}(Q) = Q$.

As was the case for regular metric spaces, the axiom (AA1) encodes the reasonable requirement that if the input space is already a dioid metric space then it must remain unaltered by the projection.

From the identity property of dioid spaces (cf. Definition 10), it follows that the function f achieves its top value T when evaluated in the pair (x, x) for all $x \in X$. In this way, we can interpret f as evaluating the similarity between the input pair of nodes. Consequently, if another function g is such that $g(x, x') \succeq f(x, x')$ for all $x, x' \in X$, then the nodes in the space (X, g) are more similar than the nodes in (X, f). Based on this, we define the notion of similarity increasing maps – the natural counterpart of dissimilarity reducing maps – as follows: given two dioid spaces (X, f) and (Y, g), the *injective* map $\phi : X \to Y$ is termed a *similarity increasing* map if it holds that $f(x, x') \preceq g(\phi(x), \phi(x'))$ for all $x, x' \in X$. With this concept in place, we extend the Axiom of Injective Transformation introduced in Section 8.2 to the dioid domain.

> *(AA2) Axiom of Injective Transformation.* Consider any two dioid spaces $Q = (X, f)$ and $Q' = (Y, g)$ and any (injective) similarity increasing map $\phi : X \to Y$. Then, for all $x, x' \in X$, the output dioid metric spaces $(X, f_{\mathrm{m}}) = \mathcal{P}(Q)$ and $(Y, g_{\mathrm{m}}) = \mathcal{P}(Q')$ satisfy
>
> $$f_{\mathrm{m}}(x, x') \preceq g_{\mathrm{m}}(\phi(x), \phi(x')). \tag{11.8}$$

The axiom (AA2) states that if we increase the similarities between every pair of nodes in X then the nodes in the corresponding dioid metric space must be more similar to each other. As done throughout the thesis, we say that a dioid metric projection $\mathcal{P}$ is *admissible* if it satisfies axioms (AA1)-(AA2).

We define the canonical metric projection $\mathcal{P}^*$ on dioid spaces as $\mathcal{P}^*(X, f) = (X, f^*)$ where, for all $x, x' \in X$, we have that

$$f^*(x, x') := \bigoplus_{P_{xx'}} \bigotimes_{i | x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.9}$$

In (11.9), the value of f^* between two nodes is obtained by summing over all paths linking these nodes the product of the values of f encountered when traversing these paths in order. The map $\mathcal{P}^*$ is in fact a metric inducing map under some assumptions of the underlying dioid. Moreover, $\mathcal{P}^*$ is the *only* admissible dioid metric inducing map as the following theorem shows. In the statement of the theorem and from now on, we say that a dioid $(E, \oplus, \otimes)$ is selective (idempotent) if the monoid $(E, \oplus)$ is selective (idempotent).

Theorem 20 *Given a complete and selective dioid $\mathfrak{A} = (E, \oplus, \otimes)$ where the monoid $(E, \otimes)$ is commutative and the top element is equal to the $\otimes$ neutral element, $T = e$, the canonical projection $\mathcal{P}^*$ defined in* (11.9) *is the only admissible dioid metric projection in the set $\mathcal{Q}^{\mathfrak{A}}$.*

The following two lemmas are instrumental in showing the above result.

Lemma 4 *Under the conditions in Theorem 20, the monoid $(E, \otimes)$ is positive, i.e., if $a \otimes b = e$ for $a, b \in E$ then it must be that $a = b = e$.*

Proof: First notice that if $a = e$ then it must be that $b = e$ since $e = a \otimes b = e \otimes b = b$ where we used the multiplicative neutrality of e in the last equality. We now use the distributive property of dioids (cf. Definition 9) to write

$$(a \otimes b) \oplus (a \otimes T) = a \otimes (b \oplus T) = a \otimes T = a, \tag{11.10}$$

where we used (11.4) in the next-to-last equality, and the fact that $T = e$ in the last one. Moreover, this latter fact can be used to write

$$e = e \oplus a = (a \otimes b) \oplus (a \otimes e) = (a \otimes b) \oplus (a \otimes T). \tag{11.11}$$

Finally, by combining (11.10) and (11.11) the statement of the lemma follows. ■

Lemma 5 *Given elements $a, b, a', b' \in E$ such that $a \succeq a'$ and $b \succeq b'$ then $a \oplus b \succeq a' \oplus b'$ and $a \otimes b \succeq a' \otimes b'$.*

Proof: We first show the inequality for the operation $\oplus$. From (11.1) we know that $a \succeq a'$ implies the existence of an element $a'' \in E$ such that $a = a' \oplus a''$, and similarly regarding the existence of an element b'' such that $b = b' \oplus b''$. Hence,

$$a \oplus b = (a' \oplus a'') \oplus (b' \oplus b'') = (a' \oplus b') \oplus (a'' \oplus b'') \succeq a' \oplus b', \tag{11.12}$$

where we used the commutativity of $\oplus$ for the second equality. Similarly, regarding $\otimes$ we can claim that

$$a \otimes b = (a' \oplus a'') \otimes (b' \oplus b'') = (a' \otimes b') \oplus (a' \otimes b'') \oplus (a'' \otimes b') \oplus (a'' \otimes b'') \succeq a' \oplus b', \tag{11.13}$$

where the distributive property of $\otimes$ with respect to $\oplus$ was used for the second equality. ■

With these results in place, we now delve into the proof of Theorem 20.

Proof of Theorem 20: The proof of this theorem is divided into three claims showing the following intermediate results: i) $\mathcal{P}^*$ is a valid dioid metric projection, ii) $\mathcal{P}^*$ satisfies axioms (AA1)-(AA2), and iii) any other admissible metric projection coincides with $\mathcal{P}^*$. The joint consideration of these three claims implies the statement of the theorem.

Claim 7 *The output f^* of the canonical projection $\mathcal{P}^*$ defined in* (11.9) *is a valid dioid metric.*

Proof: We need to show that f^* satisfies the identity and symmetry properties in Definition 10 as well as the dioid triangle inequality in (11.5). We begin by showing the identity property. In computing $f^*(x, x')$ for $x = x'$ using (11.9), we may divide the summation into the particular path $[x, x]$ and the rest of the paths starting and finishing at x to obtain

$$f^*(x, x) = f(x, x) \ \oplus \bigoplus_{P_{xx} \setminus [x,x]} \ \bigotimes_{i | x_i \in P_{xx}} f(x_i, x_{i+1}) = T, \tag{11.14}$$

where the last equality follows from combining (11.4) and the fact that $f(x, x) = T$. For the opposite implication of the identity property, consider nodes $x \neq x'$ and focus on the definition of f^* in (11.9). From Lemma 4 and the fact that f satisfies the identity property, it follows that $\bigotimes_{i|x_i \in P_{xx}} f(x_i, x_{i+1}) \neq T$ for all paths $P_{xx'}$. Furthermore, since $\oplus$ is assumed to be selective and every term in the sum $\bigoplus_{P_{xx'}}$ is different from T, we conclude that $f^*(x, x') \neq T$ whenever $x \neq x'$. The symmetry of f^* follows immediately from the symmetry of f and the commutativity of the operation $\otimes$. Lastly, we show fulfillment of the dioid triangle inequality. Splitting the summation in the definition of $f^*(x, x')$ between the paths that contain x'' and those that do not, we obtain that

$$f^*(x,x') = \bigoplus_{P_{xx'}|x''\in P_{xx'}} \bigotimes_{i|x_i\in P_{xx'}} f(x_i,x_{i+1}) \ \oplus \bigoplus_{P_{xx'}|x''\notin P_{xx'}} \bigotimes_{i|x_i\in P_{xx'}} f(x_i,x_{i+1}). \tag{11.15}$$

Using the distributive property of dioids we may rewrite the first summand in (11.15) as

$$\begin{aligned}\bigoplus_{P_{xx'}|x''\in P_{xx'}} \bigotimes_{i|x_i\in P_{xx'}} f(x_i,x_{i+1}) &= \Big(\bigoplus_{P_{xx''}} \bigotimes_{i|x_i\in P_{xx''}} f(x_i,x_{i+1})\Big) \otimes \Big(\bigoplus_{P_{x''x'}} \bigotimes_{i|x_i\in P_{x''x'}} f(x_i,x_{i+1})\Big)\\ &= f^*(x,x'')\otimes f^*(x'',x'). \end{aligned}\tag{11.16}$$

Finally, upon substitution of (11.16) into (11.15), it follows that $f^*(x,x') \succeq f^*(x,x'')\otimes f^*(x'',x')$, as wanted. ■

Claim 8 *The canonical projection $\mathcal{P}^*$ defined in* (11.9) *satisfies axioms (AA1)-(AA2).*

Proof: We begin by showing that $\mathcal{P}^*$ abides by the Axiom of Projection (AA1). Denoting by $(X,f^*) = \mathcal{P}^*(X,f_{\mathrm{m}})$ the output of canonically projecting a dioid metric space (X,f_{m}), we have to show that $f^*(x,x') = f_{\mathrm{m}}(x,x')$ for all $x,x'\in X$. Splitting the summation in (11.9) between the direct path $[x,x']$ and the rest of the possible paths we obtain that

$$f^*(x,x') = f_{\mathrm{m}}(x,x') \ \oplus \bigoplus_{P_{xx'}\setminus[x,x']} \bigotimes_{i|x_i\in P_{xx'}} f(x_i,x_{i+1}), \tag{11.17}$$

from where it immediately follows that $f^*(x,x')\succeq f_{\mathrm{m}}(x,x')$. Furthermore, since f_{m} satisfies the triangle inequality, we have that $\bigotimes_{i|x_i\in P_{xx'}} f(x_i,x_{i+1}) \preceq f_{\mathrm{m}}(x,x')$ for all paths $P_{xx'}$. Hence, from (11.17) and Lemma 5 we obtain that

$$f^*(x,x') \preceq f_{\mathrm{m}}(x,x') \ \oplus \bigoplus_{P_{xx'}\setminus[x,x']} f_{\mathrm{m}}(x,x') = f_{\mathrm{m}}(x,x'), \tag{11.18}$$

where the last equality follows from selectivity – in fact idempotency is sufficient – of $\oplus$. By combining (11.18) with $f^*(x,x')\succeq f_{\mathrm{m}}(x,x')$, fulfillment of axiom (AA1) follows.

Regarding axiom (AA2), consider any two dioid spaces $Q=(X,f)$ and $Q'=(Y,g)$ and a similarity increasing map $\phi: X\to Y$, and denote by (X,f^*) and (Y,g^*) the canonical projections of Q and Q'. From (11.9) it then follows that

$$f^*(x,x') = \bigoplus_{P_{xx'}} \bigotimes_{i|x_i\in P_{xx'}} f(x_i,x_{i+1}) \preceq \bigoplus_{P_{\phi(x)\phi(x')}} \bigotimes_{i|\phi(x_i)\in P_{\phi(x)\phi(x')}} g(\phi(x_i),\phi(x_{i+1})) = g^*(\phi(x),\phi(x')), \tag{11.19}$$

where the inequality follows from Lemma 5 and the fact that ϕ is similarity increasing. ■

Claim 9 *If $\mathcal{P}$ is an arbitrary admissible dioid metric projection in $\mathcal{Q}^{\mathfrak{A}}$, then $\mathcal{P} \equiv \mathcal{P}^*$.*

Proof: Given an arbitrary dioid space $Q = (X, f)$, denote by $(X, f_{\mathrm{m}}) = \mathcal{P}(X, f)$ and by $(X, f^*) = \mathcal{P}^*(X, f)$ the output dioid metric spaces obtain after applying an arbitrary admissible projection and the canonical projection, respectively. We show that $\mathcal{P} \equiv \mathcal{P}^*$ by first showing that $f^*(x, x') \preceq f_{\mathrm{m}}(x, x')$ for all $x, x' \in X$ and then showing the inequality in the opposite sense.

Consider an arbitrary path $P_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$, a series of two-node dioid spaces $(\{p, q\}, f^i)$ such that $f^i(p, q) = f(x_i, x_{i+1})$ for all pairs of consecutive nodes in $P_{xx'}$, and a series of maps $\phi_i : \{p, q\} \to X$ such that $\phi_i(p) = x_i$ and $\phi_i(q) = x_{i+1}$. Notice that, by construction, the maps ϕ_i are similarity increasing and the two-node spaces $(\{p, q\}, f^i)$ are, in fact, dioid metric spaces. Consequently, from axiom (AA1) we have that the spaces $(\{p, q\}, f^i)$ are not altered by the admissible projection $\mathcal{P}$ and from axiom (AA2) we have that $f_{\mathrm{m}}(x_i, x_{i+1}) \succeq f(x_i, x_{i+1})$ for all i. Combining this inequality with the fact that f_{m} satisfies the dioid triangle inequality it follows that

$$f_{\mathrm{m}}(x, x') \succeq \bigotimes_{i | x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.20}$$

Since (11.20) is true for all paths $P_{xx'}$, we may sum over all such paths to obtain

$$\bigoplus_{P_{xx'}} f_{\mathrm{m}}(x, x') \succeq \bigoplus_{P_{xx'}} \bigotimes_{i | x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.21}$$

That $f_{\mathrm{m}}(x, x') \succeq f^*(x, x')$ follows by noting that the RHS of (11.21) is the definition of $f^*(x, x')$ [cf. (11.9)] and the LHS is equal to $f_{\mathrm{m}}(x, x')$ due to selectivity – in fact, it is enough with idempotency – of $\oplus$.

We are left to show that $f_{\mathrm{m}}(x, x') \preceq f^*(x, x')$. Consider the diagram of relations between dioid spaces in Fig. 11.1 where the top (blue) and left (red) maps illustrate the definitions of projections $\mathcal{P}^*$ and $\mathcal{P}$, respectively, and the right (green) map follows from the fact that (X, f^*) is a dioid metric space and that $\mathcal{P}$ satisfies the Axiom of Projection (AA1). Moreover, following a reasoning analogous to that used in obtaining (11.17) we have that $f^*(x, x') \succeq f(x, x')$ for all $x, x' \in X$ so that the identity map from (X, f) to (X, f^*) is a similarity increasing map. Hence, from a direct application of axiom (AA2), it follows that $f_{\mathrm{m}}(x, x') \preceq f^*(x, x')$. ■

Claim 7 shows that $\mathcal{P}^*$ is a valid metric inducing map whereas Claim 8 shows that it is admissible. Finally, in Claim 9 we show that a dioid metric projection which is admissible must be equivalent to $\mathcal{P}^*$, thus, completing the proof of the theorem. ■

Under the premise that axioms (AA1) and (AA2) are reasonable properties to require

$$
\begin{tikzcd}
(\mathbf{X}, \mathbf{f}) \arrow[r, "\mathcal{P}^*"] \arrow[r, bend right, "\phi = \mathrm{Id}"'] \arrow[d, "\mathcal{P}"'] & (\mathbf{X}, \mathbf{f}^*) \arrow[d, "\mathcal{P}"] \\
(\mathbf{X}, \mathbf{f}_{\mathrm{m}}) & (\mathbf{X}, \mathbf{f}^*)
\end{tikzcd}
$$

Figure 11.1: Diagram of maps between spaces for the proof of Theorem 20. Since $\mathcal{P}$ satisfies axiom (AA2), the existence of the similarity increasing map ϕ allows us to relate f_{m} and f^*.

from an admissible dioid metric projection, Theorem 20 shows that there is one, and only one, admissible projection and this is the canonical projection introduced in (11.9). Notice that a similar result was shown in Theorem 17 for (regular) metric projections. As a matter of fact, the result in Theorem 17 can be recovered from Theorem 20 by specializing this latter theorem to a particular dioid algebra as discussed in Section 11.4.

It is possible to obtain a result similar to Theorem 20 for the less restrictive case where the underlying dioid $\mathfrak{A}$ is idempotent instead of selective. In this case, the result can be framed in terms of dioid semi-spaces. Formally, we study dioid semi-metric projections $\hat{\mathcal{P}} : \hat{\mathcal{Q}}^{\mathfrak{A}} \to \hat{\mathcal{Q}}^{\mathfrak{A}}_m$ from the set $\hat{\mathcal{Q}}^{\mathfrak{A}}$ of dioid semi-spaces (cf. Definition 10) onto the structured subset $\hat{\mathcal{Q}}^{\mathfrak{A}}_m$ of dioid semi-metric spaces.

When the canonical projection $\mathcal{P}^*$ in (11.9) is applied to a dioid semi-space the output can be shown to be a dioid semi-metric space. Moreover, under some conditions of the underlying dioid, $\mathcal{P}^*$ is the only admissible projection as we state in the following proposition.

Proposition 36 *Relaxing the assumptions in Theorem 20 to admit an idempotent (not necessarily selective) underlying dioid $\mathfrak{A}$, the canonical projection $\mathcal{P}^*$ defined in* (11.9) *is the only admissible dioid semi-metric projection in the set $\hat{\mathcal{Q}}^{\mathfrak{A}}$.*

Proof: Selectivity was used to show Lemma 4 which, in turn, was used to show that $f^*(x, x') \neq T$ whenever $x \neq x'$. This argument is not valid when the underlying dioid is idempotent but not selective, hence, the dioid space (X, f^*) can be shown to be dioid semi-metric but not dioid metric. The remainder of the proof of Theorem 20 can be replicated solely depending on idempotency of $\mathfrak{A}$ in order to show the statement of the proposition. ■

The advantage of stating the unicity results in Theorem 20 and Proposition 36 at the level of algebraic abstraction of dioids is that, by specializing the dioid, we can both recover established unicity results as well as unveiling new results in other domains without the need of an additional proof in that specific domain.

11.4 Specializing the underlying dioid

We analyze the implications of Theorem 20 and Proposition 36 in dioid spaces associated with six different dioids. For the first dioid explored $\mathfrak{A}_{\min,+} = (\bar{\mathbb{R}}_+, \min, +)$, we formally show in Proposition 37 its validity as well as the fact that it satisfies the hypotheses in Theorem 20. For the rest of the dioids analyzed we omit a formal proof of these facts in order to avoid repetition.

11.4.1 $\mathfrak{A}_{\min,+} = (\bar{\mathbb{R}}_+, \min, +)$

Even though we have introduced $\mathfrak{A}_{\min,+} = (\bar{\mathbb{R}}_+, \min, +)$ with $\bar{\mathbb{R}}_+ = \mathbb{R}_+ \cup \infty$ in Section 11.2 to justify the dioid extension of the triangle inequality, we have not formally established that it is, indeed, a valid dioid. In the following proposition we show the validity of $\mathfrak{A}_{\min,+}$ as well as the fact that it satisfies the hypotheses in Theorem 20.

Proposition 37 *The dioid $\mathfrak{A}_{\min,+}$ is valid and satisfies the hypotheses in Theorem 20.*

Proof: We show the validity of the dioid by checking the properties in Definition 9: i) The monoid $(\bar{\mathbb{R}}_+, \min)$ is commutative with neutral element $\epsilon = \infty$, ii) the monoid $(\bar{\mathbb{R}}_+, +)$ has neutral element $e = 0$, iii) the monoid $(\bar{\mathbb{R}}_+, \min)$ is canonically ordered since min is an idempotent operation (cf. discussion following Proposition 35), iv) The element $\epsilon = \infty$ is absorbing for operation $+$, and v) the operation $+$ is distributive with respect to min, i.e., $a + \min(b, c) = \min(a + b, a + c)$ and analogously for right distributivity.

Regarding the hypotheses in Theorem 20, the dioid $(\bar{\mathbb{R}}_+, \min)$ is complete with top element $T = \min(\bar{\mathbb{R}}_+) = 0$ and the operation min is selective since $\min(a, b) \in \{a, b\}$. Moreover, $+$ is commutative and its neutral element $e = 0$ coincides with T. ■

Proposition 37 indicates that the uniqueness result in Theorem 20 is valid when the underlying dioid is $\mathfrak{A}_{\min,+}$. To fully grasp the implications of this result, let us first particularize other dioid concepts to the dioid of interest $\mathfrak{A}_{\min,+}$. A dioid space (cf. Definition 10) is composed of a set of nodes and a function between pairs of nodes (x, x') that takes values in $\bar{\mathbb{R}}_+$, is symmetric, and achieves zero if and only if $x = x'$. Notice that this coincides with our notion of (symmetric) weighted network. Moreover, as was stated in Section 11.2, the dioid triangle inequality for $(\bar{\mathbb{R}}_+, \min, +)$ boils down to the regular triangle inequality meaning that a dioid metric space in this case is just a regular metric space. Consequently, Theorem 20 can be reinterpreted as stating a uniqueness result for the projection of symmetric networks onto metric spaces and this unique projection $\mathcal{P}^*$ outputs a metric given by [cf. (11.9)]

$$f^*(x, x') = \min_{P_{xx'}} \sum_{i | x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.22}$$

Notice that in (11.22) we are computing the shortest path distance between x and x'. Hence, the following result is an immediate corollary of Theorem 20.

Corollary 10 *The only admissible way of inducing a metric in a symmetric network is by computing pairwise shortest paths as in* (11.22)*.*

At this point, the result in Corollary 10 is not surprising since we have encountered the same result in Section 8.3.1 when specifically analyzing metric projections. However, the value of Theorem 20 resides in the fact that we can obtain a series of results similar to Corollary 10 when analyzing different underlying dioids, as we do in the remainder of this chapter.

11.4.2 $\mathfrak{A}_{\min,\max} = (\bar{\mathbb{R}}_+, \min, \max)$

Since the dioids $\mathfrak{A}_{\min,\max} = (\bar{\mathbb{R}}_+, \min, \max)$ and $\mathfrak{A}_{\min,+}$ share the monoid $(\bar{\mathbb{R}}_+, \min)$, the specialization of the concept of dioid space to $\mathfrak{A}_{\min,\max}$ also leads to a symmetric weighted network. Nevertheless, when specializing the dioid triangle inequality (11.5) we use (11.1) to write

$$\exists\ c \in \bar{\mathbb{R}}_+ : \quad \min\big(c, \max(f(x,x''), f(x'',x'))\big) = f(x,x'). \tag{11.23}$$

Notice that expression (11.23) can be rewritten into the more familiar form $f(x,x') \leq \max(f(x,x''), f(x'',x'))$, which is exactly the strong triangle inequality. Hence, from Definition 1 it follows that a dioid metric space is equal to an ultrametric space when the underlying dioid is $\mathfrak{A}_{\min,\max}$. Moreover, the canonical dioid metric projection $\mathcal{P}^*$ in (11.9) boils down to

$$f^*(x,x') = \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} f(x_i, x_{i+1}), \tag{11.24}$$

which coincides with the definition of single linkage hierarchical clustering [cf. (2.15)]. Consequently, the following result can be extracted as a corollary of Theorem 20.

Corollary 11 *The only admissible hierarchical clustering method for symmetric networks is single linkage as defined in* (11.24)*.*

The main difference between metric and ultrametric spaces is that the former are constrained by the triangle inequality – based on adding dissimilarities – and the latter by the strong triangle inequality – based on maximum dissimilarities. By using dioid algebras we see how by literally exchanging the + operation by the max operation as the multiplication in the dioid, we can recast a uniqueness result valid for metric spaces into a result valid for ultrametric spaces.

11.4.3 $\mathfrak{A}_{\max,\times} = ([0,1], \max, \times)$

The dioid $\mathfrak{A}_{\max,\times}$ is defined on the interval $[0,1]$ with the max operation as its addition and the regular multiplication playing the role of the dioid multiplication. It is not hard to show, similar to Proposition 37, that the dioid $\mathfrak{A}_{\max,\times}$ is valid and it satisfies the hypotheses in Theorem 20. For example, 0 is the neutral element of max which is absorbing for $\times$ (as required in Definition 9) and its top element $T = \max([0,1]) = 1$ is equal to the neutral element of $\times$ (as in the statement of Theorem 20). Moreover, a dioid space underlain by $\mathfrak{A}_{\max,\times}$ is one where the weight between two different nodes $x \neq x'$ is a number in $[0,1)$ whereas the weight between x and itself is equal to 1. We interpret this as a probability space where each node is an event and the edge weights encode probabilities of co-occurrence. In this way, two events x and x' are indistinguishable if and only if they always occur jointly $f(x,x') = 1$. Moreover, to specialize the dioid triangle inequality (11.5) we again resort to (11.1) to write

$$\exists\ c \in [0,1]: \quad \max\big(c, f(x,x'') \times f(x'',x')\big) = f(x,x'), \tag{11.25}$$

which is equivalent to writing $f(x,x') \geq f(x,x'') \times f(x'',x')$. Under the interpretation of a probability space, this inequality is stating the reasonable property that the probability of x and x' co-occurring is at least equal to the probability of x, x'' *and* x'', x' co-occurring, assuming independence of the events. Hence, we denominate this dioid triangle inequality as the *event independence inequality*. In this way, a dioid metric space is a probability space abiding by the event independence inequality. Finally, if we specialize the canonical projection $\mathcal{P}^*$ in (11.9) to dioid $\mathfrak{A}_{\max,\times}$ we obtain

$$f^*(x,x') = \max_{P_{xx'}} \prod_{i | x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.26}$$

In words, $\mathcal{P}^*$ fixes the probability of co-occurrence of x and x' to the maximum probability among all chains of intermediate events that would ensure the co-occurrence of x and x'. Moreover, as a corollary to Theorem 20 we obtain the following result.

Corollary 12 *The only admissible way of inducing an event independence inequality on a probability space is via the maximum probability of co-occurrences as given by* (11.26)*.*

11.4.4 $\mathfrak{A}_{\max,\min} = (\{0,1\}, \max, \min)$

In Sections 11.4.1 through 11.4.3 we have analyzed the particularization of dioid metric projections and leveraged the selectivity of the operation $\oplus$ to apply Theorem 20. We now shift the focus to dioid semi-metric projections by analyzing projections for dioid semi-spaces underlain by $\mathfrak{A}_{\max,\min}$. More specifically, the edges in these dioid semi-spaces take values $f(x,x') \in \{0,1\}$ where $f(x,x)=1$ but different nodes $x \neq x'$ might still achieve the top value $T=1$. In this sense, we can interpret dioid semi-metric spaces as unweighted and undirected graphs, where a weight equal to 1 signalizes the existence of an edge and a weight of 0 indicates the absence of it. When specializing the dioid triangle inequality we have that

$$\exists\, c \in \{0,1\} : \quad \max\big(c, \min(f(x,x''), f(x'',x'))\big) = f(x,x'), \tag{11.27}$$

which can be rewritten as $f(x,x') \geq \min(f(x,x''), f(x'',x'))$. In terms of edge existence, this inequality implies that if two edges are present in a triangle, then the third edge must be present as well. This means that a dioid semi-metric space is entirely composed of full cliques disconnected between them. Hence, we denominate this inequality as the *clique inducing inequality*. Furthermore, for this dioid the canonical projection in (11.9) boils down to

$$f^*(x,x') = \max_{P_{xx'}} \min_{i | x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.28}$$

From (11.28) it follows that the canonical projection turns every connected component in the input graph into a clique. Furthermore, the following corollary follows from Proposition 36.

Corollary 13 *The only admissible way of inducing a graph whose connected components are clicks is via the canonical projection in* (11.28).

11.4.5 $\mathfrak{A}_{\cup,\cap} = (\mathcal{P}(A), \cup, \cap)$

The set E in a dioid $(E, \oplus, \otimes)$ need not be composed of numbers as we illustrate through the dioid $\mathfrak{A}_{\cup,\cap}$. Consider a finite set $A = \{a_1, \ldots, a_n\}$ where the objects a_i can be interpreted as topics. Hence, the set E consists of the power set $\mathcal{P}(A)$, i.e., all possible subsets of topics chosen from the finite collection A. A dioid space underlain by $\mathfrak{A}_{\cup,\cap}$ is one where the edge weight $f(x,x')$ is a subset of A and can be interpreted as the set of topics on which agents x and x' agree. It is clear that each agent x shares all the opinions with itself, i.e., $f(x,x)=A$ and we allow the possibility that two distinct agents also coincide in their opinion for all issues, thus considering the set of dioid semi-spaces. Regarding the dioid triangle inequality, we leverage (11.1) to express it as

$$\exists\, c \in \mathcal{P}(A) : \quad c \cup (f(x,x'') \cap f(x'',x')) = f(x,x'), \tag{11.29}$$

which immediately implies that $f(x,x') \supseteq f(x,x'') \cap f(x'',x')$. Hence, a dioid semi-metric space is one in which if a topic appears in two edges of a triangle then it must be contained in the remaining edge. Thus, we denominate this inequality as the *agreement inequality*.

Moreover, the canonical projection in (11.9) particularizes to

$$f^*(x,x') = \bigcup_{P_{xx'}} \bigcap_{i|x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.30}$$

The projection in (11.30) checks all paths connecting x and x' and if for one path there is a topic a_i for which all intermediate links in the path agree, then it must be that x and x' also agree and the projection forces $a_i \in f^*(x,x')$. Further, the following corollary can be obtained from Proposition 36.

Corollary 14 *The only admissible way of inducing a network of agreements is via the canonical projection in* (11.30)*.*

11.4.6 $\mathfrak{A}_{\cap,\cup} = (\mathcal{P}(A), \cap, \cup)$

The last dioid that we analyze is the counterpart of the previous one where we exchange the order of the operations $\cap$ and $\cup$ and still consider weights from the power set of a prescribed A. In this case, the dioid triangle inequality has the opposite implication since

$$\exists\, c \in \mathcal{P}(A): \quad a \cap (f(x,x'') \cup f(x'',x')) = f(x,x') \tag{11.31}$$

can be rewritten as $f(x,x') \subseteq f(x,x'') \cup f(x'',x')$. A dioid semi-metric is then one in which if a topic appears in one edge of a triangle then it must appear at least in one other edge of the triangle. Hence, we give this inequality the name *disagreement inequality*. Moreover, the canonical projection in (11.9) boils down to to

$$f^*(x,x') = \bigcap_{P_{xx'}} \bigcup_{i|x_i \in P_{xx'}} f(x_i, x_{i+1}). \tag{11.32}$$

If there is at least one path joining x and x' along which a given topic a_i does not appear – i.e., is not disagreed upon – then the canonical projection forces x and x' not to disagree, meaning $a_i \notin f^*(x,x')$. Moreover, the following corollary of Proposition 36 can be stated.

Corollary 15 *The only admissible way of inducing a network of disagreements is via the canonical projection in* (11.32)*.*

11.4.7 Other dioids

Additional valid dioids can be shown to abide by the hypotheses of Theorem 20 or Proposition 36. For instance, consider the dioid $\mathfrak{A}_{\gcd,\times} = (\bar{\mathbb{N}}, \gcd, \times)$ defined in the extended strictly positive integers where $\gcd(a,b)$ returns the greatest common divisor of a and b. Furthermore notice that gcd is an idempotent (but not selective) operation, $\times$ is commutative, and the top element $\gcd(\bar{\mathbb{N}}) = 1$ coincides with the neutral element of $\times$, thus, complying with the hypotheses of Proposition 36. A dioid space, in this case, is a network with edges taking values in the positive integers and, mimicking the developments in Sections 11.4.1 to 11.4.6, it can be inferred that the dioid triangle inequality implies that for every triangle the product of the weights of two edges must be divisible by the third one. Moreover, Proposition 36 guarantees that particularizing the canonical projection for this dioid provides the only admissible way of imposing such a structure to a network with positive integers as weights. Nevertheless, an interpretation for the need of such a requirement is not as clear as for the dioids previously introduced. Another dioid that satisfies the hypothesis of Proposition 36 is $\mathfrak{A}_{\mathrm{gcd,lcm}} = (\bar{\mathbb{N}}, \mathrm{gcd}, \mathrm{lcm})$ where the operation $\mathrm{lcm}(a, b)$ returns the least common multiple between the positive integers a and b.

In Table 11.1 we summarize the particularization of the dioid uniqueness result in Section 11.3 for the six dioids considered in Sections 11.4.1 to 11.4.6.

Table 11.1: Dioid spaces considered

Dioid	Dioid space	Triangle inequality	Metric space	Canonical projection
$\mathfrak{A}_{\min,+}$	weighted network	$f(x,x') \leq f(x,x'') + f(x'',x')$	regular metric	$\min_{P_{xx'}} \sum_{i \mid x_i \in P_{xx'}} f(x_i, x_{i+1})$
$\mathfrak{A}_{\min,\max}$	weighted network	$f(x,x') \leq \max(f(x,x''), f(x'',x'))$	ultrametric	$\min_{P_{xx'}} \max_{i \mid x_i \in P_{xx'}} f(x_i, x_{i+1})$
$\mathfrak{A}_{\max,\times}$	probability space	$f(x,x') \geq f(x,x'')\, f(x'',x')$	independent event	$\max_{P_{xx'}} \prod_{i \mid x_i \in P_{xx'}} f(x_i, x_{i+1})$
$\mathfrak{A}_{\max,\min}$	unweighted network	$f(x,x') \geq \min(f(x,x''), f(x'',x'))$	clique graph	$\max_{P_{xx'}} \min_{i \mid x_i \in P_{xx'}} f(x_i, x_{i+1})$
$\mathfrak{A}_{\cup,\cap}$	topic network	$f(x,x') \supseteq f(x,x'') \cap f(x'',x')$	agreement	$\cup_{P_{xx'}} \cap_{i \mid x_i \in P_{xx'}} f(x_i, x_{i+1})$
$\mathfrak{A}_{\cap,\cup}$	topic network	$f(x,x') \subseteq f(x,x'') \cup f(x'',x')$	disagreement	$\cap_{P_{xx'}} \cup_{i \mid x_i \in P_{xx'}} f(x_i, x_{i+1})$

Chapter 12

Conclusions and future directions

The taxonomic analysis in Chapter 7 served a double purpose also as a concluding summary of the first part of this thesis. In a nutshell, the goal of Part I was to develop an axiomatic theory of hierarchical clustering for asymmetric networks. Part II can be seen as a natural extension of Part I where, instead of considering projections of networks onto ultrametric spaces – equivalent to hierarchical clustering –, we widen our scope to consider projections onto more general classes of metric-like spaces. Chapter 8 formally defined this problem and then focused on the projection of *symmetric* networks onto q-metric spaces, these being a parametric family of spaces containing the ultrametric and the (regular) metric spaces. The main contributions of this chapter were three. First, it formally laid the axiomatic foundations for the theory developed in all of Part II by introducing the Axioms of Projection (AA1) and Injective Transformation (AA2). Second, it defined the canonical q-metric projection and showed that, for a given q, this is the only admissible way of inducing a q-metric structure on a symmetric network. By specializing the result for $q = 1$, we proved that shortest paths are the only admissible (regular) metric in graphs. Lastly, this chapter presented three important properties of the canonical projections including their robustness to noise, the possibility to be used in the determination of lower bounds for combinatorial graph problems, and their invariance to order when sequentially applying multiple projections.

Chapter 9 leveraged the insights of the preceding chapter and examined a more challenging problem: the projection of *asymmetric* networks. Nevertheless, in order to facilitate understanding, the image of the projections was constrained to 1-metric spaces. In this sense, Chapter 9 resembles the core of Part I but with asymmetric networks projected onto 1-metric – instead of ultrametric – spaces. Hence, not surprisingly, the contributions in Chapter 9 mimicked the developments in Part I. More specifically, after modifying the axiomatic framework to accommodate asymmetric inputs, the SymPro and ProSym projections were introduced. These two methods induce metric structure in networks by imposing the properties of symmetry and triangle inequality in a sequential manner: SymPro first symmetrizes the network and then induces the triangle inequality whereas the opposite is true for ProSym. We further showed that SymPro and ProSym bound in a well-defined sense all other admissible metric projection methods. More precisely, the pairwise distances output by any other method when applied to an arbitrary network are lower bounded by the ProSym distances and upper bounded by SymPro distances. Lastly, we proved the existence of intermediate methods contained between ProSym and SymPro by effectively constructing the family of PSP methods, whose action can be intuitively understood as first inducing a partial fulfillment of the triangle inequality, then symmetrizing, and then completing the triangle inequality attainment.

In Chapter 10 we presented extensions and applications of the theory developed in the previous two chapters. In terms of extensions, we analyzed the projections of asymmetric networks onto quasi-metric and q-metric spaces. Regarding the former, we showed uniqueness of a canonical projection, reminiscent to the findings on quasi-clustering in Part I. Concerning the latter, we showed that under mild additional assumptions the theory in Chapter 9 developed for $q = 1$ can be extended for arbitrary q. In terms of applications, our focus was around the efficient search in networks and the visualization of asymmetric data. The structure imposed by the triangle inequality can be used to accelerate nearest-neighbor search in metric spaces as opposed to general weighted networks. Thus, we analyzed the benefits of performing search in networks by first projecting them onto (quasi-)q-metric spaces and then leveraging the additional structure in these spaces. Lastly, regarding data visualization, we incorporated metric projections as preprocessing steps before the application of existing low-dimensional embedding procedures and empirically showed that different metric projections unveil different aspects of the asymmetric data.

Chapter 11 presented the first results in an attempt to study metric representations of networks from an algebraic viewpoint via the introduction of dioids. A dioid is an *ordered* algebraic structure, and this order was essential in the generalization of the triangle inequality to this more abstract domain. In this way, we studied projections from dioid spaces – natural extensions of symmetric networks – onto dioid *metric* spaces, i.e., dioid spaces governed by the generalized triangle inequality. A uniqueness result similar to the one obtained in Chapter 8 was shown and, by specializing the underlying dioid, we particularized this result to different domains of interest.

In the following sections we present three potential directions for future research. Although not necessarily sorted in increasing order of difficulty, we consider them to be sorted in decreasing order of perspicuity given the current state of the art.

12.1 Intermediate q-metric projections and properties

The depth with which Parts I and II treat the problems stated therein is different, and there is a natural reason for this disparity. In Part I we considered a well-established problem – hierarchical clustering in networks – with decades' worth of prior work in the area, thus, to achieve a significant contribution we delved into more advanced concepts like representability and excisiveness. By contrast, a substantial portion of the value in Part II came from the problem formulation itself and from the idea of generalizing tools originally conceived to study hierarchical clustering and extending them for the determination of metric representations of networks. Consequently, the current state of this more general theory is not as sophisticated as that of hierarchical clustering and, as a first direction for future research, we propose expanding the general theory in Part II to the level of depth attained in Part I. To be more specific and excluding for now the concept of dioids (see Section 12.3 for related future directions), we propose to extend the study of intermediate admissible methods – contained between ProSym and SymPro – for the projection of asymmetric networks onto q-metric spaces. Expecting a wide gamut of intermediate methods akin to those found in Section 3.3 for hierarchical clustering, we envision the generalization of the properties of scale preservation, excisiveness, and representability to further winnow the set of admissible methods. Although we do not foresee major obstacles in this direction, the generalization of the aforementioned properties is not straightforward either. For example, the natural generalization of scale preservation describes a projection method for which the metric induced on a previously rescaled network equates rescaling the metric obtained from the original network. However, ingrained in this definition is the fact that the change of scale should preserve the fulfillment of the triangle inequality, thus imposing, e.g., that the rescaling must correspond to a subadditive function. This kind of problems were not encountered in the analysis of hierarchical clustering since monotonicity of the rescaling was sufficient to preserve fulfillment of the strong triangle inequality. Similar challenges are expected to be found when generalizing excisiveness and representability.

12.2 Milder axiomatic constructions

The general landscape of admissibility for the different axiomatic constructions considered is that for symmetric networks we tend to have a unique admissible projection – single linkage for ultrametrics, shortest path for metrics – whereas for asymmetric networks we have an infinite set of admissible methods bounded by two extreme ones – reciprocal and nonreciprocal for ultrametrics, SymPro and ProSym for metrics. However, the stringency of the unicity result for symmetric networks might be objectionable in some contexts. In hierarchical clustering, e.g., single linkage has shown to have some undesirable features like the so-called chaining effect [75]. As a second research avenue we propose the exploration of alternative (weaker) axiomatic constructions so that density aware methods, that do not suffer from the chaining effect, become admissible. To be more precise, recall that in Remark 19 we stated that the Symmetric Axiom of Value (B1) and the Axiom of Transformation (A2) compared to the Axiom of Projection (AA1) and the Axiom of Injective Transformation (AA2) impose the same constraints on the set of admissible methods when considered jointly. Nevertheless, (B1) is weaker than (AA1) since the former imposes a requirement solely on two-node networks and (AA2) is weaker than (A2) since the former imposes a condition only for injective maps. Therefore, by combining the weaker axiom of each pair, i.e. (B1) and (AA2), we can achieve the laxer axiomatic framework wanted. Moreover, further alternatives can be studied such as modifying axiom (AA2) to consider surjective instead of injective maps. This direction has been proposed in [11] for the study of hierarchical clustering methods. Here we propose to deepen the analysis and extend it to the projection of networks onto general q-metric spaces.

12.3 Further exploration of dioid spaces

The use of dioids to attain an algebraic generalization of the problems of hierarchical clustering and metric projections was discussed in Chapter 11. Nevertheless, we still deem this promising area as mostly unexplored and propose three avenues for potential exploration: expanding the axiomatic framework and the admissibility conditions, gaining an algebraic understanding of the algorithms, and unveiling unconventional applications for dioid spaces.

In terms of the extension of the axiomatic theory, the direction here suggested is similar to the one proposed for q-metric projections in Section 12.1. More specifically, we propose to extend to the level of detail attained in Part I the axiomatic study of projections for dioid spaces. The envisioned challenges, however, exceed those in Section 12.1. For example, the first step in the extension of the theory for dioids would be to encompass *asymmetric* dioid spaces. However, as was seen throughout the thesis, in generating a symmetric output from an asymmetric input, a symmetrizing function must be utilized. Nevertheless, our intuitive understanding of a symmetrizing function in, e.g., the real numbers, is lost when considering general dioids. To be more specific, if a dioid takes values in the power set $\mathcal{P}(A)$ of a prescribed vocabulary set A as studied in Sections 11.4.5 and 11.4.6, what would be a reasonable notion of symmetrization? Given that the elements in $\mathcal{P}(A)$ are only partially and not totally ordered, finding a symmetrized value contained between two input elements is not always feasible. As is the case for the symmetrization function, in extending the theory for dioids we foresee several instances in which we will have to depart from the intuition gained in real-valued networks and root the developments in the algebraic properties of the dioid at hand.

Dioids were briefly introduced in Section 3.5 as a tool for the design of algorithms for hierarchical clustering. In the mentioned section, matrix powers in a $(\bar{\mathbb{R}}_+, \min, \max)$ dioid algebra were essential in devising algorithms since these computed the minimum cost of paths of a prescribed edge-length joining two nodes. In the same way, we may use dioid powers in $(\bar{\mathbb{R}}_+, \min, +)$ to find the length of the shortest paths between two points. If we focus on symmetric networks, this implies that the output corresponding to single linkage and the all-pairs shortest paths can be found using the same procedure with a change in the specification of the underlying dioid. Nevertheless, employing alternative algorithms, single linkage can be implemented in $O(n^2)$ time [75] whereas the shortest paths require $O(n^3)$ [27, 90]. This points towards the fact that, even though dioid matrix powers can be used to design algorithms for *any* dioid, additional properties of *particular* dioids can be further leveraged in the algorithm design. The computational gain of single linkage in comparison with shortest paths might come from the fact that $(\bar{\mathbb{R}}_+, \min, \max)$ is a doubly-selective dioid – both min and max are selective operations – whereas $(\bar{\mathbb{R}}_+, \min, +)$ is singly-selective. A better understanding of this fact as well as a design of algorithms that best fit the properties of the underlying dioid is an appealing and useful research direction.

The generality of dioids renders the development of the associated theory more challenging but, at the same time, more powerful. Nevertheless, the extent of the applicability of such a theory is still to be revealed. In the future, we are particularly interested in studying the relation between dioid spaces taking values in the power set $\mathcal{P}(A)$ and the concept of knowledge graphs [54, 78, 89]. Knowledge graphs are structures encoding entities and relations between them. E.g., a mother and her daughter could be related via 'parent of' in one direction and 'children of' in the opposite one. Notice however that some relations can be inferred or completed from existing ones, e.g., if entity a is 'parent of' b and b is 'parent of' c, then a must be 'grandparent of' c. At a high level of abstraction we could conceive this triadic constraint as a triangle inequality and, thus, interpret a dioid metric space as a complete knowledge graph. Consequently, the problem of knowledge graph completion [54, 78] could be posed as one of metric projections in dioid spaces. A tempting goal, although not clear if achievable at this point, is to develop this theory to the point in which the computation of shortest paths in classical weighted graphs and the completion of knowledge graphs would be achieved by the same algorithm by simply modifying the underlying dioid.

Appendix A

Appendix

A.1 Proof of Theorem 11

We begin by showing sufficiency of the statement. From Propositions 15 and 16, we know that any representable method outputs a valid ultrametric and satisfies the Axiom of Transformation (A2). Consequently, in order to show admissibility, it is enough to prove fulfillment of the Axiom of Value (A1).

Consider the clustering method $\mathcal{H}^{\Omega}$ where Ω is a collection of strongly connected, structure representers. Also, consider a two-node network $\vec{\Delta}_2(\alpha, \beta) = (\{p, q\}, A_{p,q})$ with $A_{p,q}(p, q) = \alpha$ and $A_{p,q}(q, p) = \beta$. Denote by $(\{p, q\}, u^{\Omega}_{p,q}) = \mathcal{H}^{\Omega}(\vec{\Delta}_2(\alpha, \beta))$ the ultrametric output of applying $\mathcal{H}^{\Omega}$ to $\vec{\Delta}_2(\alpha, \beta)$. We want to show that

$$u^{\Omega}_{p,q}(p, q) = \max(\alpha, \beta) = m. \tag{A.1}$$

Take any $\omega = (X_\omega, A_\omega) \in \Omega$ and consider the network $m * \omega = (X_\omega, mA_\omega)$. Pick any node $\bar{x} \in X_\omega$ and construct a map $\phi_{p,q} : X_\omega \to \{p, q\}$ such that $\phi_{p,q}(\bar{x}) = p$ and $\phi_{p,q}(\bar{x}') = q$ for all $\bar{x}' \in X_\omega$ such that $\bar{x}' \neq \bar{x}$. We are assured that at least one $\bar{x}'$ exists since $|X_\omega| \geq 2$ from the definition (5.38) of structure representer.

To see that $\phi_{p,q}$ going from network $m * \omega$ to $\vec{\Delta}_2(\alpha, \beta)$ is dissimilarity reducing, note that

$$m * A_\omega(x, x') = m, \tag{A.2}$$

for $(x, x') \in \mathrm{dom}(A_\omega)$ with $x \neq x'$ since ω is a structure representer. Furthermore, from the definition of m in (A.1) we have that

$$A_{p,q}\big(\phi_{p,q}(x), \phi_{p,q}(x')\big) \leq m, \tag{A.3}$$

for $x, x' \in X_\omega$. Hence, from (A.2) and (A.3) we conclude that $\phi_{p,q}$ is a dissimilarity reducing map and, from (5.24) we know that

$$u^{\Omega}_{p,q}(p, q) \leq m, \tag{A.4}$$

since m is a multiple of ω that allows the construction of the dissimilarity reducing map $\phi_{p,q}$ with $p, q \in \mathrm{Im}(\phi_{p,q})$. Nevertheless, we still need to show that m is the minimum possible multiple of any $\omega \in \Omega$ that permits the construction of such map, to prove that (A.4) is in fact an equality.

Suppose that for $\delta < m$ we can build a dissimilarity reducing map $\phi'_{p,q} : X_\omega \to \{p, q\}$ with $p, q \in \mathrm{Im}(\phi'_{p,q})$ from $\delta * \omega = (X_\omega, \delta A_\omega)$ to $\vec{\Delta}_2(\alpha, \beta)$ for some $\omega \in \Omega$. Let us partition the set X_ω into two blocks depending on the image of its nodes, i.e. $X_\omega = \{B_p, B_q\}$ where $\phi'_{p,q}(x_p) = p$ for all $x_p \in B_p$ and $\phi'_{p,q}(x_q) = q$ for all $x_q \in B_q$.

Since we assume $\phi'_{p,q}$ to be dissimilarity reducing, we must have either

$$\delta A_\omega(x_p, x_q) \geq m, \quad \text{or} \quad \delta A_\omega(x_q, x_p) \geq m, \tag{A.5}$$

for all $x_p \in B_p$ and $x_q \in B_q$ where the dissimilarities are defined. Nevertheless, from our hypothesis of structure representers, we know that $\delta A_\omega(x, x') = \delta < m$ for all $(x, x') \in \text{dom}(A_\omega)$ such that $x \neq x'$. Consequently, (A.5) implies that either

$$(x_p, x_q) \notin \text{dom}(A_\omega), \quad \text{or} \quad (x_q, x_p) \notin \text{dom}(A_\omega), \tag{A.6}$$

for all $x_p \in B_p$ and $x_q \in B_q$. However, (A.6) implies that components B_p and B_q are not strongly connected in $\delta * \omega$ which means that ω is not strongly connected, reaching a contradiction.

Since no dissimilarity reducing map $\phi'_{p,q}$ can be found, we conclude that

$$u^{\Omega}_{p,q}(p, q) = m. \tag{A.7}$$

Because the two-node network and the collection Ω were picked arbitrarily, the fulfillment of axiom (A1) is shown.

To complete the proof of this direction of the implication in the theorem, we need to show that $\mathcal{H}^\Omega$ is a scale preserving method as defined in (P1). To do so, consider an arbitrary network $N_X = (X, A_X)$ and a nondecreasing function ψ satisfying the conditions in (P1) and define $N_Y = (Y, A_Y)$ with $Y = X$ and $A_Y = \psi \circ A_X$. Then, we want to show that

$$\psi \circ u^{\Omega}_X = u^{\Omega}_Y, \tag{A.8}$$

where $\mathcal{H}^\Omega(X, A_X) = (X, u^\Omega_X)$ and $\mathcal{H}^\Omega(X, A_Y) = (Y, u^\Omega_Y)$.

Take an arbitrary pair of nodes $x, x' \in X$. According to (5.24), there exists a minimizing path $P^*_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$, a sequence of corresponding multiples $\lambda^\Omega_X(x_i, x_{i+1})$ such that

$$\lambda^\Omega_X(x_i, x_{i+1}) \leq u^\Omega_X(x, x'), \tag{A.9}$$

and a sequence of representers $\omega_i \in \Omega$ for $i = 0, ..., l-1$ such that we can construct l dissimilarity reducing maps $\phi_{x_i,x_{i+1}} : X_{\omega_i} \to X$ from $\lambda^\Omega_X(x_i, x_{i+1}) * \omega_i$ to N_X where $x_i, x_{i+1} \in \text{Im}(\phi_{x_i,x_{i+1}})$.

Construct the path $P_{yy'} = [y = y_0, y_1, \ldots, y_l = y']$ in N_Y such that $x_i = y_i$ for all i. We now argue that the aforementioned maps $\phi_{x_i,x_{i+1}}$ when interpreted as maps $\phi_{x_i,x_{i+1}} : X_{\omega_i} \to Y$ containing y_i and y_{i+1} in their image are dissimilarity reducing maps from $\psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big) * \omega_i$ to N_Y. To see this, note that since ω_i is a structure representer,

$$\lambda_X^\Omega(x_i, x_{i+1})\, A_{\omega_i}(x_\omega, x'_\omega) = \lambda_X^\Omega(x_i, x_{i+1}), \tag{A.10}$$

for all $x_\omega \neq x'_\omega$ where the dissimilarity function A_{ω_i} is defined. Since $\phi_{x_i,x_{i+1}}$ is dissimilarity reducing from $\lambda_X^\Omega(x_i, x_{i+1}) * \omega_i$ to N_X, (A.10) implies that

$$A_X(\phi_{x_i,x_{i+1}}(x_\omega), \phi_{x_i,x_{i+1}}(x'_\omega)) \leq \lambda_X^\Omega(x_i, x_{i+1}), \tag{A.11}$$

for all $x_\omega, x'_\omega \in \mathrm{dom}(A_{\omega_i})$. If we apply ψ to (A.11) the inequality is preserved since ψ is nondecreasing. Thus, recalling $A_Y = \psi \circ A_X$ we have that

$$A_Y(\phi_{x_i,x_{i+1}}(x_\omega), \phi_{x_i,x_{i+1}}(x'_\omega)) \leq \psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big). \tag{A.12}$$

In a similar way as we obtained (A.10), since ω_i is a structure representer, we have that

$$\psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big)\, A_{\omega_i}(x_\omega, x'_\omega) = \psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big), \tag{A.13}$$

for all $x_\omega \neq x'_\omega$ where the dissimilarity function is defined. Hence, substituting (A.13) in (A.12) we obtain that

$$A_Y(\phi_{x_i,x_{i+1}}(x_\omega), \phi_{x_i,x_{i+1}}(x'_\omega)) \leq \psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big)\, A_{\omega_i}(x_\omega, x'_\omega), \tag{A.14}$$

for all $x_\omega, x'_\omega \in \mathrm{dom}(A_{\omega_i})$. Consequently, $\phi_{x_i,x_{i+1}} : X_{\omega_i} \to Y$ is a dissimilarity reducing map from $\psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big) * \omega_i$ to N_Y as we wanted to show. Moreover, by construction $x_i = y_i, x_{i+1} = y_{i+1} \in \mathrm{Im}(\phi_{x_i,x_{i+1}})$. This implies that,

$$u_Y^\Omega(y_i, y_{i+1}) \leq \psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big), \tag{A.15}$$

for every pair y_i, y_{i+1} of consecutive nodes in $P_{yy'}$. Thus, from the strong triangle inequality we have that,

$$u_Y^\Omega(y, y') \leq \max_i u_Y^\Omega(y_i, y_{i+1}) \leq \max_i \psi\big(\lambda_X^\Omega(x_i, x_{i+1})\big) \leq \psi\big(u_X^\Omega(x, x')\big), \tag{A.16}$$

where we used (A.15) for the second inequality and (A.9) for the third one.

We still have to prove that the inequalities in (A.16) cannot be strict, in order to conclude that

$$u_Y^\Omega(y, y') = \psi\big(u_X^\Omega(x, x')\big), \tag{A.17}$$

which, since the network choice was arbitrary, would show scale preservation of $\mathcal{H}^\Omega$ [cf.

(A.8)].

From the definition of representability, there must exist a path $P^*_{yy'} = [y = y_0, y_1, ..., y_{l'} = y']$, a sequence of multiples $\lambda^{\Omega}_Y(y_i, y_{i+1})$ such that

$$\lambda^{\Omega}_Y(y_i, y_{i+1}) \leq u^{\Omega}_Y(y, y'), \tag{A.18}$$

and a sequence of representers $\omega_i \in \Omega$ such that we can construct l' dissimilarity reducing maps $\phi_{y_i,y_{i+1}}$ from $\lambda^{\Omega}_Y(y_i, y_{i+1}) * \omega_i$ to N_Y. Construct the path $P_{xx'} = [x = x_0, x_1, \ldots, x_{l'} = x']$ in N_X such that $x_i = y_i$ for all i.

Let us define the inverse function $\eta : \mathrm{Im}(A_Y) \to \mathbb{R}_+$ from the set of dissimilarities in the network N_Y to the nonnegative reals such that

$$\eta(\beta) = \{\max \alpha | \psi(\alpha) = \beta, \exists\, x, x' \in X \text{ s.t. } A_X(x, x') = \alpha\}. \tag{A.19}$$

In (A.19), $\eta(\beta)$ returns the value of the maximum dissimilarity in A_X that is transformed to β by ψ. Notice that η is a nondecreasing function due to the nondecreasing nature of ψ. Moreover, η satisfies

$$\psi \circ \eta = \mathrm{Id}, \tag{A.20}$$

$$\eta \circ \psi \geq \mathrm{Id} \tag{A.21}$$

Since each ω_i has unit dissimilarities, each optimal multiple $\lambda^{\Omega}_Y(y_i, y_{i+1}) \in \mathrm{dom}(A_Y)$ must be equal to some dissimilarity value in A_Y. Moreover, from (A.21), we have that

$$\eta\big(A_Y(y, y')\big) \geq A_X(x, x'), \tag{A.22}$$

for all $x = y$ and $x' = y'$. In the same manner that we wrote (A.10) and (A.13), we may again rely on the fact that ω_i is a structure representer to state

$$\lambda^{\Omega}_Y(y_i, y_{i+1})\, A_{\omega_i}(x_\omega, x'_\omega) = \lambda^{\Omega}_Y(y_i, y_{i+1}), \tag{A.23}$$

$$\eta\big(\lambda^{\Omega}_Y(y_i, y_{i+1})\big)\, A_{\omega_i}(x_\omega, x'_\omega) = \eta\big(\lambda^{\Omega}_Y(y_i, y_{i+1})\big), \tag{A.24}$$

for all $x_\omega \neq x'_\omega$ where the dissimilarity function A_{ω_i} is defined. Since $\phi_{y_i,y_{i+1}}$ is a dissimilarity reducing map from $\lambda^{\Omega}_Y(y_i, y_{i+1}) * \omega_i$ to N_Y, we have that

$$A_Y(\phi_{y_i,y_{i+1}}(x_\omega), \phi_{y_i,y_{i+1}}(x'_\omega)) \leq \lambda^{\Omega}_Y(y_i, y_{i+1})\, A_{\omega_i}(x_\omega, x'_\omega), \tag{A.25}$$

for all $x_\omega, x'_\omega \in \mathrm{dom}(A_{\omega_i})$. Substituting (A.23) in (A.25) we obtain that

$$A_Y(\phi_{y_i,y_{i+1}}(x_\omega), \phi_{y_i,y_{i+1}}(x'_\omega)) \leq \lambda^{\Omega}_Y(y_i, y_{i+1}). \tag{A.26}$$

If we apply η to (A.26) the inequality is preserved since η is increasing. Moreover, using inequality (A.22) we may write

$$A_X(\phi_{y_i,y_{i+1}}(x_\omega), \phi_{y_i,y_{i+1}}(x'_\omega)) \leq \eta\big(\lambda_Y^\Omega(y_i, y_{i+1})\big). \tag{A.27}$$

We substitute (A.24) in (A.27) to obtain

$$A_X(\phi_{y_i,y_{i+1}}(x_\omega), \phi_{y_i,y_{i+1}}(x'_\omega)) \leq \eta\big(\lambda_Y^\Omega(y_i, y_{i+1})\big)\, A_{\omega_i}(x_\omega, x'_\omega), \tag{A.28}$$

for all $x_\omega, x'_\omega \in \mathrm{dom}(A_{\omega_i})$. Consequently, $\phi_{y_i,y_{i+1}} : X_{\omega_i} \to X$ is a dissimilarity reducing map from $\eta\big(\lambda_Y^\Omega(y_i, y_{i+1})\big) * \omega_i$ to N_X. Furthermore, by construction $x_i = y_i, x_{i+1} = y_{i+1} \in \mathrm{Im}(\phi_{y_i,y_{i+1}})$ implying that,

$$u_X^\Omega(x_i, x_{i+1}) \leq \eta\big(\lambda_Y^\Omega(y_i, y_{i+1})\big), \tag{A.29}$$

for every pair x_i, x_{i+1} of consecutive nodes in $P_{xx'}$. Thus, from the strong triangle inequality we have that,

$$u_X^\Omega(x, x') \leq \max_i u_X^\Omega(x_i, x_{i+1}) \leq \max_i \eta\big(\lambda_Y^\Omega(y_i, y_{i+1})\big) \leq \eta\big(u_Y^\Omega(y, y')\big), \tag{A.30}$$

where we used (A.29) for the second inequality and (A.18) for the third one. If we apply the function ψ to (A.30) and use equivalence (A.20), we obtain that

$$\psi\big(u_X^\Omega(x, x')\big) \leq \psi\Big(\eta\big(u_Y^\Omega(y, y')\big)\Big) = u_Y^\Omega(y, y'), \tag{A.31}$$

which, combined with (A.16), shows equality (A.17) concluding the scale preservation proof.

In order to prove the necessity statement, we start with an admissible, scale preserving, representable method $\mathcal{H}^\Omega$ and we want to show that it can be represented by a collection of strongly connected, structure representers. We begin by showing the strong connectedness and the condition on the cardinality of X_ω for the structure representers $\omega = (X_\omega, A_\omega) \in \Omega$. The condition that $|X_\omega| \geq 2$ is immediate. Suppose Ω contains some single node networks, then generate a new collection $\Omega' = \{\omega \in \Omega \big| |X_\omega| \geq 2\}$. It should be clear that $\mathcal{H}^\Omega \equiv \mathcal{H}^{\Omega'}$ since no dissimilarity reducing map can be constructed from a point to a network where the image set contains two different nodes [cf. (5.20)]. Hence, every representable method can be represented without using single node networks.

Strong connectedness is implied by admissibility, in particular, by the fulfillment of the Axiom of Value (A1). From the definition of representable clustering method, all representers must be weakly connected. However, assume that a particular representer, say $\omega = (X_\omega, A_\omega)$, is not strongly connected. We use the result in the following claim.

Claim 10 *Let $\omega = (X_\omega, A_\omega)$ be a representer which is not strongly connected then the node set X_ω can be partitioned into two blocks $X_\omega = \{B, B'\}$ such that*

$$(x, x') \notin dom(A_\omega), \tag{A.32}$$

for all $x \in B$, $x' \in B'$.

Proof: Since ω is not strongly connected then there must exist two nodes $x, x' \in X_\omega$ such that x' cannot be reached by a directed path from x. Focus on the partition $P_1 = \{B_1, B_1'\}$ where $B_1 = \{x\}$ and $B_1' = X \backslash \{x\}$. If this partition satisfies the statement of the claim, we are done. Otherwise, there must be a node $x_2 \in B_1'$ such that $(x, x_2) \in \text{dom}(A_\omega)$. Thus, consider now the partition $P_2 = \{B_2, B_2'\}$ where $B_2 = \{x, x_2\}$ and $B_2' = X \backslash B_2$, and apply the same reasoning. If P_2 satisfies the conditions of the claim, we are done. Otherwise, there must exist $x_3 \in B_2'$ such that either (x, x_3) or (x_2, x_3) belong to $\text{dom}(A_\omega)$. Either way, x_3 is reachable from x – directly or through x_2. Then, consider the partition $P_3 = \{B_3, B_3'\}$ where $B_3 = \{x, x_2, x_3\}$ and $B_3' = X \backslash B_3$. It must be that a partition P_i for $i \in \{1, \ldots, n-1\}$ satisfies the claim. Otherwise, x_i would equal x' for some i since all x_i are distinct implying that x' is reachable from x and contradicting the statement of the claim. ■

Claim 10 ensures that the node set X_ω can be partitioned into two blocks $X_\omega = \{B, B'\}$ such that

$$(x, x') \notin \text{dom}(A_\omega) \ , \ \min_{x,x'} A_\omega(x', x) = \gamma, \tag{A.33}$$

for all $x \in B$, $x' \in B'$ and some $\gamma > 0$. Consider a two-node network $\vec{\Delta}_2(\alpha, \beta) = (\{p, q\}, A_{p,q})$ with $A_{p,q}(p, q) = \alpha$ and $A_{p,q}(q, p) = \beta$. Assume that $\alpha > \beta$. Since axiom (A1) must be true for all two-node networks, we can assume this particular case. Denote by $(\{p, q\}, u^\Omega_{p,q}) = \mathcal{H}^\Omega(\vec{\Delta}_2(\alpha, \beta))$ the ultrametric output of applying $\mathcal{H}^\Omega$ to $\vec{\Delta}_2(\alpha, \beta)$. Construct a map $\phi : X_\omega \to \{p, q\}$ such that $\phi(x) = p$ for all $x \in B$ and $\phi(x') = q$ for all $x' \in B'$. This map ϕ is dissimilarity reducing when going from $\frac{\beta}{\gamma} * \omega$ to $N_{p,q}$ since

$$\min_{x,x'} \frac{\beta}{\gamma} A_\omega(x', x) = \beta, \tag{A.34}$$

for all $x \in B$ and $x' \in B'$. But this means that $u^\Omega_{p,q}(p, q) \leq \beta/\gamma$. By choosing β small enough we obtain that

$$u^\Omega_{p,q}(p, q) < \alpha = \max(\alpha, \beta), \tag{A.35}$$

which contradicts the Axiom of Value (A1). Hence, all $\omega \in \Omega$ must be strongly connected.

The last condition we need to show is that the given admissible, representable, and scale preserving method $\mathcal{H}^\Omega$ can be represented by a collection of representers Ω', i.e. $\mathcal{H}^\Omega \equiv \mathcal{H}^{\Omega'}$, where every representer $\omega' \in \Omega'$ has all its positive defined dissimilarities equal to 1. We show this by transforming the representers $\omega \in \Omega$ into structure representers, i.e. with unit dissimilarities, and checking that the represented method does not change. A useful concept in this proof will be the set of maps $\Phi^{\omega}_{x,x'}$ that go from a representer $\omega = (X_\omega, A_\omega)$ to a network $N = (X, A_X)$ having x and x' in its image, i.e.

$$\Phi^{\omega}_{x,x'} = \{\phi : X_\omega \to X \quad \text{s.t.} \quad x, x' \in \mathrm{Im}(\phi)\}. \tag{A.36}$$

Given an arbitrary representer $\omega = (X_\omega, A_\omega)$, define the sets

$$D^{\omega}_{u} = \{(z, z') \in \mathrm{dom}(A_\omega) | A_\omega(z, z') > 1\}, \quad D^{\omega}_{l} = \{(z, z') \in \mathrm{dom}(A_\omega) | A_\omega(z, z') \leq 1\}, \tag{A.37}$$

containing the pairs of nodes in $\mathrm{dom}(A_\omega)$ whose dissimilarity is strictly greater than 1 for D^{ω}_{u} and not greater than 1 for D^{ω}_{l}. Denote by $\mathcal{R}$ the set of all possible collections of representers. Then, define the map $\bar{\eta} : \mathcal{R} \to \mathcal{R}$ such that for every collection $\Omega \in \mathcal{R}$ and every representer $\omega = (X_\omega, A_\omega) \in \Omega$, the mapped collection $\bar{\eta}(\Omega)$ contains $\bar{\eta}(\omega) = (X_\omega, A_{\bar{\eta}(\omega)})$ where

$$A_{\bar{\eta}(\omega)} = \begin{cases} A_\omega & \text{on} \quad D^{\omega}_{l}, \\ 1 & \text{on} \quad D^{\omega}_{u}. \end{cases} \tag{A.38}$$

Note that the representer $\bar{\eta}(\omega)$ is obtained by truncating at 1 every positive dissimilarity in ω which is greater than 1. In the following claim, we state that the method represented by an arbitrary collection $\Omega \in \mathcal{R}$ and by its image under the map $\bar{\eta}$ are equivalent.

Claim 11 *Given an arbitrary collection $\Omega \in \mathcal{R}$, the methods represented by Ω and $\bar{\eta}(\Omega)$ are equivalent, i.e., $\mathcal{H}^{\Omega} \equiv \mathcal{H}^{\bar{\eta}(\Omega)}$.*

Proof: First note that, by definition (A.38), for every representer $\omega \in \Omega$ we have that $\mathrm{dom}(A_{\bar{\eta}(\omega)}) = \mathrm{dom}(A_\omega) =: D^{\omega}$ and

$$A_\omega \geq A_{\bar{\eta}(\omega)}, \tag{A.39}$$

in D^{ω}. To prove the claim, we need to show that

$$u^{\Omega}_{X}(x, x') = u^{\bar{\eta}(\Omega)}_{X}(x, x'), \tag{A.40}$$

for every $x, x' \in X$ in an arbitrary network $N = (X, A_X)$. To do this, we first show that

$$u^{\Omega}_{X}(x, x') \leq u^{\bar{\eta}(\Omega)}_{X}(x, x'). \tag{A.41}$$

Given definition (5.24), it is immediate that if we show that

$$\lambda_X^{\Omega}(x,x') \leq \lambda_X^{\bar{\eta}(\Omega)}(x,x') \tag{A.42}$$

for every $x, x' \in X$ in an arbitrary network $N = (X, A_X)$, then (A.41) is true. For any network $N = (X, A_X)$ and arbitrary nodes $x, x' \in X$, pick any map $\phi^\omega \in \Phi^\omega_{x,x'}$ among the maps that go from $\omega \in \Omega$ to N having x and x' in its image [cf. (A.36)]. Then,

$$L(\phi^\omega;\bar{\eta}(\omega), N) = \max_{\substack{(z,z')\in D^\omega \\ z\neq z'}} \frac{A_X(\phi^\omega(z), \phi^\omega(z'))}{A_{\bar{\eta}(\omega)}(z,z')} \geq \max_{\substack{(z,z')\in D^\omega \\ z\neq z'}} \frac{A_X(\phi^\omega(z), \phi^\omega(z'))}{A_\omega(z,z')} = L(\phi^\omega;\omega, N), \tag{A.43}$$

where the inequality is implied by (A.39). Suppose a given map $\phi_*^{\bar{\eta}(\omega)}$ is an optimal map from the representer $\bar{\eta}(\omega)$ to the network N with x and x' in its image, that is,

$$\phi_*^{\bar{\eta}(\omega)} \in \operatorname*{argmin}_{\phi\,\in\,\Phi^{\bar{\eta}(\omega)}_{x,x'}} L(\phi;\bar{\eta}(\omega), N). \tag{A.44}$$

From the definition of optimal multiples as the minimal Lipschitz constants (5.20), we can write that

$$\lambda_X^{\bar{\eta}(\omega)}(x,x') = L(\phi_*^{\bar{\eta}(\omega)};\bar{\eta}(\omega), N) \geq L(\phi_*^{\bar{\eta}(\omega)};\omega, N) \geq \min_{\phi\,\in\,\Phi^{\omega}_{x,x'}} L(\phi;\omega, N) = \lambda_X^\omega(x,x'), \tag{A.45}$$

where we used (A.43) in the first inequality. Since (A.45) is true for every $\omega \in \Omega$, (5.23) implies (A.42), which in turn implies (A.41), as we wanted.

We now show the inequality opposite to (A.41),

$$u_X^{\Omega}(x,x') \geq u_X^{\bar{\eta}(\Omega)}(x,x'), \tag{A.46}$$

for all $x, x' \in X$ in an arbitrary network $N_X = (X, A_X)$. To prove this, we assume that for some x, x' such that $x \neq x'$,

$$\lambda = u_X^{\Omega}(x,x') < u_X^{\bar{\eta}(\Omega)}(x,x') =: \lambda', \tag{A.47}$$

and derive a contradiction. From the definition of $u_X^\Omega(x,x')$ in (5.24), there must exist a path $P_{xx'} = [x = x_0, x_1, \ldots, x_l = x']$ and an ordered set of representers $C(\Omega) = [\omega_0, \omega_1, \ldots, \omega_{l-1}]$ with $\omega_i \in \Omega$ for all i such that

$$\lambda_X^{\omega_i}(x_i, x_{i+1}) \leq \lambda, \tag{A.48}$$

for all i. This is equivalent to having a series of maps $\phi_i : X_{\omega_i} \to X$ such that $x_i, x_{i+1} \in \mathrm{Im}(\phi_i)$ and $L(\phi_i; \omega_i, N) \leq \lambda$ for all i.

Define the collection $\Phi_{P_{xx'}}^{C(\Omega)} \subseteq \Phi_{x_0,x_1}^{\omega_0} \times \Phi_{x_1,x_2}^{\omega_1} \times \ldots \times \Phi_{x_{l-1},x_l}^{\omega_{l-1}}$ of vectors of maps from representers ω_i with consecutive pair of nodes of $P_{xx'}$ in their image such that the maximum Lipschitz constant does not exceed λ. Formally,

$$\Phi_{P_{xx'}}^{C(\Omega)} = \{(\phi_0, \phi_1, \ldots, \phi_{l-1}) \subseteq \Phi_{x_0,x_1}^{\omega_0} \times \ldots \times \Phi_{x_{l-1},x_l}^{\omega_{l-1}} \quad \text{s.t. } L(\phi_i; \omega_i, N) \leq \lambda \text{ for all } i\}. \tag{A.49}$$

Given a minimizing path $P_{xx'}$ and the ordered set of representers $C(\Omega)$, we use the notation $\vec{\phi}$ to refer to an element of $\Phi_{P_{xx'}}^{C(\Omega)}$. Find the set of maps that minimize the largest dissimilarity in their image and denote this dissimilarity by δ. Precisely,

$$\delta = \inf_{\{P_{xx'}, C(\Omega)\}} \min_{\vec{\phi} \in \Phi_{P_{xx'}}^{C(\Omega)}} \max_{\phi_i \in \vec{\phi}} \max_{(z,z') \in D^{\omega_i}} A_X(\phi_i(z), \phi_i(z')). \tag{A.50}$$

Given $\beta > 0$, pick a minimizing path $P'_{xx'}$ and an ordered set of maps $C'(\Omega)$ such that there exists $\vec{\phi}'_\beta \in \Phi_{P'_{xx'}}^{C'(\Omega)}$ a vector of maps achieving a value not more than β away from the infimum in (A.50), i.e.,

$$\max_{\phi_i \in \vec{\phi}'_\beta} \max_{(z,z') \in D^{\omega_i}} A_X(\phi_i(z), \phi_i(z')) \leq \delta + \beta, \tag{A.51}$$

Notice that a vector $\vec{\phi}'_\beta$ can be found for *any* $\beta > 0$. Recall the partition of the domain D^ω into two disjoint sets D_u^ω and D_l^ω in (A.37) for every representer $\omega \in \Omega$.

We want to show that $\delta > \lambda$. In order to show this, we define

$$\delta_u^\beta = \max_{\phi_i \in \vec{\phi}'_\beta} \max_{(z,z') \in D_u^{\omega_i}} A_X(\phi_i(z), \phi_i(z')), \tag{A.52}$$

which, since the maximization domain for δ_u^β in (A.52) is smaller than that in (A.51), we must have that $\delta_u^\beta \leq \delta + \beta$.

Since we assumed that $\lambda' > \lambda$ [cf. (A.47)], if we show that $\lambda' \leq \max(\delta_u^\beta, \lambda)$ then we would have that $\lambda' \leq \delta_u^\beta$ and $\lambda < \delta_u^\beta$. Since $\delta_u^\beta \leq \delta + \beta$, we would have that $\lambda < \delta_u^\beta \leq \delta + \beta$. However, since this is true for arbitrary $\beta > 0$, we would have that $\delta > \lambda$, as wanted. Hence, we show that $\lambda' \leq \max(\delta_u^\beta, \lambda)$ as follows. By definition (5.24),

$$\lambda' \leq \max_{i | x_i \in P'_{xx'}} \lambda_X^{\bar{\eta}(\Omega)}(x_i, x_{i+1}), \tag{A.53}$$

where the inequality follows from the fact that we are looking at one particular path instead of minimizing over all possible paths. By writing the multiples $\lambda_X^{\bar{\eta}(\Omega)}(x_i, x_{i+1})$ in terms of Lipschitz constants as in (5.20) and combining this with (5.23), we obtain

$$\lambda' \leq \max_{i|x_i \in P'_{xx'}} \inf_{\omega \in \Omega} \min_{\phi \in \Phi^{\omega}_{x_i, x_{i+1}}} \max_{\substack{(z,z') \in D^{\omega} \\ z \neq z'}} \frac{A_X(\phi(z), \phi(z'))}{A_{\bar{\eta}(\omega)}(z, z')}. \tag{A.54}$$

We may rewrite (A.54) by splitting the domain D^{ω} into D^{ω}_u and D^{ω}_l, to obtain

$$\lambda' \leq \max_{i|x_i \in P'_{xx'}} \inf_{\omega \in \Omega} \min_{\phi \in \Phi^{\omega}_{x_i, x_{i+1}}} \max \left(\max_{\substack{(z,z') \in D^{\omega}_u \\ z \neq z'}} \frac{A_X(\phi(z), \phi(z'))}{A_{\bar{\eta}(\omega)}(z, z')}, \max_{\substack{(z,z') \in D^{\omega}_l \\ z \neq z'}} \frac{A_X(\phi(z), \phi(z'))}{A_{\bar{\eta}(\omega)}(z, z')} \right). \tag{A.55}$$

Notice that from definition (A.38), within domain D^{ω}_u we have that $A_{\bar{\eta}(\omega)} = 1$, and within domain D^{ω}_l we have that $A_{\bar{\eta}(\omega)} \equiv A_{\omega}$. With this, we rewrite (A.55) and obtain

$$\lambda' \leq \max_{i|x_i \in P'_{xx'}} \inf_{\omega \in \Omega} \min_{\phi \in \Phi^{\omega}_{x_i, x_{i+1}}} \max \left(\max_{(z,z') \in D^{\omega}_u} A_X(\phi(z), \phi(z')), \max_{\substack{(z,z') \in D^{\omega}_l \\ z \neq z'}} \frac{A_X(\phi(z), \phi(z'))}{A_{\omega}(z, z')} \right). \tag{A.56}$$

The second argument within the maximization in (A.56) is not greater than $L(\phi; \omega, N)$, since in the definition of the Lipschitz constant [cf. (5.19)] the maximum is computed over the whole domain D^{ω} whereas in (A.56), the maximum is taken over $D^{\omega}_l \subseteq D^{\omega}$. This implies that

$$\lambda' \leq \max_{i|x_i \in P'_{xx'}} \inf_{\omega \in \Omega} \min_{\phi \in \Phi^{\omega}_{x_i, x_{i+1}}} \max \Big(\max_{(z,z') \in D^{\omega}_u} A_X(\phi(z), \phi(z')) \ , \ L(\phi; \omega, N) \Big). \tag{A.57}$$

If instead of minimizing over all possible maps joining consecutive nodes in $P'_{xx'}$, we only consider the maps contained in $\vec{\phi}'_{\beta}$, we decrease the domain of minimization. Thus, the resulting value is an upper bound of the minimization. That is,

$$\begin{aligned} \lambda' &\leq \max_{\phi_i \in \vec{\phi}'_{\beta}} \max \Big(\max_{(z,z') \in D^{\omega_i}_u} A_X(\phi_i(z), \phi_i(z')) \ , \ L(\phi_i; \omega_i, N) \Big) \\ &= \max \Big(\max_{\phi_i \in \vec{\phi}'_{\beta}} \max_{(z,z') \in D^{\omega_i}_u} A_X(\phi_i(z), \phi_i(z')) \ , \ \max_{\phi_i \in \vec{\phi}'_{\beta}} L(\phi_i; \omega_i, N) \Big). \end{aligned} \tag{A.58}$$

We can now see that the first argument is exactly δ^{β}_u [cf. (A.52)] and the second one is not greater than λ since $\vec{\phi}'_{\beta} \in \Phi^{C'(\Omega)}_{P'_{xx'}}$ [cf. (A.49)]. Consequently, $\lambda' \leq \max(\delta^{\beta}_u, \lambda)$, implying that $\delta > \lambda$ as explained in the paragraph preceding (A.53).

Let $\epsilon = (\delta - \lambda)/3 > 0$, so that we are ensured that $\lambda < \delta - \epsilon$. Define the function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ as depicted in Fig. A.1, where M is an upper bound of all the dissimilarities

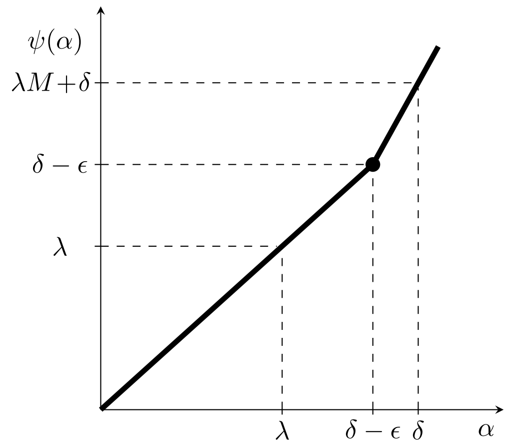


Figure A.1: Nondecreasing function ψ that transforms network N into $\psi(N)$. The parameter ϵ is picked small enough to ensure that $\lambda < \delta - \epsilon$. M is an upper bound on the maximum dissimilarity found in all representers $\omega \in \Omega$.

in every representer $\omega \in \Omega$, that is

$$M \geq \sup_{\omega \in \Omega} \max_{(z,z') \in D^\omega} A_\omega(z, z'). \tag{A.59}$$

The Property of Representability (P2) ensures the existence of such constant M. From Fig. A.1, it is clear that

$$\psi \circ A_X \geq A_X. \tag{A.60}$$

Denote by $\Theta^{C(\Omega)}_{P_{xx'}}$ the collection of vectors of optimal maps from a set of representers $C(\Omega)$ to the transformed network $\psi(N)$ in the same way that $\Phi^{C(\Omega)}_{P_{xx'}}$ was defined for the original network in (A.49). We now show that every vector of optimal maps in the transformed network $\psi(N)$ is also optimal in the original network N. Equivalently, we want to show that $\Theta^{C(\Omega)}_{P_{xx'}} \subseteq \Phi^{C(\Omega)}_{P_{xx'}}$ for all minimizing paths $P_{xx'}$ and ordered sets of representers $C(\Omega)$. To see this, from Fig. A.1 and the scale preservation property of the clustering method $\mathcal{H}^\Omega$, we may write that

$$\lambda = \psi(\lambda) = \psi(u^\Omega_X(x, x')) = u^\Omega_{\psi(X)}(x, x') = \max_{\phi_i \in \vec{\phi}} L(\phi_i; \omega_i, \psi(N)), \tag{A.61}$$

for every vector $\vec{\phi} \in \Theta^{C(\Omega)}_{P_{xx'}}$ for minimizing $P_{xx'}$ and $C(\Omega)$ where $\mathcal{H}^\Omega(\psi(N)) = (X, u^\Omega_{\psi(X)})$.

Using the definition of the Lipschitz constant and (A.60), we obtain that

$$\max_{\phi_i \in \vec{\phi}} L(\phi_i; \omega_i, \psi(N)) = \max_{\phi_i \in \vec{\phi}} \max_{\substack{(z,z') \in D^{\omega_i} \\ z \neq z'}} \frac{\psi\big(A_X(\phi_i(z), \phi_i(z'))\big)}{A_\omega(z,z')} \tag{A.62}$$

$$\geq \max_{\phi_i \in \vec{\phi}} \max_{\substack{(z,z') \in D^{\omega_i} \\ z \neq z'}} \frac{A_X(\phi_i(z), \phi_i(z'))}{A_\omega(z,z')} = \max_{\phi_i \in \vec{\phi}} L(\phi_i; \omega_i, N) \geq \min_{P_{xx'}} \max_{i | x_i \in P_{xx'}} \lambda_X^{\Omega}(x_i, x_{i+1}) = \lambda.$$

By concatenating the expressions (A.61) and (A.62), we see that the inequalities stated must be equalities. Thus, for all $\vec{\phi} \in \Theta_{P_{xx'}}^{C(\Omega)}$ we have that $\lambda = \max_{\phi_i \in \vec{\phi}} L(\phi_i; \omega_i, N)$. This means that $\vec{\phi} \in \Phi_{P_{xx'}}^{C(\Omega)}$, implying that $\Theta_{P_{xx'}}^{C(\Omega)} \subseteq \Phi_{P_{xx'}}^{C(\Omega)}$, as desired.

Given a minimizing path $P_{xx'}$ and set $C(\Omega)$, for all $\vec{\phi} \in \Theta_{P_{xx'}}^{C(\Omega)}$ it holds that

$$\lambda = \max_{\phi_i \in \vec{\phi}} \max_{\substack{(z,z') \in D^{\omega_i} \\ z \neq z'}} \frac{\psi\big(A_X(\phi_i(z), \phi_i(z'))\big)}{A_\omega(z,z')} \geq \frac{1}{M}\, \psi\Big(\max_{\phi_i \in \vec{\phi}} \max_{(z,z') \in D^{\omega_i}} A_X(\phi_i(z), \phi_i(z'))\Big), \tag{A.63}$$

where the first equality comes from (A.61) and we used the definition of M [cf. (A.59)] and the fact that ψ is a nondecreasing function for the inequality. Since (A.63) is true for all $\vec{\phi} \in \Theta_{P_{xx'}}^{C(\Omega)}$ and it is true for any minimizing $P_{xx'}$ and $C(\Omega)$, we can minimize over all paths $P_{xx'}$, over all series of maps $C(\Omega)$ and over all vectors in $\Theta_{P_{xx'}}^{C(\Omega)}$ while preserving the inequality, that is

$$\lambda M \geq \psi\Big(\inf_{\{P_{xx'}, C(\Omega)\}} \min_{\vec{\phi} \in \Theta_{P_{xx'}}^{C(\Omega)}} \max_{\phi_i \in \vec{\phi}} \max_{(z,z') \in D^{\omega_i}} A_X(\phi_i(z), \phi_i(z'))\Big), \tag{A.64}$$

and, since $\Theta_{P_{xx'}}^{C(\Omega)} \subseteq \Phi_{P_{xx'}}^{C(\Omega)}$, we have that

$$\lambda M \geq \psi\Big(\inf_{\{P_{xx'}, C(\Omega)\}} \min_{\vec{\phi} \in \Phi_{P_{xx'}}^{C(\Omega)}} \max_{\phi_i \in \vec{\phi}} \max_{(z,z') \in D^{\omega_i}} A_X(\phi_i(z), \phi_i(z'))\Big). \tag{A.65}$$

Comparing (A.65) with the definition of δ in (A.50), we obtain that

$$\lambda M \geq \psi(\delta) = \lambda M + \delta, \tag{A.66}$$

which is a contradiction since we showed that $\delta > \lambda \geq 0$. This implies that (A.46) must be true which, combined with inequality (A.41), proves equality (A.40) and the proof of Claim 11 is completed. ■

Claim 11 shows that, given an arbitrary collection of representers, if we alter this collection by truncating at 1 the dissimilarities of all the representers in the collection, the represented method does not change. We now show that, given a collection of representers with no dissimilarity greater than 1, we can generate a different collection where all its positive dissimilarities are exactly 1 and the represented method remains unchanged.

Let $\omega = (X_\omega, A_\omega)$ be a representer such that $A_\omega(z, z') \leq 1$ for all $(z, z') \in \mathrm{dom}(A_\omega) =: D^\omega$. Define the equivalence relation $\sim_\omega$ on X_ω where $z, z' \in X_\omega$ are such that $z \sim_\omega z'$ if and only if there exists a path $P_{zz'} = [z = z_0, z_1, \ldots, z_l = z']$ such that $A_\omega(z_i, z_{i+1}) < 1$ or $A_\omega(z_{i+1}, z_i) < 1$ for all $i = 0, \ldots, l-1$. That is, $z \sim_\omega z'$ if and only if they can be linked with a path such that for every hop the dissimilarity in at least one direction is strictly smaller than 1. Define the quotient node space

$$X_{\tilde{\omega}} = X_\omega \backslash \sim_\omega \tag{A.67}$$

and let $\alpha^\omega : X_\omega \to X_{\tilde{\omega}}$ be the surjective map that sends every node z to its equivalence class under $\sim_\omega$. Further, define the domain

$$\tilde{D}^\omega = \{(\bar{z}, \bar{z}') \mid \exists\, (z, z') \in D^\omega \quad \text{s.t.}\ \ \alpha^\omega(z) = \bar{z} \ \text{ and } \ \alpha^\omega(z') = \bar{z}'\}, \tag{A.68}$$

and define the dissimilarity function $A_{\tilde{\omega}}$

$$A_{\tilde{\omega}}(\bar{z}, \bar{z}') = 1, \tag{A.69}$$

for all distinct $(\bar{z}, \bar{z}') \in \tilde{D}^\omega$ and $A_{\tilde{\omega}}(\bar{z}, \bar{z}) = 0$ for all $\bar{z} \in X_{\tilde{\omega}}$. By construction, every positive dissimilarity in $\tilde{\omega}$ is exactly 1. Notice that α^ω is dissimilarity reducing by construction. Define the map $\tilde{\eta} : \mathcal{R} \to \mathcal{R}$ such that for every representer $\omega = (X_\omega, A_\omega) \in \Omega$, the mapped collection $\tilde{\eta}(\Omega)$ contains $\tilde{\eta}(\omega) = (X_{\tilde{\omega}}, A_{\tilde{\omega}})$ as defined in (A.67) to (A.69). In the following claim, we state that the method represented by a collection $\Omega \in \mathcal{R}$ with dissimilarities not greater than 1 and by its image under the map $\tilde{\eta}$ are equivalent.

Claim 12 *Given a collection $\Omega \in \mathcal{R}$ where every $\omega \in \Omega$ has dissimilarities upper bounded by 1, the methods represented by Ω and $\tilde{\eta}(\Omega)$ are equivalent, i.e., $\mathcal{H}^\Omega \equiv \mathcal{H}^{\tilde{\eta}(\Omega)}$.*

Proof: Let $N = (X, A_X)$ be an arbitrary network and pick two distinct nodes $x \neq x' \in X$. Denote by $\lambda = u_X^\Omega(x, x')$ the ultrametric value between x and x' given by the method $\mathcal{H}^\Omega$. Further, denote by $\Phi_{P_{xx'}}^{C(\Omega)}$ the collection of vectors of maps from representers $\omega_i \in C(\Omega)$ with consecutive nodes of $P_{xx'}$ in their image such that the maximum multiple does not exceed λ as defined in (A.49). For arbitrary minimizing $P_{xx'}$ and $C(\Omega)$, for each $\vec{\phi} \in \Phi_{P_{xx'}}^{C(\Omega)}$ we define $\delta_{\vec{\phi}}$ as the maximum dissimilarity in N to which some dissimilarity in some representer in $C(\Omega)$ is mapped by $\phi_i \in \vec{\phi}$. That is,

$$\delta_{\vec{\phi}} = \max_{\phi_i \in \vec{\phi}} \ \max_{(z,z') \in D^{\omega_i}} A_X(\phi_i(z), \phi_i(z')). \tag{A.70}$$

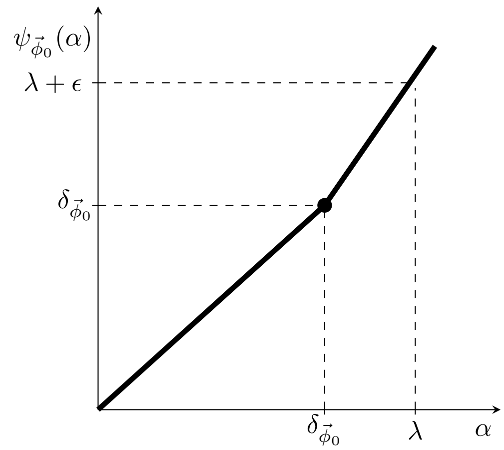


Figure A.2: Nondecreasing function $\psi_{\vec{\phi}_0}$ that transforms network N into $\psi_{\vec{\phi}_0}(N)$. The parameter ϵ is an arbitrary positive real.

First, observe that $\lambda \geq \delta_{\vec{\phi}}$ for all $\vec{\phi} \in \Phi_{P_{xx'}}^{C(\Omega)}$. To see this, from the definition of the Lipschitz constant (5.19) and the fact that no dissimilarity in $\omega_i \in \Omega$ is greater than 1 we may write

$$\lambda = \max_{\phi_i \in \vec{\phi}} L(\phi_i; \omega_i, N) = \max_{\phi_i \in \vec{\phi}} \max_{\substack{(z,z') \in D^{\omega_i} \\ z \neq z'}} \frac{A_X(\phi_i(z), \phi_i(z'))}{A_{\omega_i}(z,z')}$$
$$\geq \max_{\phi_i \in \vec{\phi}} \max_{(z,z') \in D^{\omega_i}} A_X(\phi_i(z), \phi_i(z')) = \delta_{\vec{\phi}}. \tag{A.71}$$

More importantly, we can show that $\lambda = \delta_{\vec{\phi}}$ for all $\vec{\phi} \in \Phi_{P_{xx'}}^{C(\Omega)}$. We show this by assuming that it is not true and then deriving a contradiction. Suppose that $\lambda \neq \delta_{\vec{\phi}_0}$ for some $\vec{\phi}_0 \in \Phi_{P_{xx'}}^{C(\Omega)}$ then, by (A.71) it must be that $\lambda > \delta_{\vec{\phi}_0}$. Pick any positive constant $\epsilon > 0$ and consider the nondecreasing function $\psi_{\vec{\phi}_0}$ depicted in Fig. A.2. If we denote by $(X, u^{\Omega}_{\psi_{\vec{\phi}_0}(X)})$ the ultrametric obtained by applying the method $\mathcal{H}^{\Omega}$ to the transformed network $\psi_{\vec{\phi}_0}(N)$, then by scale preservation of the $\mathcal{H}^{\Omega}$ clustering method we may write that

$$u^{\Omega}_{\psi_{\vec{\phi}_0}(X)}(x,x') = \psi_{\vec{\phi}_0}(u^{\Omega}_X(x,x')) = \psi_{\vec{\phi}_0}(\lambda) = \lambda + \epsilon > \lambda. \tag{A.72}$$

Since $\vec{\phi}_0 \in \Phi_{P_{xx'}}^{C(\Omega)}$, we have that $L(\phi_i; \omega_i, N) \leq \lambda$ for all $\phi_i \in \vec{\phi}_0$ [cf. (A.49)], which implies that

$$\lambda A_{\omega_i}(z,z') \geq A_X(\phi_i(z), \phi_i(z')), \tag{A.73}$$

for all $\phi_i \in \vec{\phi}_0$ and for all $(z,z') \in D^{\omega_i}$. But since $A_X(\phi_i(z), \phi_i(z')) \leq \delta_{\vec{\phi}_0}$ [cf. (A.70)], from

Fig. A.2 it follows that

$$\psi_{\vec{\phi}_0}(A_X(\phi_i(z), \phi_i(z'))) = A_X(\phi_i(z), \phi_i(z')), \tag{A.74}$$

for all $\phi_i \in \vec{\phi}_0$ and for all $(z, z') \in D^{\omega_i}$. Substituting (A.74) in (A.73) we have that

$$\lambda A_{\omega_i}(z, z') \geq \psi_{\vec{\phi}_0}(A_X(\phi_i(z), \phi_i(z'))), \tag{A.75}$$

for all $\phi_i \in \vec{\phi}_0$ and for all $(z, z') \in D^{\omega_i}$. Since the maps $\phi_i \in \vec{\phi}_0$ contain x and x' in the union of their images, it follows that the multiple λ is an upper bound for the ultrametric value between these nodes in the transformed network $\psi_{\vec{\phi}_0}(N)$. That is,

$$\lambda \geq u^{\Omega}_{\psi_{\vec{\phi}_0(X)}}(x, x'). \tag{A.76}$$

However, (A.72) and (A.76) contradict each other, implying that $\lambda = \delta_{\vec{\phi}}$ for all $\vec{\phi} \in \Phi^{C(\Omega)}_{P_{xx'}}$ as we wanted to show. Also, this shows that $\delta_{\vec{\phi}}$ does not depend on the ordered set of maps $\vec{\phi} \in \Phi^{C(\Omega)}_{P_{xx'}}$ chosen, i.e. $\delta = \delta_{\vec{\phi}}$ for any $\vec{\phi} \in \Phi^{C(\Omega)}_{P_{xx'}}$ and we have that

$$\delta = \lambda. \tag{A.77}$$

Given arbitrary $P_{xx'}$ and $C(\Omega)$, consider any vector of maps $\vec{\phi} \in \Phi^{C(\Omega)}_{P_{xx'}}$ and define the image network $N_{\vec{\phi}} = (X_{\vec{\phi}}, A_{X_{\vec{\phi}}})$ where

$$X_{\vec{\phi}} = \bigcup_{\phi \in \vec{\phi}} \operatorname{Im}(\phi), \tag{A.78}$$

and $A_{X_{\vec{\phi}}} = A_X|_{X_{\vec{\phi}} \times X_{\vec{\phi}}}$. Notice that, by construction, $x, x' \in X_{\vec{\phi}}$. Thus, since $\vec{\phi} \in \Phi^{C(\Omega)}_{P_{xx'}}$ we can assert that

$$u^{\Omega}_{X_{\vec{\phi}}}(x, x') \leq \lambda, \tag{A.79}$$

where $\mathcal{H}^{\Omega}(X_{\vec{\phi}}, A_{X_{\vec{\phi}}}) = (X_{\vec{\phi}}, u^{\Omega}_{X_{\vec{\phi}}})$. Moreover, since the network $N_{\vec{\phi}}$ can be embedded into N, axiom (A2) implies that $u^{\Omega}_{X_{\vec{\phi}}}(x, x') \geq \lambda$, which, combined with (A.79) ensures that

$$u^{\Omega}_{X_{\vec{\phi}}}(x, x') = \lambda, \tag{A.80}$$

i.e. that ultrametric value between nodes x and x' in the image network $N_{\vec{\phi}}$ is the same as the ultrametric value in the network N. Consider the nondecreasing function ψ' depicted in Fig. A.3. If we denote by $(X, u^{\Omega}_{\psi'(X_{\vec{\phi}})})$ the ultrametric obtained by applying the method

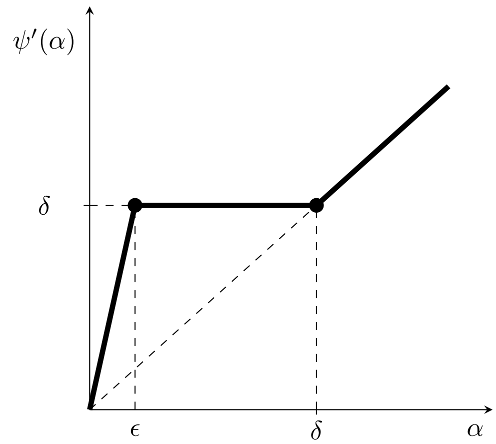


Figure A.3: Nondecreasing function ψ' that transforms network $N_{\vec{\phi}}$ into $\psi'(N_{\vec{\phi}})$. The parameter ϵ is chosen small enough to ensure that every positive dissimilarity in $N_{\vec{\phi}}$ is greater than ϵ.

$\mathcal{H}^\Omega$ to the transformed network $\psi'(N_{\vec{\phi}})$, then by scale preservation of the clustering method $\mathcal{H}^\Omega$, it follows that

$$u^\Omega_{\psi'(X_{\vec{\phi}})}(x, x') = \psi'\big(u^\Omega_{X_{\vec{\phi}}}(x, x')\big) = \psi'(\lambda) = \psi'(\delta) = \delta, \tag{A.81}$$

where we used equalities (A.80) and (A.77).

Given a minimizing path $P_{xx'}$ and an ordered set of representers $C(\Omega) = [\omega_0, \ldots, \omega_{l-1}]$, let $\vec{\theta}$ be a vector of maps into the transformed network $\psi'(N_{\vec{\phi}})$ achieving the ultrametric value λ. This implies that, for all maps $\theta_i \in \vec{\theta}$,

$$L(\theta_i; \omega_i, \psi'(N_{\vec{\phi}})) \leq \lambda = \delta. \tag{A.82}$$

Notice that $\psi'(\alpha) \geq \alpha$ for any real $\alpha \geq 0$. Thus, we may write,

$$\begin{aligned} \lambda &= \psi'(\lambda) = \psi'(u^\Omega_{X_{\vec{\phi}}}(x, x')) = u^\Omega_{\psi'(X_{\vec{\phi}})}(x, x') \\ &= \max_{\theta_i \in \vec{\theta}'} L(\theta_i; \omega_i, \psi'(N_{\vec{\phi}})) \geq \max_{\theta_i \in \vec{\theta}'} L(\theta_i; \omega_i, N_{\vec{\phi}}) \geq u^\Omega_{X_{\vec{\phi}}}(x, x') = \lambda. \end{aligned} \tag{A.83}$$

The fact that λ appears in both sides of the inequalities in (A.83), forces them to be equalities, meaning that $\vec{\theta}$ is a vector of optimal maps for the original image network $N_{\vec{\phi}}$ as well and, as discussed in the sentence following (A.80), $\vec{\theta}$ is optimal for the original network N too. More precisely, $\vec{\theta} \in \Phi^{C(\Omega)}_{P_{xx'}}$. Moreover, for every map $\theta_i \in \vec{\theta}$ it must be true that for all $(z, z') \in D^{\omega_i}$, it holds that

$$A_{\omega_i}(z, z') < 1 \quad \Rightarrow \quad \theta_i(z) = \theta_i(z'). \tag{A.84}$$

Otherwise, if $\theta_i(z_0) \neq \theta_i(z_0')$ with $0 < A_{\omega_i}(z_0, z_0') < 1$, then

$$\lambda \geq \frac{\psi'\big(A_X(\theta_i(z_0), \theta_i(z_0'))\big)}{A_{\omega_i}(z_0, z_0')} > \delta, \tag{A.85}$$

where the left inequality comes from (A.82) and the right inequality comes from the fact that the denominator is less than 1 and the numerator is at least δ from the choice of ϵ in the function ψ'. But (A.85) contradicts (A.77), thus, (A.84) must hold.

We now regard every $\theta_i \in \vec{\theta}$ as a map from ω_i to N and we show that we can construct a map $\nu_{\theta_i} : \tilde{\omega}_i \to N$ from the modified representer $\tilde{\omega}_i$ [cf. (A.67)-(A.69)] to the network N such that

$$\nu_{\theta_i} \circ \alpha^{\omega_i} = \theta_i, \tag{A.86}$$

$$L(\theta_i; \omega_i, N) = L(\nu_{\theta_i}; \tilde{\omega}_i, N). \tag{A.87}$$

Once we show the above equalities, notice that (A.87) implies that the map ν_{θ_i} achieves a Lipschitz constant equal to that of the optimal map θ_i. Since a better vector of maps from $\tilde{\omega}_i$ to N could exist, this implies that

$$u_X^{\Omega}(x, x') \geq u_X^{\tilde{\eta}(\Omega)}(x, x'). \tag{A.88}$$

Conversely, since the map $\alpha^{\omega_i} : \omega_i \to \tilde{\omega}_i$ is dissimilarity reducing, given any map ϕ from $\tilde{\omega}_i$ to N, the composed map $\phi \circ \alpha^{\omega_i}$ from ω_i to N has a Lipschitz constant not greater than ϕ, to see this,

$$\begin{aligned}
L(\phi \circ \alpha^{\omega_i}; \omega_i, N) &= \max_{\substack{(z,z') \in D^{\omega_i} \\ z \neq z'}} \frac{A_X(\phi \circ \alpha^{\omega_i}(z), \phi \circ \alpha^{\omega_i}(z'))}{A_{\omega_i}(z, z')} \\
&= \max_{\substack{(z,z') \in D^{\omega_i} \\ \alpha^{\omega_i}(z) \neq \alpha^{\omega_i}(z')}} \frac{A_X(\phi \circ \alpha^{\omega_i}(z), \phi \circ \alpha^{\omega_i}(z'))}{A_{\omega_i}(z, z')} \leq \max_{\substack{(z,z') \in D^{\omega_i} \\ \alpha^{\omega_i}(z) \neq \alpha^{\omega_i}(z')}} \frac{A_X(\phi \circ \alpha^{\omega_i}(z), \phi \circ \alpha^{\omega_i}(z'))}{A_{\tilde{\omega}_i}(\alpha^{\omega_i}(z), \alpha^{\omega_i}(z'))} \\
&= \max_{\substack{(\bar{z},\bar{z}') \in \tilde{D}^{\omega_i} \\ \bar{z} \neq \bar{z}'}} \frac{A_X(\phi(\bar{z}), \phi(\bar{z}'))}{A_{\tilde{\omega}_i}(\bar{z}, \bar{z}')} = L(\phi; \tilde{\omega}_i, N),
\end{aligned} \tag{A.89}$$

where the inequality is implied by the fact that α^{ω_i} is a dissimilarity reducing map. Moreover, since α^{ω_i} is surjective, the images of ϕ and $\phi \circ \alpha^{\omega_i}$ are the same, implying that

$$u_X^{\Omega}(x, x') \leq u_X^{\tilde{\eta}(\Omega)}(x, x'). \tag{A.90}$$

Inequalities (A.88) and (A.90) imply that

$$u_X^{\Omega}(x,x') = u_X^{\tilde{\eta}(\Omega)}(x,x'), \tag{A.91}$$

showing the desired equivalence between the clustering methods $\mathcal{H}^{\Omega}$ and $\mathcal{H}^{\tilde{\eta}(\Omega)}$.

Hence, once we construct a map ν_{θ_i} satisfying (A.86) and (A.87) we conclude the proof. Define $\nu_{\theta_i} : \tilde{\omega}_i \to N$ as

$$\nu_{\theta_i}(\bar{z}) = \theta_i(z), \tag{A.92}$$

for some z where $\bar{z} = \alpha^{\omega_i}(z)$ [cf. (A.68)]. Such z always exists since α^{ω_i} is surjective. Furthermore, this definition does not depend on the choice of z. Indeed, if $z' \neq z$ is another element in X_{ω_i} such that $\alpha^{\omega_i}(z') = \bar{z} = \alpha^{\omega_i}(z)$, then there exists a path $P_{zz'} = [z = z_0, z_1, \ldots, z_l = z']$ in X_{ω_i} such that $A_{\omega_i}(z_j, z_{j+1}) < 1$ or $A_{\omega_i}(z_{j+1}, z_j) < 1$ for all j. Thus, from (A.84), $\theta_i(z_j) = \theta_i(z_k)$ for all $j, k \in \{0, \ldots, l\}$, in particular, $\theta_i(z) = \theta_i(z')$. Consequently, our definition of $\nu_{\theta_i}(\bar{z})$ is independent of the choice of z in the preimage of $\bar{z}$ and $\nu_{\theta_i}(\alpha^{\omega_i}(z)) = \theta_i(z)$ for all $z \in X_{\omega_i}$ by construction, proving (A.86).

To show equality (A.87), note that from the definition of the Lipschitz constant (5.19), condition (A.84) and the decomposition (A.86) we may write that

$$\begin{aligned} L(\theta_i;\omega_i,N) &= \max_{\substack{(z,z')\in D^{\omega_i} \\ z\neq z'}} \frac{A_X(\theta_i(z),\theta_i(z'))}{A_{\omega_i}(z,z')} = \max_{\substack{(z,z')\in D^{\omega_i} \\ \theta_i(z)\neq\theta_i(z')}} \frac{A_X(\theta_i(z),\theta_i(z'))}{A_{\omega_i}(z,z')} \\ &= \max_{\substack{(z,z')\in D^{\omega_i} \\ \theta_i(z)\neq\theta_i(z')}} A_X(\theta_i(z),\theta_i(z')) = \max_{\substack{(z,z')\in D^{\omega_i} \\ \nu_{\theta_i}(\alpha^{\omega_i}(z))\neq\nu_{\theta_i}(\alpha^{\omega_i}(z'))}} A_X\big(\nu_{\theta_i}(\alpha^{\omega_i}(z)),\nu_{\theta_i}(\alpha^{\omega_i}(z'))\big). \end{aligned} \tag{A.93}$$

From (A.68) and the fact that α^{ω_i} is a surjective map, we may replace the generic $\alpha^{\omega_i}(z)$ in the last term of (A.93) by a generic $\bar{z}$ in $X_{\tilde{\omega}_i}$. Combining this with the fact that every positive dissimilarity in $\tilde{\omega}_i$ is exactly 1 we may write

$$L(\theta_i;\omega_i,N) = \max_{\substack{(\bar{z},\bar{z}')\in \tilde{D}^{\omega_i} \\ \nu_{\theta_i}(\bar{z})\neq\nu_{\theta_i}(\bar{z}')}} A_X(\nu_{\theta_i}(\bar{z}),\nu_{\theta_i}(\bar{z}')) = \max_{\substack{(\bar{z},\bar{z}')\in \tilde{D}^{\omega_i} \\ \bar{z}\neq\bar{z}'}} \frac{A_X(\nu_{\theta_i}(\bar{z}),\nu_{\theta_i}(\bar{z}'))}{A_{\tilde{\omega}_i}(\bar{z},\bar{z}')} = L(\nu_{\theta_i};\tilde{\omega}_i,N), \tag{A.94}$$

showing (A.87) and completing the proof of Claim 12. ■

Claim 12 ensures that given a method represented by a collection of representers with dissimilarities upper bounded by 1, we can collapse the representers in the collection in a way that the only remaining positive dissimilarities are exactly 1 and the corresponding represented method does not change.

By concatenating Claims 11 and 12, we have that given an arbitrary collection of representers, we may first truncate it and then collapse it such that the remaining positive dissimilarities in every representer within the collection are exactly 1 and the represented method does not change. Thus, every representer in the resulting collection is a structure representer. Finally, since the method remains unchanged, in particular it remains admissible, and the argument that concludes in (A.35) can be used to show that the representers must be strongly connected, concluding the proof of Theorem 11.

Bibliography

[1] F. Bach and M. Jordan, "Learning spectral clustering," *Advances Neural Info. Process. Syst. 16*, pp. 305–312, 2004.

[2] M. Barr and C. Wells, *Category Theory for Computing Science.* Prentice Hall, 1990.

[3] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer-Verlag, 2006.

[5] L. M. Blumenthal, *Distance Geometry: A Study of the Development of Abstract Metrics.* University of Missouri Studies, 1938, vol. 13.

[6] J. Boyd, "Asymmetric clusters of internal migration regions of France," *IEEE Trans. Syst., Man, Cybern.*, no. 2, pp. 101–104, 1980.

[7] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, and N. Zhang, "Topological structure analysis of the protein–protein interaction network in budding yeast," *Nucleic Acids Res.*, vol. 31, no. 9, pp. 2443–2450, 2003.

[8] T. Bui, S. Chaudhuri, F. Leighton, and M. Sipser, "Graph bisection algorithms with good average case behavior," *Combinatorica*, vol. 7, no. 2, pp. 171–191, 1987.

[9] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry.* American Mathematical Soc., 2001.

[10] G. Carlsson and F. Mémoli, "Characterization, stability and convergence of hierarchical clustering methods," *J. Mach. Learn. Res.*, vol. 11, pp. 1425–1470, 2010.

[11] ——, "Classifying clustering schemes," *Found. Comp. Math.*, vol. 13, no. 2, pp. 221–252, 2013.

[12] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Alternative axiomatic constructions for hierarchical clustering of asymmetric networks," in *IEEE Global Conf. Signal and Info. Process. (GlobalSIP)*, 2013, pp. 791–794.

[13] ——, "Axiomatic construction of hierarchical clustering in asymmetric networks," in *IEEE Intl. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, 2013, pp. 5219–5223.

[14] ——, "Hierarchical clustering methods and algorithms for asymmetric networks," in *Asilomar Conf. Signals, Systems and Computers*, 2013, pp. 1773–1777.

[15] ——, "Hierarchical quasi-clustering methods for asymmetric networks," *Intl. Conf. Mach. Learn. (ICML)*, vol. 32, no. 1, pp. 352–360, 2014.

[16] ——, "Admissible hierarchical clustering methods and algorithms for asymmetric networks," *arXiv preprint arXiv:1607.06335*, 2016.

[17] ——, "Excisive hierarchical clustering methods for network data," *arXiv preprint arXiv:1607.06339*, 2016.

[18] ——, "Hierarchical clustering of asymmetric networks," *arXiv preprint arXiv:1607.06294*, 2016.

[19] H. Caron et al., "The human transcriptome map: Clustering of highly expressed genes in chromosomal domains," *Science*, vol. 291, no. 5507, pp. 1289–1292, 2001.

[20] N. Christofides, "Worst-case analysis of a new heuristic for the travelling salesman problem," DTIC Document, Tech. Rep., 1976.

[21] F. Chung, *Spectral Graph Theory.* American Mathematical Soc., 1997.

[22] A. G. Constantine and J. C. Gower, "Graphical representation of asymmetric matrices," *J. Royal Stat. Soc. Ser. C (App. Stat.)*, vol. 27, no. 3, pp. 297–304, 1978.

[23] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling.* CRC Press, 2000.

[24] R. Duan and S. Pettie, "Fast algorithms for (max, min)-matrix multiplication and bottleneck shortest paths," in *Symp. Disc. Algo. (SODA)*, 2009, pp. 384–391.

[25] U. Feige and R. Krauthgamer, "A polylogarithmic approximation of the minimum bisection," *SIAM J. Comput.*, vol. 31, no. 4, pp. 1090–1118, 2002.

[26] W. Fernandez de la Vega, M. Karpinski, and C. Kenyon, "Approximation schemes for metric bisection and partitioning," in *Symp. Disc. Algo. (SODA)*, 2004, pp. 506–515.

[27] R. W. Floyd, "Algorithm 97: shortest path," *Commun. ACM*, vol. 5, no. 6, p. 345, 1962.

[28] M. Frechet, "Sur quelques points du calcul fonctionnel," *R. Circ. Mate. Palermo*, vol. 22, no. 1, pp. 1–72, 1906.

[29] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[30] H. Frigui and R. Krishnapuram, "A robust competitive clustering algorithm with applications in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 450–465, May 1999.

[31] H. N. Gabow, Z. Galil, T. Spencer, and R. E. Tarjan, "Efficient algorithms for finding minimum spanning trees in undirected and directed graphs," *Combinatorica*, vol. 6, no. 2, pp. 109–122, 1986.

[32] M. Gondran and M. Minoux, *Graphs, Dioids and Semi Rings: New Models and Algorithms*. Springer, 2008.

[33] V. Gurvich and M. Vyalyi, "Characterizing (quasi-) ultrametric finite spaces in terms of (directed) graphs," *Disc. App. Math.*, vol. 160, no. 12, pp. 1742–1756, 2012.

[34] I. Guyon, U. von Luxburg, and R. Williamson, "Clustering: Science or art?" in *NIPS Workshop on Clustering Theory*, 2009.

[35] M. S. Handcock, A. E. Raftery, and J. M. Tantrum, "Model-based clustering for social networks," *J. Royal Stat. Soc. Ser. A (Stat. Soc.)*, vol. 170, no. 2, pp. 301–354, 2007.

[36] R. A. Harshman, P. E. Green, Y. Wind, and M. E. Lundy, "A model for the analysis of asymmetric data in marketing research," *Mark. Sc.*, vol. 1, no. 2, pp. 205–242, 1982.

[37] E. Harzheim, *Ordered Sets*. Springer, 2005.

[38] J. Hopcroft and R. Tarjan, "Algorithm 447: Efficient algorithms for graph manipulation," *Commun. ACM*, vol. 16, no. 6, pp. 372–378, 1973.

[39] T. C. Hu, "The maximum capacity route problem," *Oper. Res.*, vol. 9, no. 6, pp. 898–900, 1961.

[40] L. Hubert, "Min and max hierarchical clustering using asymmetric similarity measures," *Psychometrika*, vol. 38, no. 1, pp. 63–72, 1973.

[41] P. Indyk, "Sublinear time algorithms for metric space problems," in *Symp. Theo. Comp. (STOC)*, 1999, pp. 428–434.

[42] J. R. Isbell, "Six theorems about injective metric spaces," *Comment. Math. Helv.*, vol. 39, no. 1, pp. 65–76, 1964.

[43] A. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Prentice Hall, 1988.

[44] N. Jardine and R. Sibson, *Mathematical Taxonomy*. John Wiley & Sons, 1971.

[45] T. R. Jensen and B. Toft, *Graph Coloring Problems*. John Wiley & Sons, 2011.

[46] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[47] ——, "An impossibility theorem for clustering," in *Advances Neural Info. Process. Syst. 15*, 2003, pp. 463–470.

[48] A. Kolmogorov and S. Fomin, *Introductory Real Analysis*. Dover Publications, 1975.

[49] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[50] J. B. Kruskal and M. Wish, *Multidimensional Scaling.* Sage, 1978.

[51] G. N. Lance and W. T. Williams, "A general theory of classificatory sorting strategies 1. Hierarchical systems," *Computer J.*, vol. 9, no. 4, pp. 373–380, 1967.

[52] E. Lawler, J. Lenstra, and A. Rinnooy Kan, "The traveling salesman problem: a guided tour of combinatorial optimization," *Wiley Intersc. Ser. in Disc. Math.*, 1985.

[53] E. Lieberman, C. Hauert, and M. Nowak, "Evolutionary dynamics on graphs," *Nature*, vol. 433, no. 7023, pp. 312–316, 2005.

[54] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *AAAI Conf. Artif. Intell.*, 2015, pp. 2181–2187.

[55] M. Meila and W. Pentney, "Clustering by weighted cuts in directed graphs," in *Intl. Conf. Data Min.*, 2007, pp. 135–144.

[56] M. Meila, "Comparing clusterings: an axiomatic view," in *Intl. Conf. Mach. Learn. (ICML)*, 2005, pp. 577–584.

[57] ——, "Comparing clusterings – an information based distance," *J. Multivar. Anal.*, vol. 98, no. 5, pp. 873 – 895, 2007.

[58] S. J. Messick and R. P. Abelson, "The additive constant problem in multidimensional scaling," *Psychometrika*, vol. 21, no. 1, pp. 1–15, 1956.

[59] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint arXiv:1109.2378*, 2011.

[60] F. Murtagh, "Multidimensional clustering algorithms," *Physika Verlag*, vol. 1, 1985.

[61] M. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, no. 3, p. 036104, 2006.

[62] M. Newman and M. Girvan, "Community structure in social and biological networks," *Proc. Ntnl. Acad. Sci.*, vol. 99, no. 12, pp. 7821–7826, 2002.

[63] ——, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, p. 026113, 2004.

[64] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances Neural Info. Process. Syst. 14*, pp. 849–856, 2002.

[65] W. Pentney and M. Meila, "Spectral clustering of biological sequence data," in *AAAI Conf. Artif. Intell.*, 2005, pp. 845–850.

[66] G. Punj and D. W. Stewart, "Cluster analysis in marketing research: Review and suggestions for application," *J. Mark. Res.*, vol. 20, no. 2, pp. 134–148, 1983.

[67] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059 – 1069, 2010.

[68] X. Rui and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, 2005.

[69] S. Sahni and T. Gonzalez, "P-complete approximation problems," *J. ACM*, vol. 23, no. 3, pp. 555–565, 1976.

[70] T. Saito and H. Yadohisa, *Data Analysis of Asymmetric Structures: Advanced Approaches in Computational Statistics*. CRC Press, 2004.

[71] S. Segarra and A. Ribeiro, "Stability and continuity of centrality measures in weighted graphs," *IEEE Trans. Signal Process.*, vol. 64, no. 3, pp. 543–555, 2016.

[72] S. Segarra, G. Carlsson, F. Mémoli, and A. Ribeiro, "Metric representations of network data," *Preprint available at https://www.seas.upenn.edu/~ssegarra/*, 2016.

[73] R. N. Shepard, "The analysis of proximities: Multidimensional scaling with an unknown distance function. I." *Psychometrika*, vol. 27, no. 2, pp. 125–140, 1962.

[74] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[75] R. Sibson, "SLINK: an optimally efficient algorithm for the single-link cluster method," *Computer J.*, vol. 16, no. 1, pp. 30–34, 1973.

[76] P. Slater, "Hierarchical internal migration regions of France," *IEEE Trans. Syst., Man, Cybern.*, no. 4, pp. 321–324, 1976.

[77] ——, "A partial hierarchical regionalization of 3140 US counties on the basis of 1965-1970 intercounty migration," *Env. Plan. A*, vol. 16, no. 4, pp. 545–550, 1984.

[78] R. Socher, D. Chen, C. D. Manning, and A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *Advances Neural Info. Process. Syst. 26*, 2013, pp. 926–934.

[79] R. E. Tarjan, "An improved algorithm for hierarchical clustering using strong components," *Inform. Process. Lett.*, vol. 17, no. 1, pp. 37–41, 1983.

[80] R. E. Tarjan and A. E. Trojanowski, "Finding a maximum independent set," *SIAM J. Comput.*, vol. 6, no. 3, pp. 537–546, 1977.

[81] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[82] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[83] C. Traina, A. Traina, B. Seeger, and C. Faloutsos, "Slim-trees: High performance metric trees minimizing overlap between nodes," in *Adv. Database Tech. (EDBT)*. Springer Berlin Heidelberg, 2000, pp. 51–65.

[84] J. K. Uhlmann, "Satisfying general proximity/similarity queries with metric trees," *Inform. Process. Lett.*, vol. 40, no. 4, pp. 175 – 179, 1991.

[85] T. van Laarhoven and E. Marchiori, "Axioms for graph clustering quality functions," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 193–215, 2014.

[86] V. Vassilevska, R. Williams, and R. Yuster, "All pairs bottleneck paths and max-min matrix products in truly subcubic time," *Theory Comput.*, vol. 5, pp. 173–189, 2009.

[87] U. von Luxburg, "A tutorial on spectral clustering," *Stat. and Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[88] D. Walsh and L. Rybicki, "Symptom clustering in advanced cancer," *Supp. Care Cancer*, vol. 14, no. 8, pp. 831–836, 2006.

[89] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *AAAI Conf. Artif. Intell.*, 2014, pp. 1112–1119.

[90] S. Warshall, "A theorem on boolean matrices," *J. ACM*, vol. 9, no. 1, pp. 11–12, 1962.

[91] P. N. Yianilos, "Data structures and algorithms for nearest neighbor search in general metric spaces," in *Symp. Disc. Algo. (SODA)*, 1993, pp. 311–321.

[92] R. B. Zadeh and S. Ben-David, "A uniqueness theorem for clustering," in *Conf. Uncert. Artif. Intell.*, 2009, pp. 639–646.

[93] Y. Zhao and G. Karypis, "Hierarchical clustering algorithms for document datasets," *Data Min. Knowl. Discov.*, vol. 10, pp. 141–168, 2005.

[94] B. Zielman and W. J. Heiser, "Models for asymmetric proximities," *Br. J. Math. Stat. Psychol.*, vol. 49, no. 1, pp. 127–146, 1996.