Tangles and Single Linkage Hierarchical Clustering

Eva Fluck
RWTH Aachen University, Germany
fluck@cs.rwth-aachen.de

Abstract

We establish a connection between tangles, a concept from structural graph theory that plays a central role in Robertson and Seymour's graph minor project, and hierarchical clustering. Tangles cannot only be defined for graphs, but in fact for arbitrary connectivity functions, which are functions defined on the subsets of some finite universe, which in typical clustering applications consists of points in some metric space.

Connectivity functions are usually required to be submodular. It is our first contribution to show that the central duality theorem connecting tangles with hierarchical decompositions (so-called branch decompositions) also holds if submodularity is replaced by a different property that we call maximum-submodular.

We then define a natural, though somewhat unusual connectivity function on finite data sets in an arbitrary metric space and prove that its tangles are in one-to-one correspondence with the clusters obtained by applying the well-known single linkage clustering algorithms to the same data set.

The idea of viewing tangles as clusters has first been proposed by Diestel and Whittle [5] as an approach to image segmentation. To the best of our knowledge, our result is the first that establishes a precise technical connection between tangles and clusters.

2012 ACM Subject Classification Mathematics of computing → Graph theory; Theory of computation → Unsupervised learning and clustering

Keywords and phrases Tangles, Branch Decomposition, Duality, Hierarchical Clustering, Single Linkage

Digital Object Identifier 10.4230/LIPIcs.MFCS.2019.38

1 Introduction

Connectivity in graphs and connectivity systems is a widely studied topic in theoretical computer science, e.g. [11, 8, 6, 7, 5]. On the other hand similarity, especially clustering, is an important and well studied topic in Data Science, e.g. [2, 4, 3, 10, 12]. We study the connection between both concepts by interpreting similarity as connectivity, thus two points are highly connected if their data is very similar and two sets are highly connected if they contain similar data points. Both communities will benefit from such a connection, as it opens up a basis for a wide range of new results. For example connectivity systems provide us with witnesses for the absence of highly connected regions, which is not yet established for clusters, as well as tree like representations of all those highly connected regions. Additionally there is large variety of efficient algorithms to compute or approximate different kind of clusters, which can possibly be used to find algorithms for computing highly connected regions in connectivity systems.

The concept of connectivity systems is based on the notion of connectivity in graphs. Such systems consist of a universe U of usually finitely many elements and some set function on subsets of U describing the connectivity between a set and its complement. These so called connectivity functions are symmetric and submodular. In this context two complementary questions are of interest [9]:

44th International Symposium on Mathematical Foundations of Computer Science (MFCS 2019).
Editors: Peter Rossmanith, Pinar Heggernes, and Joost-Pieter Katoen; Article No. 38; pp. 38:1–38:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1. What are the highly connected regions of the universe?
2. How can we decompose the universe along low order separations?

An answer to the first question can be given by tangles, which describe highly connected regions in a non-rigorous way. For every low order separation the tangle describes on which side of the separation the large part of the region can be found. Nevertheless, a separation may cut off small parts of the region. In this way, for any single point it is not clearly defined whether it is part of the region or not. On the other hand, the orientation has to be consistent, meaning that all orientations have to point towards the same region.

The second question is addressed by branch decompositions. They contain a ternary tree and a mapping of the elements to the leafs of the tree. Then, the edges of the tree represent separations of the universe. The width of such a decomposition is the largest value of any separation induced by the tree. Both concepts have been introduced on graphs by Robertson and Seymour [11]. An overview on both branch decompositions and tangles for integer-valued functions, as well as their connection can be found in a survey of Grohe [9], which can be translated to real-valued functions. Branch decompositions and tangles address contrary questions, but they are dual. There is a tangle of a certain value if and only if there is no branch decomposition of smaller value. This duality result has first been shown on graphs in [11]. There have been other set functions, that are not submodular, for which duality has been shown. For example Adler et al. [1] have shown duality up to a constant factor for so called hypertangle number and hyperbranch width in hypergraphs. Diestel and Oum [6] developed a general duality theorem in combinatorial structures.

Besides tangles, there is another approach to identify highly similar (connected) regions, called clustering (e.g. [2, 4, 3, 10, 12]). Clustering is the umbrella term for different techniques to define sets of data points that are very similar to each other and not so similar to the data points contained in other sets. Thus, a question arises: How do clustering and tangles relate? A first approach towards this question was done by Diestel and Whittle [5], where they analyzed tangles in digital images as a way to describe the meaningful parts of the image. We consider hierarchical clustering algorithms as introduced by Carlson and Mémoli [2]. The basis of such an algorithm is a function that describes the distance between two sets. Single linkage clustering for example considers the distance of two sets to be the smallest distance between any element from one set to any element from the other set. The hierarchical clustering algorithm then merges the sets with the smallest distance, assigning the resulting partition said value. We allow more than one merge at once, if the distances are equal. The resulting sequence of partitions is represented by a so called dendogram.

1.1 Results

To find a correspondence between tangles and clustering one of the main goals is to find connectivity functions, or functions with similar properties, that represent the different clustering methods. Our main result is, that we are able to specify a function that is correspondent to hierarchical clustering using single linkage. For an arbitrary metric $\mathrm{d}\colon U \times U \to \mathbb{R}$, the set function corresponding to single linkage is the *minimum distance function* $\delta_{\mathrm{d}}\colon 2^U \to \mathbb{R}$, defined by

$$\delta_{\mathrm{d}}(X) = \max_{x \in X, y \notin X} \exp(-\,\mathrm{d}(x, y)),$$

for all $X \in 2^U \setminus \{\emptyset, U\}$ and $\delta_{\mathrm{d}}(\emptyset) = \delta_{\mathrm{d}}(U) = 0$. An introduction of this function is given in Section 3.

We show that this function is not a classical connectivity function, since it is not submodular. Therefore we define a different property, that we call maximum-submodularity. Our first theorem proves duality between branch decompositions and tangles of functions with this property.

▶ **Theorem 1.** *Let U be a finite set and let ϕ be a maximum-submodular connectivity function. The maximum order of a tangle of ϕ equals the minimum order of a branch decomposition of ϕ. The existence of one is witness to the non-existence of the other.*

A formal version of this theorem and its proof are shown in Section 4. This duality is a key result in the theory of tangles of connectivity systems and suggests that the chosen property on the functions results in a similarly deep theory. It allows us to use maximum-submodular connectivity functions to establish a connection between tangles and clustering.

Our second main result says that the tangles of the minimum distance function are in one-to-one correspondence with the resulting dendogram of single linkage hierarchical clustering. The technical notions appearing in the statement of the theorem will be explained later in this paper.

▶ **Theorem 2.** *Let (U, d) be a metric space.*

1. *For every $r \in \mathbb{R}$ and every cluster $\mathcal{B}$ of the dendogram resulting from single linkage with $|\mathcal{B}| > 1$,*

$$\mathcal{T} := \{X \subseteq U \mid \delta_{\mathrm{d}}(X) < \exp(-r),\ \mathcal{B} \subseteq X\}$$

is a δ_{d}-tangle of U of order $\exp(-r)$.

2. *For every δ_{d}-tangle $\mathcal{T}$ of U of order k we can identify a cluster $\mathcal{B}$ of the dendogram resulting from single linkage with $|\mathcal{B}| > 1$ such that*

$$\mathcal{T} = \{X \subseteq U \mid \delta_{\mathrm{d}}(X) < k,\ \mathcal{B} \subseteq X\}.$$

For every non-singular set contained in a partition of the dendogram we find a distinct δ_{d}-tangle and vice versa. This is to the best of our knowledge the first precise technical connection between tangles and clusters.

2 Preliminaries and Definitions

In our definitions we follow [9]. Our goal is to describe connectivity within some data set U. Therefore we define set functions κ, that aim to describe how strong the connection is between a set and its complement. We say κ is *normalized* if $\kappa(\emptyset) = 0$, κ is *symmetric* if $\kappa(X) = \kappa(\overline{X})$, for all $X \subseteq U$ and κ is *submodular* if $\kappa(X) + \kappa(Y) \geq \kappa(X \cap Y) + \kappa(X \cup Y)$, for all $X, Y \subseteq U$. A set function that is normalized, symmetric and submodular is called *submodular connectivity function*.

▶ **Example 3** (see [9]). Let $G = (V, E)$ be a graph with edge weights $w_E \colon E \to \mathbb{R}$. The weighted edge-connectivity function $\nu \colon 2^V \to \mathbb{R}$, defined as

$$\nu(X) := \sum_{u \in X, v \in V \setminus X, (u,v) \in E} w_E(u, v),$$

is a submodular connectivity function.

We introduce a different type of function, that also describes connectivity. To show that this type has similar properties, we first take a look at basic concepts from the theory of connectivity systems. Most of these concepts have only been studied for integer-valued functions, but for our needs all properties are translatable to real-valued functions. We start with a formal definition of tangles, which are a way to describe highly connected regions.

▶ **Definition 4.** *Let κ be a symmetric set function on the universe U. A κ-*tangle *of order* $\mathrm{ord}(\mathcal{T}) = k \geq 0$ *is a set $\mathcal{T} \subseteq 2^U$ such that:*
T.0 *$\kappa(X) < k$ for all $X \in \mathcal{T}$,*
T.1 *for all $X \subseteq U$ with $\kappa(X) < k$, either $X \in \mathcal{T}$ or $\overline{X} \in \mathcal{T}$ holds,*
T.2 *$X_1 \cap X_2 \cap X_3 \neq \emptyset$ for all $X_1, X_2, X_3 \in \mathcal{T}$ and*
T.3 *$\{x\} \notin \mathcal{T}$ for all $x \in U$.*

We define the *tangle number* $\mathrm{tn}(\kappa)$ of a symmetric set function κ to be the largest possible order for which we can still define a κ-tangle.

We use the following well-known lemma, which states that tangles are in a way closed under intersection and supersets.

▶ **Lemma 5** (see [9]). *Let $\mathcal{T}$ be a κ-tangle of order k. Then it holds that*
1. *for all $X \in \mathcal{T}$ and all $Y \supseteq X$, if $\kappa(Y) < k$ then $Y \in \mathcal{T}$ and*
2. *for all $X, Y \in \mathcal{T}$, if $\kappa(X \cap Y) < k$ then $X \cap Y \in \mathcal{T}$.*

A different way to describe connectivity in a universe is given by branch decompositions. Here we do not look for highly connected regions, but ask ourselves how we can separate the universe into its single elements, using only separations of small value.

▶ **Definition 6.** *Let U be a finite set.*
- *A* pre-decomposition *of U is a pair (T, γ) consisting of a ternary (undirected) tree T and a mapping $\gamma \colon \overrightarrow{E}(T) \to 2^U$, from the directed edges of T to subsets of U, such that*
  - *$\gamma(t, u) = \overline{\gamma(u, t)}$, for all $(t, u) \in \overrightarrow{E}(T)$, and*
  - *$\gamma(s, u_1) \cup \gamma(s, u_2) \cup \gamma(s, u_3) = U$, for all internal nodes $s \in V(T)$ with $N(s) = \{u_1, u_2, u_3\}$.*
- *For leaves $\ell \in L(T)$ with neighbor $N(\ell) = \{u\}$, we write $\gamma(\ell)$ instead of $\gamma(u, \ell)$. We call the $\gamma(\ell)$* atoms *and define $\mathrm{At}(T, \gamma) \coloneqq \{\gamma(\ell) \mid \ell \in L(T)\}$.*
- *A pre-decomposition is* complete *if $|\gamma(\ell)| = 1$, for all leaves $\ell \in L(T)$.*
- *A pre-decomposition is* exact *at an internal node $t \in V(T)$ with $N(t) = \{u_1, u_2, u_3\}$ if all $\gamma(t, u_i)$ are mutually disjoint.*
- *A* decomposition *is a pre-decomposition that is exact at all internal nodes.*
- *A* branch decomposition *is a complete decomposition.*
- *Let κ be a set function on U. The* width *of a pre-decomposition (T, γ) is*

$$\mathrm{wd}(T, \gamma) \coloneqq \max\{\kappa(\gamma(t, u)) \mid (t, u) \in \overrightarrow{E}(T)\}.$$

We define the *branch width* $\mathrm{bw}(\kappa)$ of a symmetric set function κ to be the smallest possible width $\mathrm{wd}(T, \gamma)$ of any branch decomposition (T, γ) on U. For submodular connectivity functions duality between branch decompositions and tangles has been proven. The first to find this duality in graphs were Robertson and Seymour [11]. Duality between branch decompositions and tangles states that a branch decomposition of a certain width is a witness for the non-existence of a tangle of any larger order and vice versa. It follows that for any submodular connectivity function κ it holds that

$$\mathrm{tn}(\kappa) = \mathrm{bw}(\kappa).$$

3 The Minimum Distance Function

▶ **Definition 7** (Minimum Distance). *Let* $\mathrm{d}\colon U \times U \to \mathbb{R}$ *be an arbitrary metric. For a finite data set* U*, the* minimum distance function $\delta_{\mathrm{d}}\colon 2^U \to \mathbb{R}$*, is defined as follows:*

$$\delta_{\mathrm{d}}(X) := \begin{cases} 0 & \text{if } X = \emptyset \text{ or } \overline{X} = \emptyset, \\ \max_{x \in X, x' \in \overline{X}} \exp(-\mathrm{d}(x, x')) & \text{otherwise.} \end{cases}$$

This definition yields that X has a high value if there is a point outside of X very close to a point in X. The transformation $\exp(-c \cdot \mathrm{f}(x, y))$, for some constant c and some function f, is often used in clustering applications to transform a dissimilarity function f like a metric into a similarity function. The minimum distance function is in general not submodular, as can be seen with a small example. For an arbitrary $x \in \mathbb{R}^+$ define a one-dimensional universe containing only the following four points $a_1 = x$, $a_2 = x + 1$, $a_3 = -x$ and $a_4 = -x - 1$. Let $X = \{a_1, a_2\}$, $Y = \{a_1, a_3\}$ and the metric $d(u, v) = |u - v|$ is the absolute of the difference. Then $\delta_{\mathrm{d}}(X) = \exp(-2x) < \delta_{\mathrm{d}}(Y) = \delta_{\mathrm{d}}(X \cap Y) = \delta_{\mathrm{d}}(X \cup Y) = \exp(-1)$, for all $x > \frac{1}{2}$

We define a new property, that is similar to submodularity, which allows us to develop similar theories as for submodular connectivity functions.

▶ **Definition 8.** *A set function* κ *on a finite set* U *is* maximum-submodular *if, for all* $X, Y \subseteq U$*,*

$$\max(\kappa(X), \kappa(Y)) \geq \max(\kappa(X \cap Y), \kappa(X \cup Y)).$$

This property is neither a generalization of submodularity nor a specialization. For instance ν as in Example 3 is submodular but not maximum-submodular and in the next lemma we see that the minimum distance function, which in general is not submodular, is maximum-submodular. We call a normalized, symmetric and maximum-submodular set function *maximum-submodular connectivity function*.

▶ **Lemma 9.** *The minimum distance function is a maximum-submodular connectivity function.*

Proof. The minimum distance function is normalized by definition and symmetric since metrics are symmetric. If X or Y are equal to $\emptyset$ or U, maximum-submodularity trivially holds as $\{X, Y\} = \{X \cup Y, X \cap Y\}$ in these cases. Otherwise, we choose $u \in X \cap Y$ and $v \in \overline{X \cap Y}$ such that $\mathrm{d}(u, v) = \delta_{\mathrm{d}}(X \cap Y)$. Analogously we choose $u' \in \overline{X \cup Y} = \overline{X} \cap \overline{Y}$ and $v' \in X \cup Y$. Then w.l.o.g. we distinguish four cases, depending on v and v'.

Case 1: $v \in \overline{X} \cap \overline{Y}$ and $v' \in X \cap Y$ hold: Then w.l.o.g. $v = u'$ and $u = v'$ hold. Therefore $\delta_{\mathrm{d}}(X \cap Y) = \delta_{\mathrm{d}}(X \cup Y)$ and thus $\delta_{\mathrm{d}}(X) \geq \delta_{\mathrm{d}}(X \cap Y)$ and $\delta_{\mathrm{d}}(Y) \geq \delta_{\mathrm{d}}(X \cup Y)$ hold.

Case 2: $v \in \overline{X} \cap \overline{Y}$ and $v' \in \overline{X} \cap Y$ hold: It follows that $\delta_{\mathrm{d}}(X) \geq \delta_{\mathrm{d}}(X \cap Y)$ and $\delta_{\mathrm{d}}(Y) \geq \delta_{\mathrm{d}}(X \cup Y)$ hold.

Case 3: $v, v' \in \overline{X} \cap Y$ holds: It follows that $\delta_{\mathrm{d}}(X) \geq \delta_{\mathrm{d}}(X \cap Y)$ and $\delta_{\mathrm{d}}(Y) \geq \delta_{\mathrm{d}}(X \cup Y)$ hold.

Case 4: $v \in X \cap \overline{Y}$ and $v' \in \overline{X} \cap Y$ hold: In this case it holds that $\delta_{\mathrm{d}}(Y) \geq \delta_{\mathrm{d}}(X \cap Y)$ and $\delta_{\mathrm{d}}(Y) \geq \delta_{\mathrm{d}}(X \cup Y)$. Therefore we have $\delta_{\mathrm{d}}(Y) \geq \max(\delta_{\mathrm{d}}(X \cap Y), \delta_{\mathrm{d}}(X \cup Y))$ and the inequality holds.

As all other cases are symmetric to the four cases shown above, the inequality holds for all $X, Y \subseteq U$. ◀

Next, we consider tangles of the Minimum Distance Function. Firstly, we give an example of such a tangle.

▶ **Example 10.** Let $U \subset \mathbb{R}^n$ be a finite set of points. Let $x_1, x_2 \in U$ be two points such that $\mathrm{d}(x_1, x_2) = \min\{\mathrm{d}(x, y) \mid x, y \in U, x \neq y\}$. Then, for every $k \leq \exp(-\mathrm{d}(x_1, x_2))$,

$$\mathcal{T} := \{X \subseteq U \mid \delta_{\mathrm{d}}(X) < k,\ x_1, x_2 \in X\}$$

is a δ_{d}-tangle of order k.

$\mathcal{T}$ satisfies (T.0), (T.2) and (T.3) by construction. To see that (T.1) is satisfied note that if $x_1 \in X$ and $x_2 \in \overline{X}$ holds then $\delta_{\mathrm{d}}(X) = \exp(-\mathrm{d}(x_1, x_2)) \geq k$ holds.

Having this example we realize that the tangles described are the only δ_{d}-tangles.

▶ **Lemma 11.** *Every δ_{d}-tangle of order k is of the form described as in Example 10. That is, we can identify two points $u, v \in U$ such that $\exp(-\mathrm{d}(u, v)) \geq k$ and for all $X \in \mathcal{T}$ we have $u, v \in X$.*

Proof. To prove this we define the relation ${\delta_{\mathrm{d}}}^k := \{(u, v) \mid \exp(-\mathrm{d}(u, v)) \geq k\}$ and consider the graph $G_k := (U, {\delta_{\mathrm{d}}}^k)$. For every set $X \subseteq U$ with $\delta_{\mathrm{d}}(X) < k$ and every connected component C of G_k, holds either $V(C) \subseteq X$ or $V(C) \subseteq \overline{X}$ by definition. Thus for every δ_{d}-tangle $\mathcal{T}$ of order $\leq k$ it holds that all $X \in \mathcal{T}$ are disjoint unions of connected components of G_k. Additionally, using Lemma 5 (2) every such $\mathcal{T}$ is closed under intersection, as for two sets $X, Y \in \mathcal{T}$ and all connected components C of G_k either $V(C) \subseteq X \cap Y$ or $V(C) \cap X \cap Y = \emptyset$ and thus $\delta_{\mathrm{d}}(X \cap Y) < k$. Suppose for contradiction there is a δ_{d}-tangle $\mathcal{T}$ of order k such that there is no connected component of G_k, of size at least two, that is contained in all $X \in \mathcal{T}$. Let $C_0, \dots, C_n$ be an enumeration of all connected components of G_k with $|C_i| \geq 2$. Then we can identify a sequence $X_0, \dots, X_n \in \mathcal{T}$ such that $V(C_i) \nsubseteq X_i$, thus $V(C_i) \cap X_i = \emptyset$. We set $Y_1 := X_0 \cap X_1$ and $Y_{i+1} := Y_i \cap X_{i+1}$. Since $\mathcal{T}$ is closed under intersection, we get $Y_1, \dots, Y_n \in \mathcal{T}$. As $\mathcal{T}$ is a tangle, $|Y_n| > 1$ has to hold. For every subset $Y \subseteq Y_n$ we have $\kappa(Y) < k$ as $Y_n \cap \bigcup_{i=0}^{n} C_i = \emptyset$. Take an enumeration of all elements $y_1, \dots, y_m \in Y_n$ and construct a series of sets $Z_1, \dots, Z_\ell \in \mathcal{T}$ such that $|Z_\ell| = 1$. Clearly this contradicts the existence of $\mathcal{T}$. If $\{y_1\} \in \mathcal{T}$ set $Z_1 := \{y_1\}$, else set $Z_1 := Y_n \setminus \{y_1\} = Y_n \cap \overline{\{y_1\}} \in \mathcal{T}$. If $|Z_i| = 1$ set $\ell = i$ and stop the construction. Otherwise if $\{y_{i+1}\} \in \mathcal{T}$ set $Z_{i+1} := \{y_{i+1}\}$, else set $Z_{i+1} := Z_i \setminus \{y_{i+1}\} = Z_i \cap \overline{\{y_{i+1}\}} \in \mathcal{T}$. As $|Z_i| > |Z_{i+1}|$ this construction terminates and yields the desired contradiction. ◀

From this lemma an important corollary follows. In Section 5 we use this to identify for each tangle a cluster resulting from single linkage hierarchical clustering.

▶ **Corollary 12.** *Let $\mathcal{T}$ be a δ_{d}-tangle of order k over the universe U. There is a unique connected component C of the graph $G = (U, {\delta_{\mathrm{d}}}^k)$, with ${\delta_{\mathrm{d}}}^k := \{(u, v) \mid \exp(-\mathrm{d}(u, v)) \geq k\}$, such that $C \subseteq X$, for all $X \in \mathcal{T}$.*

Proof. We already showed that there exists some component C such that $C \subseteq X$, for all $X \in \mathcal{T}$. Assume there is some component $C' \neq C$ such that $C' \subseteq X$, for all $X \in \mathcal{T}$. Then we have $C' \in \mathcal{T}$ and thus $C \subseteq C'$ which contradicts $C' \neq C$. ◀

4 Duality for Submodular Bounded Functions

Now we prove duality for all maximum-submodular connectivity functions, thus also for the minimal distance function. We achieve a result similar to the theory for submodular connectivity functions, first shown in [11]. To formulate the Duality Theorem we first need a definition.

▶ **Definition 13.** *Let κ be a symmetric set function on U and $\mathcal{A} \subseteq 2^U$.*

- *A pre-decomposition (T, γ) is* over *$\mathcal{A}$ if $At(T, \gamma) \subseteq \mathcal{A}$.*
- *A κ-tangle $\mathcal{T}$* avoids *$\mathcal{A}$ if $\mathcal{T} \cap \mathcal{A} = \emptyset$.*

The Duality Theorem states that there can not be any decomposition over a family of sets, if there is a tangle avoiding this family and vice versa. The proof yields a construction of such a decomposition. The following theorem is a precise formulation of Theorem 1.

▶ **Theorem 14** (Duality Theorem of Submodular Bounded Functions)**.** *Let κ be a maximum-submodular connectivity function on U. Let $\mathcal{A} \subseteq 2^U$ such that $\mathcal{A}$ is closed under taking subsets and $\mathrm{Sing}(U) \subseteq \mathcal{A}$, where $\mathrm{Sing}(U)$ is the set of all singletons from U. Then there is a decomposition of width less than k over $\mathcal{A}$ if and only if there is no κ-tangle of order k that avoids $\mathcal{A}$.*

Assuming the theorem holds, we can directly derive the following corollary using that every branch decomposition of U is complete, thus is over $\mathrm{Sing}(U)$ and every κ-tangle avoids $\mathrm{Sing}(U)$ by definition.

▶ **Corollary 15.** *Let κ be a maximum-submodular connectivity function on U. It holds that*

$$\mathrm{tn}(\kappa) = \mathrm{bw}(\kappa).$$

Looking at the proof of duality for submodular connectivity functions as it is presented in [9], we see that they do not use any properties of the set function, besides symmetry and a transformation from a pre-decomposition into a decomposition of equal width. Therefore, we can adapt that proof if we are able to do a similar transformation. The following lemma shows how to achieve exactness at every node of a pre-decomposition.

▶ **Lemma 16** (Exactness Lemma)**.** *Let κ be a maximum-submodular connectivity function on U and (T, γ) be a pre-decomposition of U. Then there is a mapping $\gamma' \colon \overrightarrow{E}(T) \to 2^U$ such that (T, γ') is a decomposition of U satisfying*

- *$\mathrm{wd}(T, \gamma') \leq \mathrm{wd}(T, \gamma)$ and*
- *$\gamma'(\ell) \subseteq \gamma(\ell)$, for all leaves $\ell \in L(T)$.*

Proof. We iteratively construct γ' from γ, keeping the invariants

- $\mathrm{wd}(T, \gamma') \leq \mathrm{wd}(T, \gamma)$ and
- $\gamma'(\ell) \subseteq \gamma(\ell)$, for all leaves $\ell \in L(T)$.

We pick an arbitrary leaf $\ell_{start} \in L(T)$ and set $\gamma'(\ell_{start}, s) \coloneqq \gamma(\ell_{start}, s)$ and $\gamma'(s, \ell_{start}) \coloneqq \gamma(s, \ell_{start})$ for $s \in N(\ell_{start})$. If T only consists of at most two nodes we are done, since (T, γ') is already a decomposition. Otherwise, we traverse the tree with breadth-first search starting at ℓ_{start}. If we reach a node $s \in V(T) \setminus L(T)$ with predecessor $t \in V(T)$, we do the following. Let $u_1, u_2 \in N(s)$ be the successors of s and define $X \coloneqq \gamma'(s, t)$ and $Y_i \coloneqq \gamma(s, u_i)$, for $i = 1, 2$.

If $X \cap (Y_1 \cup Y_2) \neq \emptyset$ we update Y_i to $Y_i \cap \overline{X}$, for $i = 1, 2$. This step is consistent with the invariants as $\kappa(Y_i \cap \overline{X}) \leq \max(\kappa(Y_i \cap \overline{X}), \kappa(Y_i \cup \overline{X})) \leq \max(\kappa(Y_i), \kappa(X))$, where the second inequality holds due to maximum-submodularity, and $Y_i \cap \overline{X} \subseteq \gamma(s, u_i)$, for $i = 1, 2$.

If $Y_1 \cap Y_2 \neq \emptyset$, update Y_1 to $Y_1 \cap \overline{Y_2}$. This step is again consistent with the invariants as $\kappa(Y_1 \cap \overline{Y_2}) \leq \max(\kappa(Y_1 \cap \overline{Y_2}), \kappa(Y_1 \cup \overline{Y_2})) \leq \max(\kappa(Y_1), \kappa(Y_2))$, where the second inequality holds due to maximum-submodularity, and $Y_1 \cap \overline{Y_2} \subseteq \gamma(s, u_1)$.

Set $\gamma'(s, u_i) \coloneqq Y_i$ and $\gamma'(u_i, s) \coloneqq \overline{\gamma'(s, u_i)}$, for $i = 1, 2$. After these steps we know that γ' is exact at s and we do not change γ' for any predecessor of s.

When we reach a leaf $\ell \in L(T)$, we do not change γ' and continue with the next node in the breadth-first search.

This construction yields the desired mapping. ◀

The construction above may result in a tree with empty leafs. But such a leaf can be easily removed by deleting it and its neighbor, connecting the resulting open edges.

Now, we are ready to prove the Duality Theorem for Submodular Bounded Functions.

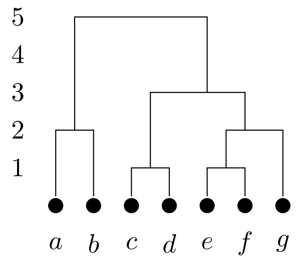
Proof of Theorem 14, see [9]. For the forward direction, we let (T, γ) be a decomposition of U over $\mathcal{A}$ of width less than k. Suppose, for contradiction, $\mathcal{T}$ is a κ-tangle of order k that avoids $\mathcal{A}$. We orient the edges $E(T)$ such that they point in the direction of the set that is contained in the tangle. Such a set always exists as the width is less then k. Thus, formally we orient $(s,t) \in E(T)$ towards t if $\gamma(s,t) \in \mathcal{T}$, and towards s if $\gamma(t,s) = \overline{\gamma(s,t)} \in \mathcal{T}$. As in every oriented tree, there is at least one node $t \in V(T)$ such that all edges incident to t are oriented towards t. If $t \in L(T)$ then $\gamma(t) \in \mathcal{A}$ and $\gamma(t) \in \mathcal{T}$ which contradicts the assumption that $\mathcal{T}$ avoids $\mathcal{A}$. Thus, t is an internal node with $N(t) = \{u_1, u_2, u_3\}$. But since all $\gamma(t, u_i)$ are mutually disjoint and all $\gamma(u_1, t) \in \mathcal{T}$ this contradicts (T.2) as $\gamma(t,u_1) \cup \gamma(t,u_2) \cup \gamma(t,u_3) = U$ and thus $\gamma(u_1,t) \cap \gamma(u_2,t) \cap \gamma(u_3,t) = \emptyset$. It follows that such a κ-tangle can not exist and the forward direction holds.

For the backward direction assume there is no κ-tangle of order k that avoids $\mathcal{A}$. We will construct a pre-decomposition (T, γ) of U over $\mathcal{A}$ of width less than k. Using the Exactness Lemma and since $\mathcal{A}$ is closed under taking subsets it follows that a decomposition of U over $\mathcal{A}$ exists.

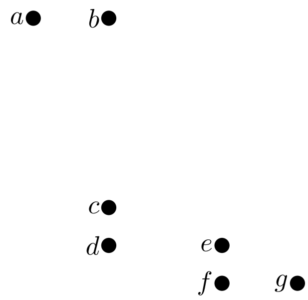
We construct such a pre-decomposition (T, γ) inductively on the number of sets $X \subseteq U$ with $\kappa(X) < k$ and neither $X \in \mathcal{A}$ nor $\overline{X} \in \mathcal{A}$.

In the base case, for all $X \subseteq U$ with $\kappa(X) < k$, holds $X \in \mathcal{A}$ or $\overline{X} \in \mathcal{A}$. We define $\mathcal{Y} := \{\overline{X} \mid X \in \mathcal{A} \text{ with } \kappa(X) < k\}$. We know that $\mathcal{Y}$ can not be a tangle, as we assumed that there is no tangle of order k. Since (T.0) and (T.1) hold by assumption on $\mathcal{A}$, either (T.2) or (T.3) have to be false. If $\mathcal{Y}$ violates (T.2) there are three sets $Y_1, Y_2, Y_3 \in \mathcal{Y}$ such that $Y_1 \cap Y_2 \cap Y_3 = \emptyset$. Then $\overline{Y_1}, \overline{Y_2}, \overline{Y_3} \in \mathcal{A}$ and $\overline{Y_1} \cup \overline{Y_2} \cup \overline{Y_3} = U$. We set $T := (\{\ell_1, \ell_2, \ell_3, t\}, \{(\ell_i, t) \mid i = 1,2,3\})$, $\gamma(t, \ell_i) := \overline{Y_i} \in \mathcal{A}$ and $\gamma(\ell_i, t) := Y_i$. Then, (T, γ) is a pre-decomposition of U over $\mathcal{A}$. If $\mathcal{Y}$ violates (T.3) there is some $x \in U$ such that $\{x\} \in \mathcal{Y}$ and thus $\overline{\{x\}} \in \mathcal{A}$. Since $\mathrm{Sing}(U) \subseteq \mathcal{A}$ we have $\{x\} \in \mathcal{A}$. We take $T := (\{s,t\}, \{(s,t)\})$ and $\gamma(s,t) := \{x\}$, $\gamma(t,s) := \overline{\{x\}}$. Then (T, γ) is a pre-decomposition of U over $\mathcal{A}$.

In the inductive step, we have some $X \subseteq U$ with $\kappa(X) < k$ and neither $X \in \mathcal{A}$ nor $\overline{X} \in \mathcal{A}$. We chose X' such that $|X'|$ is minimal with respect to the conditions above. We set $\mathcal{A}^1 := \mathcal{A} \cup 2^{X'}$ and $\mathcal{A}^2 := \mathcal{A} \cup 2^{\overline{X'}}$. By the induction hypothesis there are pre-decompositions (T^1, γ^1) over $\mathcal{A}^1$ and (T^2, γ^2) over $\mathcal{A}^2$. If $\mathrm{At}(T_i, \gamma_i) \subseteq \mathcal{A}$ than (T_i, γ_i) is a pre-decomposition over $\mathcal{A}$ and we are done. Otherwise, we can assume that (T^1, γ^1) is a decomposition, due to the Exactness Lemma 16, thus the $\gamma^1(\ell)$ for all $\ell \in L(T^1)$ are unique. There is some $\ell^1 \in L(T^1)$ such that $\gamma^1(\ell^1) \notin \mathcal{A}$. As the width of (T^1, γ^1) is less than k and no true subset $X'' \subset X'$ fulfills $\kappa(X'') < k$ and neither $X'' \in \mathcal{A}$ nor $\overline{X''} \in \mathcal{A}$, we know that $\gamma^1(\ell^1) = X'$ and that it is the only leaf with this condition. We denote its neighbor by s^1. Let us now consider all $\ell^2_1, \dots, \ell^2_m \in L(T^2)$ with $\gamma^2(\ell^2_i) \notin \mathcal{A}$. We know that, for all ℓ^2_i, we have $\gamma^2(\ell^2_i) \subseteq \overline{X'}$. We consider all s^2_i with $N(\ell^2_i) = \{s^2_i\}$. We modify γ^2 by setting $\gamma^2(s^2_i, \ell^2_i) := \overline{X'}$ and $\gamma^2(\ell^2_i, s^2_i) := X'$. The result will still be a pre-decomposition. Then, we construct a pre-decomposition (T, γ) of U over $\mathcal{A}$. We take m disjoint copies (T^1_i, γ^1_i) of (T^1, γ^1). We



(a) The dendogram corresponding to the data points using single linkage.



(b) Data points used to compute the dendogram.

Figure 1 An example of a dendogram. The distance function used is $\ell_{SL}(X,Y) := \min_{x\in X, y\in Y} \|x-y\|$, where $\|\cdot\|$ is the Euclidean norm. This distance function is used in single linkage.

define

$$V(T) := \bigcup_{1\leq i\leq m} V(T_i^1)\setminus\{\ell_i^1\} \cup V(T^2)\setminus\{\ell_1^2,\ldots,\ell_m^2\}$$

and take the union of all edge sets where ℓ_i^1 is replaced by s_i^2 and ℓ_i^2 is replaced by s_i^1. Then, we define $\gamma\colon E(T)\to 2^U$ by

$$\gamma(s,t) := \begin{cases} X' & \text{if } (s,t) = (s_i^1, s_i^2) \text{ for some } 1\leq i\leq m,\\ \overline{X'} & \text{if } (s,t) = (s_i^2, s_i^1) \text{ for some } 1\leq i\leq m,\\ \gamma^1(s,t) & \text{if } s,t\in V(T_i^1) \text{ for some } 1\leq i\leq m,\\ \gamma^2(s,t) & \text{if } s,t\in V(T^2).\end{cases}$$

Then, (T,γ) is a pre-decomposition of U over $\mathcal{A}$ of width less than k. ◀

5 Minimum Distance Function and Hierarchical Clustering

To establish the connection between the δ_{d}-tangles and hierarchical clustering, we use Agglomerative Hierarchical Clustering via single linkage on dissimilarity inputs. A dissimilarity input is an instance, where a small function value describes a large similarity between the points. The result of a hierarchical clustering algorithm is a dendogram. For an arbitrary set U, $\mathcal{P}(U)$ denotes the *set of all partitions* of U.

▶ **Definition 17** ([2]). *A* dendogram *over a finite set* $U = \{x_1,\ldots,x_n\}$ *is a function* $\theta\colon [0,\infty)\to\mathcal{P}(U)$*, satisfying the following conditions:*

1. $\theta(0) = \{\{x_1\},\ldots,\{x_n\}\}$,
2. *there exists* t_0 *such that* $\theta(t) = \{U\}$ *for all* $t\geq t_0$,
3. *if* $r\leq s$ *then* $\theta(r)$ *is a* refinement *of* $\theta(s)$*, that is for every* $\mathcal{B}\in\theta(r)$ *there is some* $\mathcal{B}'\in\theta(s)$ *such that* $\mathcal{B}\subseteq\mathcal{B}'$*, and*
4. *for all* r *there exists* $\epsilon>0$ *such that* $\theta(r)=\theta(t)$ *for all* $t\in[r,r+\epsilon]$.

The first and second condition ensure that the trivial partitions are part of the dendogram, with the single elements having the smallest possible value and the whole set giving an upper bound. The third condition states that every partition results from a merge of sets contained in a more refined partition. The last condition is a bit technical, and ensures that θ is right continuous. An example of a dendogram can be seen in Figure 1. We allow more than one cluster to merge in one step, as introduced and analyzed by Carlson and Mémoli [2]. The single linkage clustering in this framework works as follows.

▶ **Definition 18** ([2]). *Let (U, d) be a metric space and let $\ell_{SL}\colon 2^U \times 2^U \to \mathbb{R}$ be the single linkage function on U defined by*

$$\ell_{SL}(\mathcal{A}, \mathcal{B}) \coloneqq \min_{a \in \mathcal{A}, b \in \mathcal{B}} \mathrm{d}(a, b).$$

Define a sequence of distances $R_0, R_1, R_2, \ldots \in [0, \infty)$ and a corresponding sequence of partitions $\Theta_0, \Theta_1, \Theta_2, \ldots \in \mathcal{P}(U)$ by:

- *$R_0 = 0$ and $\Theta_0 = \{\{x_1\}, \ldots, \{x_n\}\}$, with $U = \{x_1, \ldots, x_n\}$,*
- *$R_{i+1} \coloneqq \min_{\mathcal{B}, \mathcal{B}' \in \Theta_i} \ell_{SL}(\mathcal{B}, \mathcal{B}')$*
- *$\Theta_{i+1} \coloneqq \Theta_i / \sim_{R_{i+1}}$, where $\mathcal{B} \sim_{R_{i+1}} \mathcal{B}'$ if there exists a sequence of blocks of distance at most R_{i+1}, thus $\mathcal{B} = \mathcal{B}_1, \ldots, \mathcal{B}_s = \mathcal{B}' \in \Theta_i$ with $\ell_{SL}(\mathcal{B}_k, \mathcal{B}_{k+1}) \leq R_{i+1}$, for $k = 1, \ldots, s-1$.*

Then the dendogram for single linkage *is defined by*

$$\theta^{\ell_{SL}}(r) \coloneqq \Theta_{i(r)},$$

where $i(r) \coloneqq \max\{i \mid R_i \leq r\}$.

A less technical way to describe this is, that we start with distance $R_0 = 0$ and the partition into single elements. Then we inductively compute the smallest pairwise distance of any two points separated by the partition Θ_i, store this as the next distance value R_{i+1} and merge the corresponding sets to achieve a new partition Θ_{i+1}. We repeat this step until all sets are merged. The resulting dendogram can be interpreted as a decomposition of the universe into its δ_{d}-tangles, where the non-singular blocks of the dendogram correspond to the tangles. The following theorem is a precise formulation of Theorem 2.

▶ **Theorem 19.** *Let (U, d) be a metric space.*

1. *For every $r \in \mathbb{R}$ and every block $\mathcal{B} \in \theta^{\ell_{SL}}(r)$ with $|\mathcal{B}| > 1$,*

   $$\mathcal{T} \coloneqq \{X \subseteq U \mid \delta_{\mathrm{d}}(X) < \exp(-r),\ \mathcal{B} \subseteq X\}$$

   is a δ_{d}-tangle of U of order $\exp(-r)$.
2. *For every δ_{d}-tangle $\mathcal{T}$ of U of order k we can identify a block $\mathcal{B} \in \theta^{\ell_{SL}}(-\log(k))$ with $|\mathcal{B}| > 1$ such that*

   $$\mathcal{T} = \{X \subseteq U \mid \delta_{\mathrm{d}}(X) < k,\ \mathcal{B} \subseteq X\}.$$

Proof. Using the same arguments as in Example 10 the first statement holds. For the second statement one needs the equivalence relation $\sim_r$ on U, where $x \sim_r y$ if and only if there is a sequence of elements $x = x_1, \ldots, x_s = y \in U$ such that $\mathrm{d}(x_i, x_{i+1}) \leq r$. Carlson and Mémoli [2] have shown that the blocks of $\theta^{\ell_{SL}}(r)$ are exactly the equivalence classes $U/{\sim_r}$.

Let $\mathcal{T}$ be a δ_{d}-tangle of order k. Using Corollary 12 we can find a connected component C in the graph $G = (U, {\delta_{\mathrm{d}}}^k)$ such that $C \subseteq X$, for all $X \in \mathcal{T}$. Looking at the definition ${\delta_{\mathrm{d}}}^k \coloneqq \{(u, v) \mid \exp(-\mathrm{d}(u, v)) \geq k\}$ we see that two elements $u, v \in U$ are connected in G if and only if for their distance holds $\mathrm{d}(u, v) \leq -\log(k)$, thus $u \sim_{-\log(k)} v$. It follows that the equivalence classes of $U/{\sim_{-\log(k)}}$ are exactly the same as the connected components of G. Thus, there is a block $C = \mathcal{B} \in \theta^{\ell_{SL}}(-\log(k))$ that fulfills the requirement. ◀

▶ Remark 20. Let us consider two popular hierarchical clustering algorithms, average linkage and complete linkage. The algorithm is the same as in Definition 18, but the linkage function ℓ changes. The distance of two sets for complete linkage equals the maximum distance of any point from one set to any point from the other set, thus $\ell_{CL}(X,Y) \coloneqq \max_{x\in X, y\in Y} \mathrm{d}(x,y)$. Using the same trick as for single linkage, a natural related connectivity function is $\kappa_{\mathrm{d}}(X) \coloneqq \min_{x\in X, y\in \overline{X}} \exp(-\mathrm{d}(x,y))$, for $X \in 2^U\setminus\{\emptyset, U\}$, and $\kappa_{\mathrm{d}} = 0$, otherwise. This function is maximum-submodular and using the Manhattan distance it is even submodular. But in general, for an arbitrary partition P, it holds that

$$\min_{X,Y\in P} \max_{x\in X, y\in Y} \mathrm{d}(x,y) \neq -\log(\max_{X\in P} \min_{x\in X, x'\in\overline{X}} \exp(-\mathrm{d}(x,x'))),$$

as $\ell_{CL}(X,\overline{X}) = \max_{Y\in P} \ell_{CL}(X,Y)$, for arbitrary $X \in P$. This is different for single linkage. It holds that for any partition P of the universe we have

$$\min_{X,Y\in P} \ell_{SL}(X,Y) = -\log(\max_{X\in P} \delta_{\mathrm{d}}(X)),$$

as $\ell_{SL}(X,\overline{X}) = \min_{Y\in P} \ell_{SL}(X,Y)$, for arbitrary $X \in P$. Thus in contrast to single linkage, the optimum of complete linkage ℓ_{CL} used to compute R_{i+1} does not correspond to the optimum according to the connectivity function κ_{d}. For average linkage ($\ell_{AL}(X,Y) \coloneqq \sum_{x\in X, y\in Y} \frac{\mathrm{d}(x,y)}{|X||Y|}$), a corresponding set function could be $\varphi_{\mathrm{d}}(X) \coloneqq \sum_{x\in X, y\in\overline{X}} \frac{\exp(-\mathrm{d}(x,y))}{|X||\overline{X}|}$, for $X \in 2^U\setminus\{\emptyset, U\}$, and $\varphi_{\mathrm{d}} = 0$, otherwise. It is neither submodular nor maximum-submodular. Additionally in general $\ell_{AL}(X,\overline{X})$ is not directly computable from $\ell_{AL}(X,Y)$, for $X, Y \in P$ with P an arbitrary partition. To compute $\ell_{AL}(X,\overline{X})$, also the size of all $Y \in P$ is needed. We have

$$\ell_{AL}(X,\overline{X}) = \frac{\sum_{Y\in P, X\neq Y} |Y| \ell_{AL}(X,Y)}{\sum_{Y\in P, X\neq Y} |Y|}.$$

This makes it even harder to find a suitable connectivity function.

6 Conclusion

We establish a precise technical connection between tangles and hierarchical clustering. We can specify this connection for the minimum distance function and single linkage clustering. It is still an open question if there are other clustering algorithms for which we can find corresponding set functions. One of the main obstacles here is, that tangles and the corresponding set functions only look at global connectivity of some set to its converse whereas hierarchical clustering looks at local connectivity between two sets. For single linkage these two notions turned out to be the same.

Our second contribution is to show duality between tangles and branch decompositions for a new class of functions. In our view, the key transformation in the proof of the Duality Theorem is the Exactness Lemma (related to „shifting“ in [6]); this is where submodularity or maximum-submodularity comes in. To broaden the theory, it will be essential to understand under which general conditions such a transformation is possible.

References

1 Isolde Adler, Georg Gottlob, and Martin Grohe. Hypertree width and related hypergraph invariants. *European Journal of Combinatorics*, 28(8):2167–2181, 2007. doi:10.1016/j.ejc.2007.04.013.

2 Gunnar Carlsson and Facundo Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11(Apr):1425–1470, 2010.

3 Vincent Cohen-Addad, Varun Kanade, Frederik Mallmann-Trenn, and Claire Mathieu. Hierarchical Clustering: Objective Functions and Algorithms. In Artur Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 378–397. SIAM, 2018. doi:10.1137/1.9781611975031.26.

4 Sanjoy Dasgupta. A cost function for similarity-based hierarchical clustering. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 118–127. ACM, 2016. doi:10.1145/2897518.2897527.

5 Reinhard Diestel, Fabian Hundertmark, and Sahar Lemanczyk. Profiles of Separations: in Graphs, Matroids, and Beyond. *Combinatorica*, 39(1):37–75, 2019. doi:10.1007/s00493-017-3595-y.

6 Reinhard Diestel and Sang-il Oum. Unifying Duality Theorems for Width Parameters in Graphs and Matroids (Extended Abstract). In Dieter Kratsch and Ioan Todinca, editors, *Graph-Theoretic Concepts in Computer Science - 40th International Workshop, WG 2014, Nouan-le-Fuzelier, France, June 25-27, 2014. Revised Selected Papers*, volume 8747 of *Lecture Notes in Computer Science*, pages 1–14. Springer, 2014. doi:10.1007/978-3-319-12340-0_1.

7 Reinhard Diestel and Geoff Whittle. Tangles and the Mona Lisa. *ArXiv*, 2016. arXiv:1603.06652 [math.CO]. arXiv:1603.06652.

8 Jim Geelen, Bert Gerards, and Geoff Whittle. Tangles, tree-decompositions and grids in matroids. *Journal of Combinatorial Theory, Series B*, 99(4):657–667, 2009. doi:10.1016/j.jctb.2007.10.008.

9 Martin Grohe. Tangled up in blue (a survey on connectivity, decompositions, and tangles). *ArXiv*, 2016. arXiv:1605.06704 [cs.DM].

10 Jon M Kleinberg. An impossibility theorem for clustering. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15 [Neural Information Processing Systems, NIPS 2002, December 9-14, 2002, Vancouver, British Columbia, Canada]*, pages 446–453. MIT Press, 2002. URL: http://papers.nips.cc/book/advances-in-neural-information-processing-systems-15-2002.

11 Neil Robertson and Paul D Seymour. Graph minors. X. Obstructions to tree-decomposition. *Journal of Combinatorial Theory, Series B*, 52(2):153–190, 1991. doi:10.1016/0095-8956(91)90061-N.

12 Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. doi:10.1007/s11222-007-9033-z.