Nine Quick Tips for Analyzing Network Data

Vincent Miele, Stephane Robin, Catherine Matias, Stéphane Dray

HAL Id: hal-02089501

https://hal.archives-ouvertes.fr/hal-02089501v2

Submitted on 19 Jul 2019

Nine Quick Tips for Analyzing Network Data

Vincent Miele[1,*], Catherine Matias[2], Stéphane Robin[3], and Stéphane Dray[1]

[1]Université de Lyon, F-69000 Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive, F-69622 Villeurbanne, France
[2]Sorbonne Université, Université Paris Diderot, Centre National de la Recherche Scientifique, Laboratoire de Probabilités, Statistique et Modélisation, F-75005 Paris, France
[3]UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, F-75005, Paris, France
[*]Corresponding author: vincent.miele@univ-lyon1.fr

Abstract

These tips provide a quick and concentrated guide for beginners in the analysis of network data.

Introduction

From the molecular to the ecosystem level, a biological system can often be represented as a set of entities that interact with other biological entities. Recent advances in data acquisition technology (e.g., high-throughput sequencing or tracking devices) open up the opportunity to quantify these interactions and call for the development of ambitious methodology to tackle these data. In this context, networks are widely used in Biology, Bioinformatics, Ecology, Neuroscience or Epidemiology to represent interaction data [1]. A network contains a set of entities (the *nodes* or *vertices*) that are connected by *edges* (or *links*) depicting some interactions or relationships. These relationships may be either directly observed or deduced from raw data. The first case encompasses protein-protein interaction (PPI) networks where interactions between two proteins are experimentally assessed or plant-pollinator interactions that are directly observed in the field. Gene regulatory networks reconstructed from gene expression data, co-occurrence networks inferred from species abundances or animal social contact networks deduced from GPS tracks are some examples of the second case. New kinds of networks are still emerging (for instance, cell-cell similarity networks [2], Hi-C networks and image similarity networks [3]).

Networks are very attractive objects and many methods have been developed to analyze their structure. However, biological networks are often analyzed by non-specialists and it may be difficult for them to navigate through the plethora of concepts and available methods. In this paper, we propose nine tips to avoid common pitfalls and enhance the analysis of network data by biologists.

1

Tip 1:   Formulate questions first, use networks later

Network theory is well established and truly powerful but it cannot be used as a "black-box". Indeed, building a network should not be considered as an end in itself. We recommend to (i) establish a list of scientific questions and hypotheses before manipulating the data and then (ii) evaluate if these questions naturally translate into a series of network analyses, rather than making network analyses first and checking whether they raise questions after (in agreement with rule 1 in [4]). Indeed, it is generally immediate to represent/model the data with a network, but much trickier to translate a question into a network-based analysis.

To this end, besides integrating the network formalism, it is important to embrace the network viewpoint. It relies on a cornerstone idea that makes the strength but also the challenge of network modelling: any interaction is considered within its context taking into account the other interactions that occur (or not). In this viewpoint, any interaction between the pair of nodes (A,B) is considered in the context of the other pairs involving A or B. For instance, the importance of a particular edge between two genes will be differently assessed if the target gene is or is not a *hub* (i.e. regulated by many genes). This viewpoint does not consider interactions as independent objects and is thus the exact opposite of examining the set of interactions one by one.

Finally, it is obviously recommended to check whether your questions and data really fit the network viewpoint before performing any analysis. If the number of nodes and/or edges is very low, network analysis can be applied but results can be disappointing as they are not enough observed interactions to identify a structure in the data. On the other hand, although any matrix can be viewed as a network (one edge per cell, see next Tip), it is often more adequate to consider using non-network methods dedicated to complete matrices. For instance a correlation matrix, possibly viewed as a correlation network, can be naturally analysed with a hierarchical clustering or a principal component analysis. In other words, network analysis is not necessarily the answer when analysing a data matrix.

Tip 2:   Categorize your network data correctly

To grab the cutting-edge concepts and methods in the networks field, learning the appropriate vocabulary from *graph theory* is a prerequisite [5]. In particular, it is important to categorize your network properly to be sure you apply suitable methods. Different network categories for different data lead to different approaches.

Links can be *directed* (from a *source* to a *target*), possibly including *self-loops* (e.g., a protein interacting with itself or cannibalism in food webs). Ignoring this information for the sake of simplicity would actually betray the original data. When dealing with edges embedding a value (a *weight*), we strongly advise you to avoid transforming the network into a *binary* one using any *ad-hoc* threshold value. Indeed it clears a significant part of the available information because some aspects of the network structure might be undetected in the binarized network [6]. This binarization could be used as an exploratory step only (for instance, to facilitate a first visualization step - see Tip 4), but it can bias your analysis (e.g., a nested pattern can be observed in binarized ecological networks but no not in weighted ones [7]). Methods handling weighted networks are usually available and therefore more efficient. Furthermore, the data analyst must be very cautious since, in the literature, weights can be considered as intensity-based (the greater the weight, the stronger the edge is) as well as distance-based (the smaller the weight, the closer the nodes are).

Nodes can belong to different categories and edges can be allowed only between nodes of different categories (*bi/tri/multi-partite* networks; e.g., nodes as hosts/parasites

or as plant/fungus/seed dispersers [8]). It is mandatory to select methods that handle this particularity. For instance, many statistical approaches rely on the expected number of edges (e.g., in the computation of *modularity*, see Tip 5) which is here clearly different compared to the unipartite case.

Finally, additional information on the nodes is often available. For instance, nodes can have spatial positions (e.g., nodes as habitat patches or farms in 2D, brain area in 3D) or can be associated to external *attributes* (e.g., species traits in a food web). This additional information can be explicitly considered in the analysis, either to understand if it contributes to organize the network [9], or to look for some remaining structure once accounted for its effect (e.g., spatial [10] or phylogenetic effect [11]). In the former case, a simpler but suboptimal alternative often consists in using this information *a posteriori* in the interpretation of results (e.g., explaining the structure of genetic networks with spatial information [12] or comparing network structure with *metadata* [13]).

Tip 3:   Use specific network analysis software

A range of versatile software is dedicated to network analysis. It is therefore a waste of time trying to use unspecific tools. These software tools belong to two distinct categories that have pros/cons: graphical user interface (mouse-based navigation) and software packages (command line interface or programming). The first category is mainly dedicated to powerful and interactive visualization (see Tip 4). It includes the two major open source software tools `Gephi` and `Cytoscape`, both supported by an active community. They also offer the computation of some network metrics (the choice of a relevant metric is discussed in Tip 5). The second category is dominated by the two leading general-purpose network packages `NetworkX` and `igraph`, but there exist plenty of more specific packages (for instance `bipartite` in `R`). Browser-based visualization [14] recently emerged as an intermediate category, mostly based on a collection of `javascript` libraries (e.g., `Sigma.js`).

That said, we strongly suggest that you learn programming and scripting your analysis (in agreement with papers in the "10 simple rules" collection about computing skills and reproducibility [15,16]). Dealing with reproducible code enhances network research: you can re-run with no effort the complete analysis on a modified version of your raw data, on different datasets and share the code with others colleagues interested in the modelling approach. Finally, there exist a limited set of common network file formats (e.g., adjacency list in the format `source target`) that you should adopt from the very beginning, to easily switch between different software tools.

Meanwhile, the data analyst should avoid a hasty use of the different functions implemented in these tools. As underlined in Tips 5 and 6, it is crucial to understand the metrics/methods before running functions, and to select the appropriate ones with respect to the questions and the data at hand.

Tip 4:   Be aware that network visualization can be useful but possibly misleading

One powerful aspect of networks is their ability to depict complex data in a single object. It can be therefore tempting to represent networks graphically in two dimensions: nodes are spread in the plane and edges drawn with the objective to achieve the most aesthetic design (the nodes'positioning is called a *layout*). This apparently simple task is in fact a very hard combinatorial problem. An active research community proposed a series of heuristics aiming at obtaining a nice network view in a

reasonable time, despite the growing size of available networks. The aforementioned tools (see Tip 3) embed a wide range of easy-to-use layouts.

Graphics are usually considered as an important tool for exploratory data analysis [17]. However, special care is required to not over-interpret network visualization. A layout does not only provide a nice representation of a network, it makes it optimal for a given set of objectives (e.g., maximizing attractions between connected nodes) that you often ignore. As a consequence, what you see with your eyes can be biased. When visualizing a network, always keep in mind that the position of a node in such a display is not part of the data, but results from an algorithm. Hence, the distance between two nodes should not be interpreted as an intrinsic measure of proximity as another display algorithm would result in a possibly very different distance (see the distances between the two red nodes in Figure 1a-b). On the other hand, network visualization can be useful as a way to illustrate the results of a network analysis (as presented in Tips 5 and 6). In this case, a layout should be chosen for its ability to highlight network properties (Figure 1c) or conclusions drawn by an analysis (Figure 1d). For instance, nodes can be positioned according to the values of some particular metrics of interest [18]. In any case, we encourage biologists to clearly describe the layout used in any graphical representation of a network in scientific publications, especially to make it reproducible.

Lastly, we also advise to consider visualizing the *adjacency matrix* as a heatmap/a colored matrix (see Figure 1 in [19] for an explanation). It allows to represent the presence or weight of edges (colored cells) but it has also the advantage to highlight edges' absence (blank matrix cells). This is particularly relevant when the matrix rows/columns are reordered in an informative manner (e.g., by increasing value of a metric [20] or according to some clustering results; see Tips 5 and 6 and Figure 1 d).
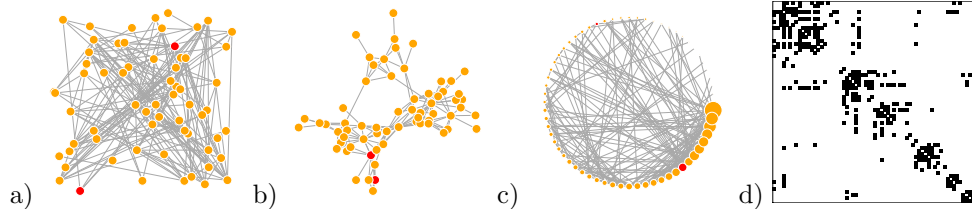
Fig 1. Four visualizations of the same network modelling interactions between 64 sociable weavers [14, 21]. a) Random layout. b) Fruchterman and Reingold layout. c) Circle layout where nodes'size and position are defined by their degree. The same two nodes are colored in red in panels a-c to show their distance varies depending on the layout. d) Representation of the adjacency matrix with row/columns ordering consistent with the clustering obtained with the Infomap algorithm (see [22] for details). Graphical representations are performed with the package `igraph`.

Tip 5:   Avoid blind use of metrics, understand formulas instead

Beside the limitations of network visualization, describing a network can also (and advantageously) consist in computing summary statistics. The beginner will immediately find the path to a series of *network metrics*: one number per node or edge (local metrics; e.g., *degree*) or one number for the whole network (global metrics; e.g., *connectance/density* or modularity). Metrics have proliferated and it is strongly advised to take time to read carefully the mathematical definition of the metrics one has at hand (see also Tip 9): the deeper the mathematical understanding, the easier the

interpretation is. For instance, the concept of *nodes'centrality* goes with a range of centrality metrics that have different meanings. Moreover, it is so easy to compute any metric with the aforementioned software tools that it can sometimes prevent the analyst to check their pros/cons. As an example, reading the definition of the widely-used *betweenness centrality*, you can understand it is based on *shortest paths*. If you intend to use this measure, it is therefore necessary to check whether the shortest path is a relevant concept associated to the process under study (such as energy fluxes in food webs) or if it is more questionable (e.g., paths in functional networks may not actually correspond to information flow [19], paths in contact networks may not be relevant when information or disease diffusion is not studied [23]). Another example consists in the analysis of directed and/or weighted networks with extensions of metrics to this case. It is important to note that the formula of the weighted degree accounts for two effects: how many *neighbours* and how large the weights are, two effects that are impossible to disentangle (a weighted degree of 2 can correspond to a single edge of weight 2 or four edges of weights 0.5). A similar problem can also be raised for the weighted path (potential pitfalls highlighted in [24]). Lastly, global metrics are often used to compare networks (networks measured from different data or conditions, or simulated networks as mentioned in Tip 7). In this case, special care should be taken when comparing values because metrics differences can be a side effect of differences in simple network characteristics such as the number of nodes or edges (see common pitfalls mentioned in [25] for brain networks and a discussion on co-variation of metrics with characteristics of ecological networks in [26]). For instance, modularity, number of modules and network size are known to be intertwined [27].

It is not unusual that authors, instead of choosing a given metric adapted to a particular question, compute a high number of metrics among the available ones. However, many metrics are correlated (see a correlation study in [23]) and it becomes necessary to deal with this redundancy to interpret the results (e.g., with an ordination method [28]). This approach is not hypothesis-driven as recommended in Tip 1 and can undeniably be replaced by an incremental approach where metrics are selected one at a time for their ability to check particular hypothesis associated to the fundamental questions on the data (as for many statistical analysis, see rule 5 in [4]).

Tip 6: Avoid blind use of clustering methods, check their difference instead

With the data avalanche arising this decade, leading to larger networks, clustering has become one of the most popular tool to get a comprehensive view of the network structure. Its general purpose is to aggregate nodes into *clusters* in order to identify a *meso-scale* structure in the network (i.e. zooming out the network). Choosing a network clustering raises similar issues than for choosing a network metric (Tip 5). It is much more than using one of the functions available in a software. Indeed, likewise any clustering methods, the ones dedicated to networks aim at gathering similar objects (i.e. nodes) and thus relies on a specific definition of node similarity. What does the analyst want to be similar in a network? Discussing the pros/cons of the different methods is beyond the scope of this article whereas a massive literature on the topic exists (see Tip 9). However, we illustrate the impact of choosing a specific definition for node similarity with three classical proposals (among others).

A first and natural definition for the similarity between nodes is the existence of a connection between them. Based on this definition, network clustering consists in finding a *modular* structure, i.e. identifying dense clusters of nodes (also called *modules* or *communities*) poorly connected with others. *Community detection* methods [22]

implement this approach, which implicitly assumes the existence of modules in the network. They were successfully applied in many studies in Biology (for instance to identify chromatin domains [29]). A second approach considers that two nodes are similar when they tend to be connected (or unconnected) with the same type of nodes. Hence, species in a food web are considered similar if they have similar preys and predators [30]. This definition can accommodate networks with non-modular structure [31] since it assumes that the nodes are involved in a "diversity of meso-scale architectures" [32]. The *stochastic block model* (SBM) is a popular method based on this definition [31, 33], which has shown to be relevant for the analysis of some biological networks (to highlight the complex architecture of connectomes [32] or functional groups in ecological networks [34]). One important feature is that it allows to model explicitly edge directions and weights by means of different statistical distributions [11]. A third approach consists in associating a vector of characteristics to each node and then to gather nodes with similar characteristics. This includes *motifs*-based approaches [35] and a wide range of innovative *node's embedding* techniques [36, 37]. Nodes are described as points in a space with reasonable low dimension, which allows to apply the huge variety of existing clustering methods for multivariate data. It is important to realize that each of these similarity concepts naturally results in different nodes clustering. The choice between these alternatives must be driven by biological questions not by their availability in software tools (Tip 1).

Tip 7: Don't choose the easy way when simulating networks

To highlight the properties specific to an observed network (for instance a peculiar metric value), a common practice consists in comparing with simulated networks. These properties are detected as *significant* deviation (or not) from a *typical* behaviour implemented in simulated networks. However, there is no generic definition of a typical network and, as a consequence, the features that can be detected depends dramatically on the *null model* used to simulate networks. This null model must be chosen for a given purpose, fitting expected behaviours whereas contrasting those we are interested in. In other words, it must fit the data reasonably well to avoid numerous false discoveries, but not too well so that deviations can emerge.

A natural option could consist in selecting a null model among the series of *random graph models* (e.g., *Erdös-Rényi, small-world, scale-free, SBM, Exponential Random Graph, configuration model*). However, we recommend not to use them too hastily because they are often too general. For example, the Erdös-Rényi model (all edges independent and having the same probability of occurrence) is often a poor null model to detect nodes having an unexpectedly high degree. Indeed, it induces a Poisson degree distribution which is so far from the one observed in most networks that many nodes appear to be unexpectedly connected. On the opposite, no node can display an unexpectedly high degree with respect to the configuration model, as this null model precisely fits to the degree of each node. Moreover, the analyst is usually aware of a series of properties that should be displayed by a simulated network: imbalanced degree distribution, different nodes' roles associated with available side information, forbidden interactions (e.g., depending on body-mass in food webs [38]), etc. Such expected properties must be encoded in the simulation process (for instance a fixed degree sequence [34]), otherwise they will emerge and be detected as significant, or contribute to detect false significant effects as side effects. As an example, when assessing whether the number of feed-forward loops is unexpected in a given transcription network, the simulation procedure must rely on fixed number of nodes and degrees whereas the

number of these loops remains free.

Lastly, when the network under study is not directly observed but built from raw data interpretation, it can be relevant to simulate the whole construction process. Consider the case of contact networks inferred from movement data [23]: one can either simulate trajectories keeping some properties of the original data and then build a contact network, or directly simulate a "realistic" contact network. The former approach will intrinsically account for the uncertainties and biases induced by the construction steps, which are likely to be overlooked by the later approach.

Tip 8:   Reconsider the data to build multiple network layers

A network object can be the result of data aggregation. Indeed, interactions are often observed at different times, locations or for different conditions. You are therefore strongly urged to keep in mind (and at hand) the different *layers* of data (time, space, type,...) and consider networks composed by multiple layers, because *multilayer networks* can provide new insights compared to an aggregated one [39–41].

A network is called *dynamic* when it gathers a time series of network snapshots corresponding to successive rounds of data collection (the nodes' list possibly varying in time). In this case, the temporal variability of the network structure can be assessed (e.g., rewiring of interactions or changes in network metrics over time) and extensions of the concepts developed in Tip 6 now exist in the dynamic case [42, 43]. For instance, the dynamics of animal social structure can be inferred from dynamic networks to enhance the understanding of disease transmission [44]. On another hand, interactions can be observed at different spatial locations. In ecology, they are often aggregated in a *metanetwork* (or *metaweb* [45]) to study how the local networks differ from this metanetwork and explain these variations with environmental factors. In these two cases, multiple layers allows to describe a network as an evolving object and the analysis aim to identify the spatio-temporal variations of interactions and their drivers.

Different kind of interactions can also be observed between nodes. Stacking layers representing molecular interactions in different human tissues [46] or mapping extrasynaptic and synaptic connectomes [47] leads to a *multiplex* network: between any two nodes, there possibly exist more than one edge, one per interaction type at most (often visualized with different colors). Taking jointly into account the different layers enhances the understanding of the nodes' interplay. For instance, using jointly trophic and non-trophic interactions enhances the definition of species ecological roles compared to the use of single layers independently [34]. Finally, it is also possible to integrate different layers of information with different sets of nodes for each layer, such as proteins and chemical compounds [48]. In this case, different kind of interactions are defined inside and between layers. In all these cases, different information layers are integrated into a comprehensive network such that they are treated jointly rather than one after the other.

Tip 9:   Dive into the network literature, beyond your discipline

Network science has emerged "at the dawn of the 21st century" [49]. It now involves a hyper-active community of researchers from different domains such as Physics, Statistics, Computer Science or Social Science. As a result, a massive literature on networks exists and it is challenging for biologists to dive into it. Indeed, we are not

used to explore the bibliography outside our research domain. Reference books [5, 41, 49, 50] and reviews [22, 39, 51] are obviously good entry-points for developing your network skills. However, without any doubt you will highly benefit from a round trip in this literature exogenous to your field, provided that you make the effort to learn the appropriate vocabulary of this area.

Conclusion

The 9 tips presented here should be a way for the data analyst to get a foot in the door of network data analysis. These tips are not exclusive and we are aware of other network-based questions that deserve a special interest, including diffusion on networks for instance. Still, the network non-specialist must be confident in his ability to learn, step by step, the network concepts and methods with a productive effect on his scientific questions.

Acknowledgements

References

1. T. Ideker and R. Nussinov, "Network approaches and applications in biology," *PLoS computational biology*, vol. 13, no. 10, p. e1005771, 2017.

2. M. Zitnik, R. Sosi, and J. Leskovec, "Prioritizing network communities," *Nat Commun*, vol. 9, no. 1, p. 2544, 2018.

3. B. Wang, A. Pourshafeie, M. Zitnik, J. Zhu, C. D. Bustamante, S. Batzoglou, and J. Leskovec, "Network enhancement as a general method to denoise weighted biological networks," *Nat Commun*, vol. 9, no. 1, p. 3108, 2018.

4. R. E. Kass, B. S. Caffo, M. Davidian, X.-L. Meng, B. Yu, and N. Reid, "Ten simple rules for effective statistical practice," 2016.

5. R. Diestel, "Graph theory." `http://diestel-graph-theory.com`, 2016.

6. A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the national academy of sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.

7. P. P. Staniczenko, J. C. Kopp, and S. Allesina, "The ghost of nestedness in ecological networks," *Nature communications*, vol. 4, p. 1391, 2013.

8. G. A. Pavlopoulos, P. I. Kontou, A. Pavlopoulou, C. Bouyioukos, E. Markou, and P. G. Bagos, "Bipartite graphs in systems biology and medicine: a survey of methods and applications," *GigaScience*, vol. 7, no. 4, p. giy014, 2018.

9. V. Miele, F. Picard, and S. Dray, "Spatially constrained clustering of ecological networks," *Methods in Ecology and Evolution*, vol. 5, no. 8, pp. 771–779, 2014.

10. P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, "Uncovering space-independent communities in spatial networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 19, pp. 7663–7668, 2011.

11. M. Mariadassou, S. Robin, C. Vacher, *et al.*, "Uncovering latent structure in valued graphs: a variational approach," *The Annals of Applied Statistics*, vol. 4, no. 2, pp. 715–742, 2010.

12. M. A. Fortuna, R. G. Albaladejo, L. Fernández, A. Aparicio, and J. Bascompte, "Networks of spatial genetic variation across species," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 19044–19049, 2009.

13. D. Hric, R. K. Darst, and S. Fortunato, "Community detection in networks: Structural communities versus ground truth," *Physical Review E*, vol. 90, no. 6, p. 062805, 2014.

14. R. Rossi and N. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

15. G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, "Ten simple rules for reproducible computational research," *PLoS Comput. Biol.*, vol. 9, no. 10, p. e1003285, 2013.

16. M. A. Carey and J. A. Papin, "Ten simple rules for biologists learning to program," *PLoS Comput. Biol.*, vol. 14, no. 1, p. e1005871, 2018.

17. J. W. Tukey, *Exploratory data analysis*. Reading: Addison-Wesley, 1977.

18. M. Krzywinski, I. Birol, S. J. Jones, and M. A. Marra, "Hive plots—rational approach to visualizing networks," *Briefings in Bioinformatics*, vol. 13, no. 5, pp. 627–644, 2011.

19. M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

20. J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen, "The nested assembly of plant–animal mutualistic networks," *Proceedings of the National Academy of Sciences*, vol. 100, no. 16, pp. 9383–9387, 2003.

21. R. E. Van Dijk, J. C. Kaden, A. Argüelles-Ticó, D. A. Dawson, T. Burke, and B. J. Hatchwell, "Cooperative investment in public goods is kin directed in communal nests of social birds," *Ecology letters*, vol. 17, no. 9, pp. 1141–1148, 2014.

22. S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.

23. D. R. Farine and H. Whitehead, "Constructing, conducting and interpreting animal social network analysis," *J Anim Ecol*, vol. 84, no. 5, pp. 1144–1163, 2015.

24. A. Costa, A. M. M. Gonzalez, K. Guizien, A. M. Doglioli, J. M. Gomez, A. Petrenko, and S. Allesina, "Ecological networks: Pursuing the shortest path, however narrow and crooked," *bioRxiv*, p. 475715, 2018.

25. B. C. Van Wijk, C. J. Stam, and A. Daffertshofer, "Comparing brain networks of different size and connectivity density using graph theory," *PloS one*, vol. 5, no. 10, p. e13701, 2010.

26. L. Pellissier, C. Albouy, J. Bascompte, N. Farwig, C. Graham, M. Loreau, M. A. Maglianesi, C. J. Melián, C. Pitteloud, T. Roslin, *et al.*, "Comparing species interaction networks along environmental gradients," *Biological Reviews*, vol. 93, no. 2, pp. 785–800, 2018.

27. S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 36–41, 2007.

28. S. Kortsch, R. Primicerio, M. Aschan, S. Lind, A. V. Dolgov, and B. Planque, "Food-web structure varies along environmental gradients in a high-latitude marine ecosystem," *Ecography*, vol. 42, no. 2, pp. 295–308, 2019.

29. H. K. Norton, D. J. Emerson, H. Huang, J. Kim, K. R. Titus, S. Gu, D. S. Bassett, and J. E. Phillips-Cremins, "Detecting hierarchical genome folding with network modularity," *Nature methods*, vol. 15, no. 2, p. 119, 2018.

30. S. Allesina and M. Pascual, "Food web models: a plea for groups," *Ecol. Lett.*, vol. 12, no. 7, pp. 652–662, 2009.

31. M. E. Newman and E. A. Leicht, "Mixture models and exploratory analysis in networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 23, pp. 9564–9569, 2007.

32. R. F. Betzel, J. D. Medaglia, and D. S. Bassett, "Diversity of meso-scale architecture in human and non-human connectomes," *Nature Communications*, vol. 9, no. 1, p. 346, 2018.

33. J.-J. Daudin, F. Picard, and S. Robin, "A mixture model for random graphs," *Statistics and computing*, vol. 18, no. 2, pp. 173–183, 2008.

34. S. Kéfi, V. Miele, E. A. Wieters, S. A. Navarrete, and E. L. Berlow, "How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience," *PLoS Biol.*, vol. 14, no. 8, p. e1002527, 2016.

35. D. B. Stouffer, M. Sales-Pardo, M. I. Sirer, and J. Bascompte, "Evolutionary conservation of species' roles in food webs," *Science*, vol. 335, no. 6075, pp. 1489–1492, 2012.

36. B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, (New York, NY, USA), pp. 701–710, ACM, 2014.

37. A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, ACM, 2016.

38. U. Brose, T. Jonsson, E. L. Berlow, P. Warren, C. Banasek-Richter, L.-F. Bersier, J. L. Blanchard, T. Brey, S. R. Carpenter, M.-F. C. Blandenier, *et al.*, "Consumer–resource body-size relationships in natural food webs," *Ecology*, vol. 87, no. 10, pp. 2411–2417, 2006.

39. S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014.

40. S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi, "The multilayer nature of ecological networks," *Nature Ecology & Evolution*, vol. 1, no. 4, p. 0101, 2017.

41. G. Bianconi, *Multilayer Networks: Structure and Function*. Oxford university press, 2018.

42. C. Matias and V. Miele, "Statistical clustering of temporal networks through a dynamic stochastic block model," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 79, no. 4, pp. 1119–1141, 2017.

43. G. Rossetti and R. Cazabet, "Community discovery in dynamic networks: a survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, p. 35, 2018.

44. D. Farine, "The dynamics of transmission and the dynamics of networks," *Journal of Animal Ecology*, vol. 86, no. 3, pp. 415–418, 2017.

45. M. Ohlmann, V. Miele, S. Dray, L. Chalmandrier, L. O'Connor, and W. Thuiller, "Diversity indices for ecological networks: a unifying framework using Hill numbers," *Ecology letters*, 2019.

46. M. Zitnik and J. Leskovec, "Predicting multicellular function through multi-layer tissue networks," *Bioinformatics*, vol. 33, no. 14, pp. i190–i198, 2017.

47. B. Bentley, R. Branicky, C. L. Barnes, Y. L. Chew, E. Yemini, E. T. Bullmore, P. E. Vértes, and W. R. Schafer, "The multilayer connectome of caenorhabditis elegans," *PLoS computational biology*, vol. 12, no. 12, p. e1005283, 2016.

48. A. J. Berenstein, M. P. Magariños, A. Chernomoretz, and F. Agüero, "A multilayer network approach for guiding drug repositioning in neglected diseases," *PLoS neglected tropical diseases*, vol. 10, no. 1, p. e0004300, 2016.

49. A.-L. Barabási, "Network science." `http://networksciencebook.com`, 2019.

50. M. Newman, *Networks*. Oxford university press, 2018.

51. P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Systems*, vol. 151, pp. 78–94, 2018.