Descriptive Analysis of Network Structures

Deserno, M.K.; Isvoranu, A.-M.; Epskamp, S.; Blanken, T.F.

DOI
10.4324/9781003111238-4

Publication date
2022

Document Version
Final published version

Published in
Network Psychometrics with R

Chapter 3

Descriptive Analysis of Network Structures

Marie K. Deserno[1], Adela-Maria Isvoranu[2], Sacha Epskamp[1,3], & Tessa F. Blanken[2]

1. Max Planck Institute for Human Development, Berlin
2. University of Amsterdam, Department of Psychology
3. University of Amsterdam, Centre for Urban Mental Health

3.1 Introduction

This chapter introduces networks, how they can be constructed, and how they can subsequently be analyzed. Networks consist of nodes (also called *vertices*) and edges (also called *links*). Often in the literature, networks can also be called *graphs*, though the former generally refers to the visual representation, while the latter generally refers to the mathematical notation. Virtually anything can be represented as a set of nodes and edges. For example, Figure 3.1 displays interactions between characters of a book series. In real life we encounter networks constantly: transportation networks consisting of train stations (nodes) and railways between them (edges); lights (nodes) that are connected through wires (edges); or internet cables (edges) that connect homes (nodes) to the internet. Networks can also consist of edges that are not directly visible: friends (nodes) that are connected through telephone calls (edges); or a social network consisting of scientists (nodes) that are connected through co-authorships (links). Finally, and central to this book, networks can consist of estimated statistical relationships (edges) among variables (nodes).

Cite this chapter as:

Deserno, M. K., Isvoranu, A. M., Epskamp, S., & Blanken, T. F. (2022). Chapter 3. Descriptive analyses of network structures. In Isvoranu, A. M., Epskamp, S., Waldorp, L. J., & Borsboom, D. (Eds.). *Network psychometrics with R: A guide for behavioral and social scientists*. Routledge, Taylor & Francis Group.

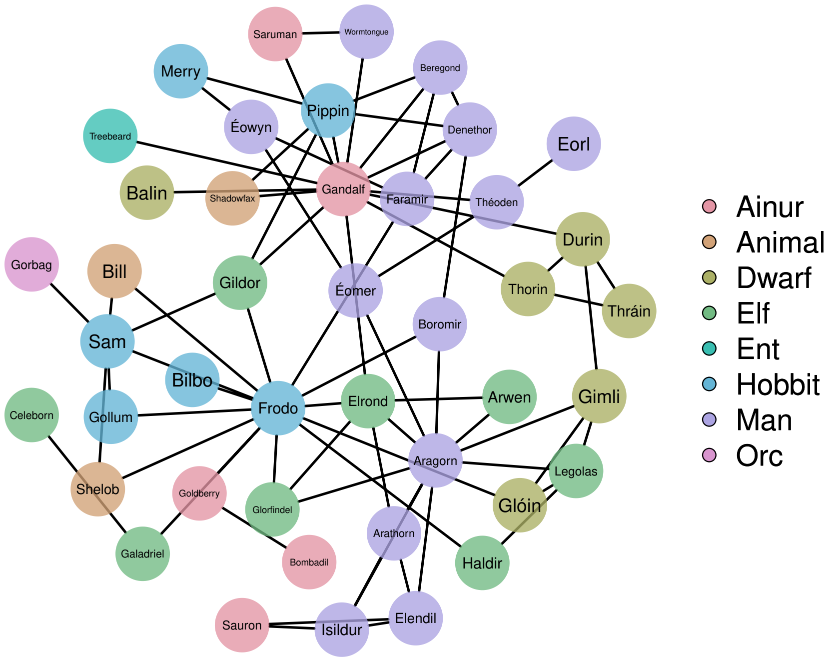

Figure 3.1. A network depicting character interactions in the *Lord of the Rings* book series: nodes represent characters, and edges represent that these characters were often mentioned together in the same paragraph. Based on the work of Calvo Tello (2016). The data can be downloaded from Github at https://github.com/morethanbooks/projects/tree/master/LotR.

The chapter starts with a description of the origin and context in which network analysis is most applied. Next, we describe the differences between the setting in which we introduce network analysis tools in this chapter, *Network Science*, compared to the setting which is the focus of the remainder of this book, *Network Psychometrics*. Subsequently, we discuss different types of networks, how they can be encoded, and several of the most prominent methods in which networks are analyzed. We conclude this chapter with an introduction to network analysis of psychological variables, which will form the core of the remainder of this book.

3.2 Complex systems and network science

Increasingly, scientific questions are targeted at the structure and evolution of complex systems. Theoretical advances lie in the conceptual integration of complex and multivariate mechanisms. As such, many of the big scientific questions of this century are rooted in a general interest in patterns within such complex system. Now, more than ever, the availability of multi-scale and multi-modal data and powerful computing capacity enable scientists to empirically test theories about how components of a complex system, such as cells, human beings, or species, interact. These advances have led to a rapidly emerging field spanning a wide range of applied disciplines: *Network Science* (Barabási, 2012).

The fundamental study of networks is by no means new. Graph theory—dating back to the solution of the Königsberg bridge problem by Euler (1736)—has been a substantial part of mathematics and adopted by the social sciences for most of the twentieth century. Recent years, however, have witnessed an unprecedented fast-paced expansion and application of the mathematical framework: ecological networks of food webs or ecosystems in ecology (Ings et al., 2009), semantic association networks in linguistics (Steyvers & Tenenbaum, 2005), social networks in sociology (Scott, 1988), and symptom networks in psychology (Borsboom, 2017). All of these networks display characteristics, organized patterns emerging from the system's diverse interconnectedness. Fuelled by the substantial questions and data from different fields, not only the network toolbox, but also network theory is rapidly expanding. There are multiple applied examples of the complexity-credo: "The whole is more than the sum of its parts." Primarily, because modern network approaches provide fundamental insights into the dynamics and emergent properties that result from the interaction of simple elements. As such, the flocking of birds, for example, produces a swarm with collective intelligence and the collective actions of nerves create consciousness—all through multi-scale network interactions. Such complex phenomena can best be understood by studying the integrative dynamics and functions of the system that produces it.

The translation of this mathematical framework into network analysis techniques has allowed scholars from different scientific fields to formalize and apply their hypotheses grounded in system-focused thinking. For example, Watts and Strogatz (1998) generalized the basic idea of network structures to a variety of data sets and demonstrated that many real-world networks display similar properties (see Section 3.5). Despite the differences in what networks represent across disciplines, the network analytical framework allows to describe and predict the inner workings of the system's interconnectedness. During the past decades, scholars across different scientific fields have invented a new language to communicate and formalize their theories grounded in systems-focused thinking and analysis. With a focus on the predictability of patterns, analytic schools based on differential equations and network analysis have advanced the translation of systems-thinking into toy models and analytic toolboxes. Its empirical application has grown even more, after Watts and Strogatz (1998) generalized the basic idea of network structures to a variety of data sets. What distinguishes network analysis from previous attempts at accommodating complex systems theory is its data-inspired methodology. From early on, network analysis has been rooted in empirical evidence for the system properties its theory predicts: the changing distribution of connections when more variables are added to the network, for example, has been reported across a diverse range of disciplines and their data. The decades-old questions fuelled by complex systems theory have been reassembled by the toolbox of network science. With its strong empirical basis, network science has allowed researchers to develop theoretical bits and pieces about not only effects and workings of complex systems and their interconnectedness but also about individual nodes and links. In other words, the rapidly advancing network methodology has become indispensable in the study of complex systems, relating local dynamics and properties to global effects and structures, and vice versa.

3.3 From network science to network psychometrics

In many networks, nodes are well-defined entities. In a transportation network, for example, it is clearly defined what makes up a node (e.g., a train station). Two nodes can be easily separated from each other, and what constitutes the network can be clearly defined. If we were to map out the subway system in London, for example, it is clear that we (a) consider subway stations; and (b) consider the subway stations that are situated in London, not those in Glasgow. These clear distinctions and definitions, however, are not a given. In the psychological networks discussed in this book, it is not always clear (a) what a node entails, and (b) what the scope of a network should be. Rather than nodes representing *entities* (e.g., people, cities, species), in such psychological networks nodes typically represent *variables* that can take more than one state (e.g., absence or presence of a psychological symptom). Moving away from representing entities as nodes, it may be unclear what the nodes then should represent. Constructing a network of depression, for example, one may choose to include ‘symptoms’ as defined in nosologies and assessed by a clinician, or to include items of a self-reported depression questionnaire. Similarly, one may choose to restrict to depression symptoms only, or to consider anxiety symptoms important to the network of depression as well.

The same holds for the edges in a network. These can be clearly defined, for example when they are directly observed (e.g., rails in a transportation network) or follow a clear rule (e.g., co-authorships), but may be less straightforward when they represent statistical relationships between variables that need be estimated from data. Edges representing statistical relationships bring several complications to the interpretation of the network. For example, the interpretation of the interaction between connected nodes depends on the type of statistical relationship that is estimated. These edges now no longer denote ways in which two entities can interact with one another (e.g., two friends meeting and potentially infecting one another with a virus), but instead represent probabilistic association (e.g., the presence of symptom A increases the probability of the presence of symptom B). Estimating edges from data through the use of statistics also adds a layer of complexity in the form of sampling variation and uncertainty in the estimated values—a topic discussed in more detail in Chapter 8.

Because there can be such striking differences between what networks encode, it is important to make the setting in which network analysis is applied explicit. This chapter introduces some of the core methods of analyzing networks in the field of *Network Science*. Network Science is a very broad field of research on network structures, but mostly relates to networks in which nodes and edges are quite well defined (often clear entities linked through observable or well-defined links). Statistical modeling in this field often takes the edges to be the variables, which can take, for example, the values ‘present’ and ‘absent.’ The remainder of this book will focus on the estimation of network models from data, in which not the edges but rather the nodes are the variables (e.g., symptoms).

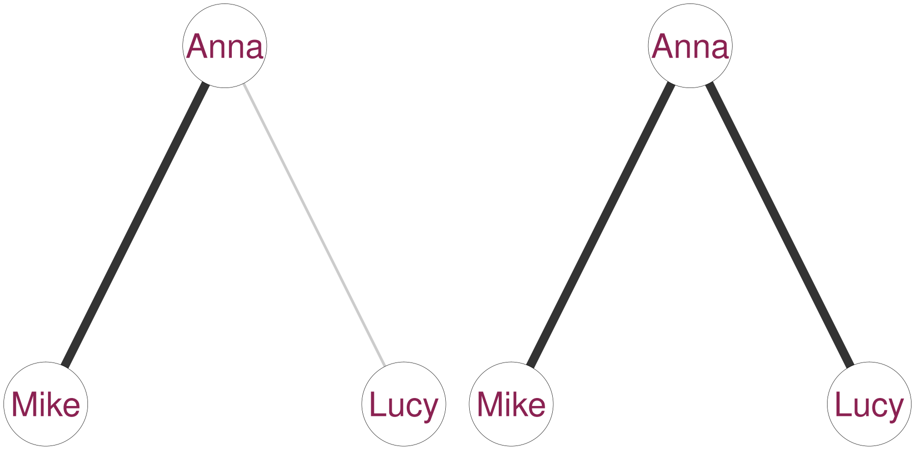

Figure 3.2. The difference between weighted (left) and unweighted (right) networks. In weighted networks, the strength of connections can differ across edges in the network. For example, Anna could interact more with Mike than with Lucy, and as such if Anna is infected with a virus she is more likely to infect Mike than Lucy.

The field of study that is concerned with the estimation of such networks from data is termed *Network Psychometrics*. While this difference between networks typically analyzed in network science and networks typically obtained in network psychometrics may seem subtle, it substantively changes the interpretation of what a network is. Methods from network science can be, and routinely are, applied to networks estimated from data, but doing so takes a certain leap of interpretation. To this end, it is vital to learn about these metrics in the setting in which they were developed, and keep a critical eye to when they are being applied outside of this scope.

3.4 Constructing networks

Constructing a network starts with the selection of a set of nodes which represent certain entities that are connected in some way through edges. The edges in a network structure can be either *weighted* or *unweighted*. Let's take the basic example in Figure 3.2, assuming the edges in the network represent *friendship*. In the network structure on the left side, where we have a weighted network, we would take the interpretation that Anna and Mike are better friends than Anna and Lucy, as the edge between Anna and Mike is stronger. In the network structure on the right side, where we have an unweighted network, we would take the general interpretation of friendships between people, without a focus on whether some people are better friends than others: Anna and Mike are friends, Anna and Lucy are also friends, and there is no friendship relationship between Mike and Lucy.

Further, edges can have both a *sign* and a *direction*. Throughout this book, as well as in many of the software packages that we will use in this course, positive links will be by

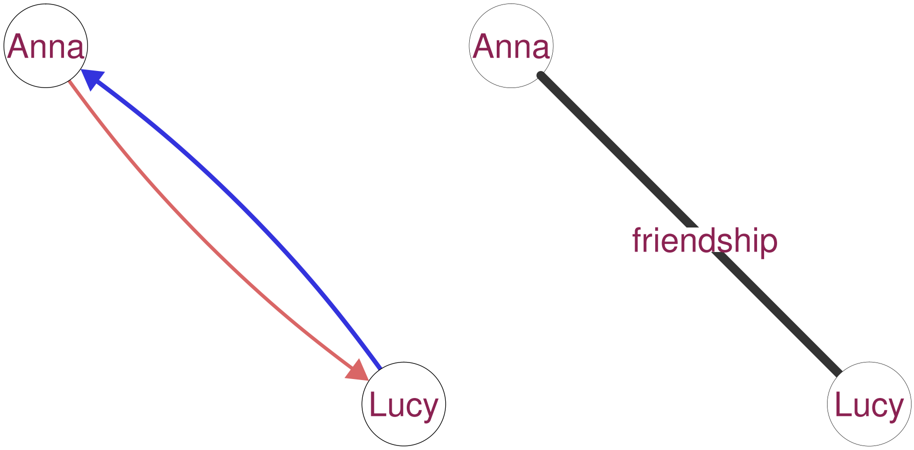

Figure 3.3. The difference between weighted directed (left) and unweighted undirected (right) networks. In addition to differing in the strength of connections, edges can also differ in the direction of the edge. For example, a relationship from *A* to *B* may be of a different strength or absent entirely in comparison with a relationship from *B* to *A*. In this example: Lucy likes Anna, but Anna does not like Lucy.

default represented by *blue edges* and negative links will be by default represented by *red edges*.[1] Let's take the basic example in Figure 3.3, assuming the edges in the network still represent *friendship*. In the network structure on the left side, we would take the interpretation that Anna and Lucy experience their relationship differently: While Lucy really likes Anna and thinks of her as a friend, Anna doesn't really like Lucy and does not reciprocate the relationship. In the network structure on the right side, where we have no direction or sign, the edge would simply represent a friendship relation between Anna and Lucy.

With the basic building blocks of a network at hand, the network can now be represented mathematically, as is further discussed in Technical Box 3.1. Such mathematical representations can be used to encode a large variety of network structures, such as large and small networks, weighted and unweighted networks, and directed and undirected networks. Throughout this book, we will mainly focus on undirected weighted networks estimated from cross-sectional data (Part II), and both directed and undirected weighted networks estimated from longitudinal data (Part III). In the following sections, we will discuss how networks can further be analyzed.

3.5 Analyzing networks

Network construction or estimation is typically followed by network inferences to quantify or summarize the network's interconnectedness. To this end, different metrics can be

[1]Some other color codings are used in the literature. Older papers often use green for positive effects instead of blue, and some papers use red for positive edges and blue for negative edges.

Mathematically, a *graph* (or network) can be defined as a set G that consists of a set of *vertices* (nodes) V and a set of *edges* (links) E:

$$G = \{V, E\}\,.$$

For example, the network on the right of Figure 3.2 could be represented with the following sets:

$$\begin{aligned} V &= \{\text{Mike}, \text{Anna}, \text{Lucy}\} \\ E &= \{(\text{Mike — Anna}), (\text{Anna — Lucy})\}\,. \end{aligned}$$

A typical way of encoding the network mathematically is through the use of an *adjacency matrix*, $\boldsymbol{A}$, which contains a row/column for each node, and in which a 1 indicates two nodes are connected and a 0 indicates two nodes are not connected:

$$\boldsymbol{A} = \begin{array}{r} \scriptstyle\text{Mike} \\ \scriptstyle\text{Anna} \\ \scriptstyle\text{Lucy} \end{array} \overset{\begin{array}{ccc} \scriptstyle\text{Mike} & \scriptstyle\text{Anna} & \scriptstyle\text{Lucy} \end{array}}{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}$$

If the network is *directed*, the adjacency matrix could be asymmetrical, in which case the rows indicate the node of origin and the columns the node of destination. For *weighted* networks, additional weights can be added to every edge in the graph. These can be encoded through a *weights matrix* $\boldsymbol{W}$, in which 0 indicates no strength of connection. For example, suppose the left network in Figure 3.2 represents the number of interactions in a week. Then the corresponding weights matrix could be:

$$\boldsymbol{W} = \begin{array}{r} \scriptstyle\text{Mike} \\ \scriptstyle\text{Anna} \\ \scriptstyle\text{Lucy} \end{array} \overset{\begin{array}{ccc} \scriptstyle\text{Mike} & \scriptstyle\text{Anna} & \scriptstyle\text{Lucy} \end{array}}{\begin{bmatrix} 0 & 4 & 0 \\ 4 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}$$

Typically, if a weights matrix is supplied an adjacency matrix is no longer needed, as a weight of 0 already indicates the absence of an edge. To this end, the weights matrix alone is typically enough to encode a network. Occasionally, the weights matrix is also called the (weighted) adjacency matrix.

Technical Box 3.1. Mathematical encoding of networks.

computed to make inferences about the structure of the network, or regarding the role that specific nodes have within the network. Roughly, these metrics can be divided into ‘local’ and ‘global’ metrics. Local metrics concern specific parts of the network, whereas global metrics relate to the network as a whole.

In network inference, it is again important to keep in mind what the networks represent, as well as the research question that is being asked. For example, when interested in the overall network structure, the appropriate metric to quantify this structure may be different for transportation than for psychological networks. Consequently, within different contexts (e.g., social networks, brain networks, psychology networks), different metrics have been proposed. Metrics originally proposed in relation to e.g., social networks can be informative for psychological networks too. Yet, it is important to critically consider how the metric is computed, and what the interpretation for your network could be. The metrics currently used to describe psychological networks are by no means set in stone but rather part of an ongoing discussion. As there are many metrics available, here we only list a selection. This list is thus by no means exhaustive, and the purpose is not to direct to all relevant metrics out there. Rather, the goal is to familiarize with the different available questions that can be asked about network properties, and match them with examples of metrics that pick up on those properties.

Local network properties

Research questions can relate to the role that specific nodes have in the estimated network. Initially, many of the local inferences made in psychological networks were based on centrality metrics developed in social network science, of which strength, closeness, and betweenness are most common (Opsahl et al., 2010). The computation and interpretation of these metrics are explained in the sections below. However, to quantify the role of specific nodes, there are many other metrics to turn to, of which some are specifically developed in the context of psychological networks, for example: expected influence (Robinaugh et al., 2016), predictability of nodes (Haslbeck et al., 2021), stabilizing and communicating nodes (Blanken et al., 2018). Alternatively, existing metrics can be adopted and adapted to suit specific study hypotheses (e.g., Letina et al., 2019). Different metrics will shed different light on the role a node might have, and as such can be used to answer different research questions.

Centrality

Centrality measures can be used to estimate the position and role of a node in a network. As detailed further in Chapter 4, a two-dimensional representation of the network cannot judge how 'central' a node is in the network: nodes placed in the center of the visualization are not necessarily important to the network architecture or well connected. Centrality metrics aim to quantify how 'central' a node is in the network. While many centrality metrics exist, we only focus here on a select few commonly used centrality metrics that have been generalized to weighted networks and directed networks (Opsahl et al., 2010; Robinaugh et al., 2016). These metrics can roughly be divided into two categories: metrics for directed connectivity, and metrics for indirect connectivity. For example, in the network $A — B — C$, a centrality metric of direct connectivity of the node A would only take its immediate connection to node B into account, whereas a centrality metric of indirect connectivity would also take its indirect connection to C (via B) into account.

Centrality metrics for direct connectivity. Metrics for direct connectivity only investigate the immediate connections of a node. The most common metrics of direct connectivity are *node degree*, tallying the number of connected nodes regardless of their weight, *node strength*, summing the absolute edge weights connected to a node, and *one-step expected*

Opsahl et al. (2010) describe several generalizations of common centrality metrics for weighted graphs. These all use a tuning parameter α that can be used to place more emphasis on the presence/absence of connection in $\boldsymbol{A}$ ($\alpha = 0$) and the strength of connection in $\boldsymbol{W}$ ($\alpha = 1$). Unweighted variants can be obtained by setting $\boldsymbol{W} = \boldsymbol{A}$ or $\alpha = 0$. For all metrics below, let n be the number of nodes and i a node of interest.

One of the main metrics used in network science is the *node degree*, which counts the number of connected edges to a node (in this case, node i):

$$\text{Degree}(i) = \sum_{j=1}^{n} a_{ij},$$

in which a_{ij} is the element at row i and column j of the adjacency matrix $\boldsymbol{A}$. The typical weighted variant is termed *node strength*, which sums over the absolute weights matrix instead of the adjacency matrix:

$$\text{Strength}(i) = \sum_{j=1}^{n} |w_{ij}|,$$

in which w_{ij} is the element at row i and column j of the weights matrix $\boldsymbol{W}$. Opsahl's variant of node degree/strength is as follows:

$$\text{WeightedDegree}(i) = \sum_{j=1}^{n} a_{ij}^{1-\alpha} |w_{ij}|^{\alpha},$$

which reduces to the degree with $\alpha = 0$ and strength with $\alpha = 1$.
An alternative measure to node strength is one-step *expected influence*, which is virtually the same as node strength except that no absolute value is used (Robinaugh et al., 2016):

$$\text{ExpectedInfluence}_1(i) = \sum_{j=1}^{n} a_{ij} w_{ij}$$

Technical Box 3.2. Explanation of how to calculate centrality metrics that encode strength of direct connectivity.

influence, summing the edge weights connected to a node without taking the absolute value. If a node ranks high on metrics of direct centrality, this implies that it has (many) strong relations to other nodes in the network. In psychological networks, researchers have used these metrics aiming to extract information on the importance of a node for the state of all other nodes. Technical Box 3.2 explains these metrics of direct centrality in more detail.

Centrality metrics for indirect connectivity. In addition to centrality metrics that are based on direct connectivity, there are also several centrality metrics that are based on indirect connectivity. That is, these centrality metrics do not only take into account the nodes a node of interest is immediately connected to (sometimes termed the 'neighbors' of a node), but also the nodes that a node of interest is connected to via other nodes (e.g., neighbors of neighbors). The metric *two-step expected influence*, for example, takes into account every node that is connected in two steps to the node of interest. Other metrics of indirect centrality rely on the concept of *distance* between two nodes: the minimum number of steps it takes to go from one node to another node. In weighted graphs, the distance is determined by the length of each edge, which is typically quantified as the inverse of the absolute edge-weight. The metric of *closeness* quantifies how far removed a node is from every other node in the network, by taking the inverse of the sum of all distances from a node of interest to all other nodes. As such, from a node with a high closeness you can reach all other nodes relatively fast. The *betweenness*

The most important indirect centrality metrics rely on the concept of *shortest path length* or *geodesic distance* between two nodes: what is the shortest distance to go from one node to another node. To compute this, we first need to translate the weight of an edge to a hypothetical length of the edge. Opsahl et al. (2010) describe the following function to define the length of an edge between nodes i and j (including the absolute value operator to make edges positive):

$$\text{Length}(i, j) = \frac{1}{|w_{ij}|^{\alpha}}.$$

For unweighted networks $\boldsymbol{A} = \boldsymbol{W}$ can be used. As such, for unweighted networks (or when $\alpha = 0$) the length of an edge is 1 when the edge is present and ∞ when the edge is absent. For weighted networks, the edge is longer for edges with weaker weights. The shortest path length is then the minimal distance between two nodes:

$$\text{Distance}(i, j) = \min\left(w_{ik} + \ldots + w_{lj}\right).$$

The common centrality metric *closeness* makes use of this metric. Closeness is computed by taking the inverse of the sum of all distances from one node to all other nodes (also termed *farness*):

$$\text{Closeness}(i) = \frac{1}{\sum_{j=1}^{n} \text{Distance}(i, j)}.$$

Another common centrality metric is *betweenness*, which is computed by investigating how many shortest paths go through a node of interest:

$$\text{Betweenness}(i) = \sum_{<j,k>} \frac{\text{\# of shortest paths between nodes } j \text{ and } k \text{ that go through } i}{\text{\# of shortest paths between nodes } j \text{ and } k}.$$

Finally, an alternative metric for indirect connectivity is two-step *expected influence*, which looks at the strength of connection two steps away from a node (Robinaugh et al., 2016):

$$\text{ExpectedInfluence}_2(i) = \sum_{j=1}^{n} a_{ij} w_{ij} + \sum_{j=1}^{n} a_{ij} w_{ij} \sum_{k=1}^{n} a_{jk} w_{jk}$$

Technical Box 3.3. Explanation of how to calculate centrality metrics that encode strength of indirect connectivity.

centrality quantifies how often a node lies on the shortest path connecting any two other nodes, which directly reflects on the extent to which that node funnels the activity or influence between any two sets of nodes (Opsahl et al., 2010). As such, a node with high betweenness is a node that tends to connect clusters of nodes together. Technical Box 3.3 explains these metrics of indirect centrality in more detail.

Bridge centrality. The above described centrality indices each quantify the connectivity of one node to potentially all other nodes in the network. Sometimes, however, it may not be interesting to look at how well a node connects to all other nodes in the network. For example, in symptom networks of multiple psychological disorders, we may expect that symptoms connect well with other symptoms from the same disorder (Epskamp et al., 2017). The topic of interest may be to only investigate how well symptoms connect to symptoms from *other* disorders, indicating that such symptoms may form bridges between clusters of disorders (Borsboom et al., 2011; Cramer et al., 2010). Jones et al. (2021) proposed a set of adaptions for the above described centrality indices that specifically look at connectivity across predefined clusters of nodes—*bridge centrality*. *Bridge strength* and *bridge expected influence* investigate how many and how strong the edges are between a node and nodes from different clusters, *bridge closeness* investigates how easy it is to reach all nodes belonging to other clusters from a particular node, and *bridge betweenness* investigates how often a node lies on the shortest paths between two different clusters.

Positive and negative edges. One topic of interest is how some of the discussed centrality metrics differ is the way in which negative edge-weights are handled when computing the centrality indices. It may not be clear how negative edge-weights should be handled when computing centrality. This is because in most network applications studied in network science, edges can typically have only positive weights. For example, two people cannot interact less than 0 times per week with one another, and the distance between two cities, usually quantified with the inverse of the edge-weight, can at most be ∞, leading to a weight of 0. As we will learn later in this book, psychological networks can feature negative edges. For example, the association between two variables can be negative, implying that higher levels of one variable (e.g., amount of coffee someone drinks in a day) is negatively related to another variable (e.g., quality of sleep). It can be noted that we could arbitrarily re-code variables in such psychological networks to make this association positive. For example, instead of ‘quality of sleep’ a node could represent ‘sleep problems,’ making the association positive. To this end, a typical manner in which negative edges are handled is by taking the absolute value (making negative values positive). This way, a negative effect of, say, -0.2 is interpreted as just as strong as a positive effect of 0.2. This interpretation is used in default computations of node strength, but also closeness and betweenness, as these metrics rely on the definition of the length of an edge, which is typically defined as the inverse of the absolute edge weight. By not taking the absolute value, expected influence interprets negative edges and positive edges to potentially cancel one another out. This interpretation is sensible if the nodes in the network cannot arbitrarily be re-coded. For example, in a network in which nodes represent symptoms, the interpretation of the scales of the nodes is the same for every node (higher values indicate worse symptomatology). This makes negative edges substantively interpretable and not just an artifact due to encoding of the variables represented by the nodes.

Clustering

In addition to centrality metrics, another local property of nodes that is often investigated —although not very often in psychological networks—is the *clustering* of nodes. That is, to what extent do nodes in the network tend to cluster together? The *local clustering coefficient* takes one node of interest as a starting point and quantifies the proportion of nodes that is connected to that specific node and also to each other. In other words, it

The *local clustering coefficient* quantifies how well the nodes connected to a node of interest are also connected to one another. To explain this, we need to make a distinction between *triplets* and *triangles* in a network. A triplet is any sequence of three nodes that are connected in a chain, such as A — B — C. It does not matter for the triplet if A and C are connected as well. A triangle is a set of three nodes that are all connected to one another (e.g., A — B — C — A). As such, every triangle leads to three triplets, each containing one of the three nodes as middle node once: A — B — C, B — C — A, and C — A — B. Let $\tau_{\Delta}(i)$ indicate the number of times node i is in a triangle, and let $\tau_3(i)$ indicate the number of times node i is the middle node of a triplet. The local clustering coefficient can then be computed as follows:

$$\text{Clustering}(i) = \frac{\tau_{\Delta}(i)}{\tau_3(i)}.$$

The *global clustering coefficient* computes how well nodes cluster together on average in a network. One way to compute this is is by simply averaging the local clustering coefficients:

$$\text{Clustering}(G) = \frac{1}{n}\sum_{i=1}^{n}\text{Clustering}(i).$$

A problem with this metric is that it is an average of averages, and does not take into account that some nodes are more often the middle node of a triplet (i.e., are more central with more directed connections). To this end, the *transitivity* metric computes the global clustering directly using the number of triangles, $\tau_{\Delta}(G)$, and the number of triplets, $\tau_3(G)$, that are present in the graph, taking into account that there are three times more potential triplets in the graph than there are triangles:

$$\text{Transitivity}(G) = \frac{3 \times \tau_{\Delta}(G)}{\tau_3(G)}.$$

Of note, the clustering coefficients here are well defined for unweighted networks, but less well defined for weighted networks. While weighted variants exist, they are not without limitations and make interpretation of clustering harder. To this end, we do not discuss these weighted variants in this book.

Technical Box 3.4. Explanation of how to calculate the local and global clustering coefficient.

quantifies the proportion of neighbors of a node that are also neighbors to each other. In contrast to centrality metrics, local clustering can be interpreted as a metric of redundancy of a node. For example, in a social network of friendships, a person (node) with a high local clustering coefficient would have friends that are also friends with each other. The so-called redundancy of this person's role in the network can be illustrated by considering the spread of a virus through their network: a virus can more easily spread in a cluster of people that are all friends with one another, as there are many ways in which two people can infect each other in such a well-connected cluster; if an individual is removed from the network (e.g., vaccinated), others can still easily infect each other. As such, the role of the individual with high local clustering is less important for the spread of the virus than the role of an individual with low clustering. The local clustering coefficient is explained in more detail in Technical Box 3.4.

Global network properties

In addition to metrics quantifying local properties of networks, there are also several metrics that quantify the network architecture as a whole, termed *global network properties*. This architecture is sometimes referred to as the networks' topological structure: the organization of nodes and edges in the network. Knowing the architecture of a network conveys information on how quickly information or activity can travel through the system. Important features of the network architecture are, for example, the amount of clustering in a network (i.e., whether neighboring nodes are also neighbors), the overall connectivity of the network (i.e., how densely or strongly the nodes are connected to each other) or the existence of hubs (i.e., whether some nodes are more interconnected than others).

The network architecture is thought to be a key property of the network itself. For example, in brain networks the architecture is thought to reflect on the efficiency of information processing (Liu et al., 2017), in social networks the structure might reveal something about how gossip spreads through the network (Doerr et al., 2012), and in symptom networks the strength of the associations might reveal something about the vulnerability of the network (van Borkulo et al., 2014). These structures can be assessed using various global metrics, such as the clustering, degree distribution, or the connectivity of the network. There are some well-known network architectures such as *scale-free* or *small-world* networks, whose characteristic properties have been extensively studied. As such, if a network can be classified to follow a certain known network architecture, the research on the behavior of networks with this architecture can become relevant to the topic studied in an applied network analysis. This section discusses some of the most prominent global network properties.

Degree distributions

As discussed above, the most important and well-studied centrality metric is node degree, which tallies the number of connected nodes to a node of interest. A typical analysis for unweighted networks is to determine how node degree is distributed over the network. This concept is termed the *degree distribution*. The degree distribution dictates the expected number of connections per node, as well as the variance around that expected number. For example, a network in which edges are present or absent at random with

There are several metrics that can be used to quantify global connectivity in a network. For these metrics, we first need to know the total number of possible edges in a network. Let m represent the total number of possible edges. This number is then a direct function of the number of nodes n and the type of network studied. For example, for undirected networks that do not feature loops (edges from and to the same node), this number becomes:

$$m = n(n-1)/2.$$

However, if the network is directed and self-loops are included, this number becomes:

$$m = n^2.$$

The most straightforward metric is the unweighted *density* of a network, which gives a proportion of the number of present edges:

$$\text{Density}(G) = \frac{1}{m} \sum_{<i,j>} a_{ij},$$

in which $\sum_{<i,j>}$ denotes the sum over all unique edges: all $n(n+1)/2$ undirected edges in an undirected network or all n^2 directed edges in a directed network.[a] The *sparsity* of a network is simply the proportion of absent edges:

$$\text{Sparsity}(G) = 1 - \text{Density}(G).$$

For weighted networks, a weighted variant of density can be computed. There are several ways in which this is done, but the most common method is to take the average absolute edge-weight:

$$\text{WeightedDensity}(G) = \frac{1}{m} \sum_{<i,j>} |w_{ij}|.$$

Another metric of global connectivity that is often studied is the *average shortest path length* (APL), which is quantified by taking the average of all shortest path lengths (geodesic distance) between each pair of nodes:

$$\text{APL}(G) = \frac{1}{m} \sum_{<i,j>} \text{Distance}(i, j).$$

[a] Of note: the inclusion of self-loops in network density is questionable, and some software packages will only compute density for edges between different nodes by default.

Technical Box 3.5. Description of measures of global connectivity.

some probability, also termed an Erdös-Rényi model, will feature a degree distribution that is binomial, which looks similar to a normal distribution (bell shaped). In such a network, we expect every node to have similar degrees: some nodes will have more connections than others, but we don't expect some nodes to have many more connections than others. Decades of literature investigating the degree distributions of natural network structures have shown that 'real' networks rarely adhere to such a distribution. More common is a *power-law* or exponential distribution, in which some nodes (hubs) have many connections and thus a high degree and many nodes (spokes) have only few connections and thus a low degree. Such networks are termed *scale-free* networks.

Connectivity

In addition to studying average properties of the nodes in a network, it may also be interesting to study average properties of the edges in a network, for example by looking at if a network features many or few connections. Important metrics that are often discussed are the *density* and *sparsity* of a network. The density of a network quantifies the proportion of present connections, whereas the sparsity quantifies the proportion of absent connections. For weighted networks, a weighted density can also be computed by averaging the absolute edge weights. Density plays an important role in simulations of network dynamics, as further discussed in Chapter 13. Sparsity, on the other hand, plays an important role in the performance of network estimation procedures, as further discussed in Chapter 7. A final metric of connectivity is the *average shortest path length* (APL), which takes the average of the shortest distances between every pair of nodes. Technical Box 3.5 introduces these metrics in more detail.

Clustering and small-worldness

In addition to investigating how well nodes cluster around an individual node using local clustering coefficients, it may also be interesting to investigate how well nodes cluster together in a network on average. Technical Box 3.4 discusses the *global clustering coefficient* and *transitivity* of a network, two metrics for determining if nodes tend to cluster together. Watts and Strogatz (1998) describe an interesting relationship between the global clustering in a network and the APL of a network. They noted that in a network generated at random (Erdös-Rényi model), both the APL and the global clustering tends to be low. That is: on average it is easy to reach one node from another node, but nodes also do not tend to cluster together. On the other hand, a network can also be constructed that is highly organized, for example a network in which people only interact with their literal neighbors and neighbors of neighbors. Such a network will have a high clustering, but also a very high APL (it will take a very long time to go from one node to another node on average). Watts and Strogatz (1998) noted that many networks have properties somewhere in between these random and organized networks: networks are overall organized, but may include some random connections. This leads to a network structure in which both APL is low and global clustering is high, which is termed a *small-world* network. The small-world architecture is named after the six-degrees of separation phenomenon; the phenomenon that all people are, on average, only six social contacts away from each other. Such a phenomenon is only possible in a social network with high clustering and short characteristic path-lengths. A technical description of how small-worldness can be quantified can be seen in Technical Box 3.6.

Community detection

A final topic of interest in network analysis is the detection of clusters in a network. When a network is highly clustered, one can aim to identify these clusters using community detection. In community detection, the goal is to identify highly connected clusters (or communities) that exhibit greater connectivity within than between clusters. Statistically, many community detection algorithms have been developed that allow identifying such a structure, see for example Fortunato (2010) for an extensive overview. In psychological network analysis, community detection has become popular through the introduction of *exploratory graph analysis* (EGA; Golino & Demetriou, 2017; Golino & Epskamp, 2017), which utilizes cluster detection on estimated psychological networks to provide an alternative to exploratory factor analysis: to detect a potential latent variable structure underlying the data.

To quantify small-worldness of a network G, the APL and transitivity of that network can be compared to the APL and transitivity of a comparable random network (Erdös-Rényi model) with the same number of nodes and edges, G_R. A network can be said to feature a small world if the network has a comparable APL as a corresponding random network:

$$\text{APL}(G) \approx \text{APL}(G_R), \tag{3.1}$$

but also a much higher clustering than the comparable random network:

$$\text{Transitivity}(G) >> \text{Transitivity}(G_R). \tag{3.2}$$

One way to summarize these quantities is through the *small-world index*:

$$\text{SmallWorld}(G) = \frac{\text{Transitivity}(G)/\text{Transitivity}(G_R)}{\text{APL}(G)/\text{APL}(G_R)}.$$

The denominator of this expression contains the quantities of Expression (3.1), and should be around 1 if that property holds. The numerator of this expression contains the quantities of Expression (3.2), and is therefore expected to be higher than 1 if this property holds. To this end, higher values of the small-world index indicate larger levels of small-worldness. A small-world index larger than 1 indicates that a network holds some small-world properties, although cutoffs of 3 and 6 are also proposed as more strict criteria.

Technical Box 3.6. Description of quantifying small-worldness.

Assuming the weights matrix of an unweighted and undirected network (in this case the *Lord of the Rings* network) is called `W`, the resulting network plot is called `graph`, and differing clusters are encoded in an object called `clusters`, we can use the code below to compute network metrics of interest. Please note that both the *qgraph* and *networktools* R packages need to be installed and loaded for the code to run:

Load packages:

```
library("qgraph"); library("networktools")
```

Number of nodes:

```
ncol(W)
```

Number of edges:

```
sum(W[lower.tri(W)])
```

Density:

```
mean(W[lower.tri(W)])
```

Centrality:

```
centrality(W)
```

Bridge centrality:

```
bridge(graph, communities = clusters,
  useCommunities = c("Elf","Dwarf"))
```

APL, transitivity and small-worldness:

```
smallworldIndex(qgraph(W))
```

N.B.: The .R file including all code and the data file to run this example are available on the *Companion Website*.

Tutorial Box 3.1. Computing network metrics for the *Lord of the Rings* network.

3.6 *Lord of the Rings* example

Figure 3.1 displays an example of a network that shows interactions between characters of the *Lord of the Rings* book series. This example is based on the work of Calvo Tello (2016), who provided the data online.[2] The original network structure was weighted with the edge weight indicating the number of times two characters were mentioned in the same paragraph. First, we transformed this network into an unweighted network by normalizing every edge: i.e., we dived the weight by the smallest number of times one of the two characters was mentioned (the maximum possible number of times two characters could be mentioned in a paragraph together). Subsequently, we connected two nodes together if the normalized number of interactions between these characters exceeded 0.15. The resulting network has 43 nodes and 75 edges, leading to a density of 0.08. By using

[2] https://github.com/morethanbooks/projects/tree/master/LotR

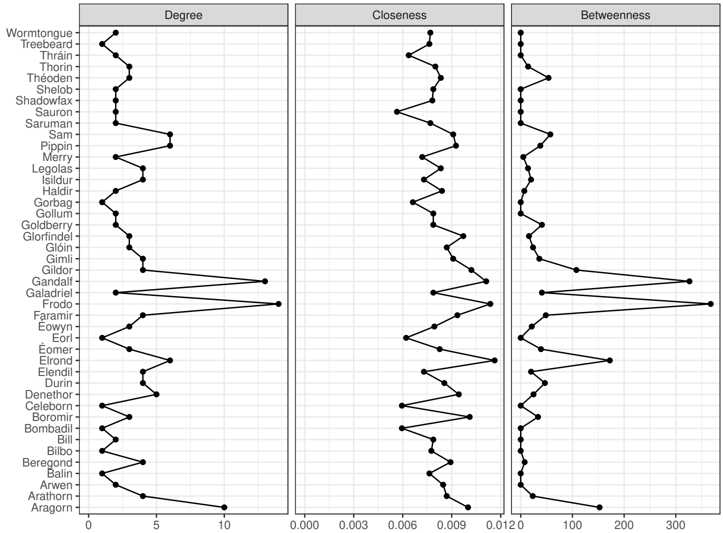

Figure 3.4. Centrality estimates of the *Lord of the Rings* network shown in Figure 3.1. This plot was created using the `centralityPlot` function in the *qgraph* package, which is further discussed in Chapter 4.

R, as further detailed in Tutorial Box 3.1, we can then compute several network metrics. The APL of the network is 2.95 and the transitivity of the network is 0.21. A comparable random network would have an APL of 3.05 and a transitivity of 0.08. The small world index becomes 2.67 (using unrounded APL and transitivity values), indicating that the network contains some small world properties. Figure 3.4 shows estimated centrality metrics (we will learn how to make this plot in the next chapter), indicating that the most central nodes are Frodo, Gandalf, and Aragorn, also the three main characters of the book series. We can in addition note the central role of Elrond, especially in terms of closeness and betweenness. Investigating the bridge centrality between the pre-defined clusters 'Dwarf' and 'Elf,' we find that the characters Legolas and Gimli feature high levels of bridge betweenness, which is in line with the theme of the friendship between these two characters forming a bridge between the rather disjoint societies of elves and dwarfs.

3.7 Conclusion

This chapter introduced networks and some of the most prominent ways of analyzing network structures in network science, including quantifying local network properties such as centrality and clustering of nodes, and global network properties such as the degree distribution and small-worldness of a network. The topics introduced here form the basics of network analysis, but do not nearly cover the full breadth of possibility network science has to offer. To this end, several books have been written that cover descriptive analysis of network structures in far more detail (e.g., Newman, 2010).

In psychological science, many scholars have voiced the need to conceptualize the cognitive and behavioral characteristics of individuals as complex and ever-evolving systems (Ferguson, 1954). The first instantiation of the data-driven methodology from network science followed from the introduction of the mutualism model by van der Maas et al. (2006). This model formalizes the idea that the cognitive system consists of many basic processes that, through their direct relations, form a network of cognitive components that cannot be understood without its interconnectedness. First taken up by the field of psychopathology, this idea has been applied to data on behavioral symptoms from individuals with diagnosed depression (Cramer et al., 2010). This work has caused a surge of empirical research on psychological systems as complex networks. As such, network science is fundamentally reshaping our approach to the ontology, etiology, and development of psychological phenomena (Borsboom, 2017). The remainder of this book will expand on both the methodological and theoretical developments in the field.

3.8 Exercises

Conceptual

3.1. What is the main difference between network structures analyzed in, e.g., social network or railroad network analysis, and the psychological networks analyzed in fields such as psychopathology?

3.2. What additional challenges does this difference present?

3.3. Describe why the absolute value is needed to translate a negative edge weight to an edge length.

3.4. In weighted networks in which the edge weights are real numbers (any continuous quantity), the betweenness metric typically is an integer (whole number). Describe why this is the case.

3.5. Describe the difference between the global clustering coefficient and transitivity.

True or false

3.6. Nodes can also be called *vertices*.

3.7. Node strength is usually calculated by summing up all the edge weights of edges connected to a given node.

3.8. In psychological networks nodes are always well-defined entities.

3.9. The local clustering coefficient can be seen as a measure of redundancy.

3.10. For a network to be a *small-world network* it is sufficient to display very high clustering.

Practical

For exercises in R, please navigate to the appropriate folder of this chapter, available on the online *Companion Website*.

References

Barabási, A.-L. (2012). The network takeover. *Nature Physics*, *8*(1), 14–16.

Blanken, T. F., Deserno, M. K., Dalege, J., Borsboom, D., Blanken, P., Kerkhof, G. A., & Cramer, A. O. J. (2018). The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks. *Scientific Reports*, *8*(1), 5854.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13.

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS One*, *6*(11), e27407.

Calvo Tello, J. (2016). Graph – network of the Lord of the Rings. http://www.morethanbooks.eu/graph-network-of-the-lord-of-the-rings/

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2-3), 137–150.

Doerr, B., Fouz, M., & Friedrich, T. (2012). Why rumors spread so quickly in social networks. *Communications of the ACM*, *55*(6), 70–75.

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS One*, *12*(6), e0179891.

Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Petropolitanae*, *8*, 128–140.

Ferguson, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *8*(2), 95–112.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3-5), 75–174.

Golino, H. F., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using exploratory graph analysis. *Intelligence*, *62*, 54–70.

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One*, *12*(6), e0174035.

Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, *56*(1), 120–149.

Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., et al. (2009). Ecological networks–beyond food webs. *Journal of Animal Ecology*, *78*(1), 253–269.

Jones, P. J., Ma, R., & McNally, R. J. (2021). Bridge centrality: A network approach to understanding comorbidity. *Multivariate Behavioral Research*, *56*(2), 353–367.

Letina, S., Blanken, T. F., Deserno, M. K., & Borsboom, D. (2019). Expanding network analysis tools in psychological networks: Minimal spanning trees, participation coefficients, and motif analysis applied to a network of 26 psychological attributes. *Complexity*, *2019*, 9424605.

Liu, J., Li, M., Pan, Y., Lan, W., Zheng, R., Wu, F.-X., & Wang, J. (2017). Complex brain network analysis and its applications to brain disorders: A survey. *Complexity*, *2017*, 8362741.

Newman, M. (2010). *Networks: An introduction*. Oxford University Press.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251.

Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, *125*(6), 747–757.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109–127.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78.

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Nature: Scientific Reports*, *4*(5918).

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com