State of the aRt personality research: A tutorial on network analysis of personality data in R

Giulio Costantini[1,a], Sacha Epskamp[1,b], Denny Borsboom[b], Marco Perugini[a], René Mõttus[c,d], Lourens J. Waldorp[b], Angélique O. J. Cramer[b]

[a] Department of Psychology, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo 1 (U6), 20126 Milan, Italy

[b] Department of Psychological methods, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands

[c] Department of Psychology, University of Edinburgh, George Square 7, EH8 9JZ Edinburgh, Scotland, UK

[d] Department of Psychology, University of Tartu, Näituse 2, 50409 Tartu, Estonia

Correspondence concerning this article should be addressed to Giulio Costantini, giulio.costantini@unimib.it

[1] The first two authors contributed equally to this work.

Abstract

Network analysis represents a novel theoretical approach to personality. Network approaches motivate alternative ways of analyzing data, and suggest new ways of modeling and simulating personality processes. In the present paper, we provide an overview of network analysis strategies as they apply to personality data. We discuss different ways to construct networks from typical personality data, show how to compute and interpret important measures of centrality and clustering, and illustrate how one can simulate on networks to mimic personality processes. All analyses are illustrated using a data set on the commonly used HEXACO questionnaire using elementary R-code that readers may easily adapt to apply to their own data.

Highlights

- Network analysis can foster novel insights in personality psychology.
- We provide an overview of network analysis.
- We show how R can be used to analyze personality networks.
- We show how to simulate personality networks in R.

Keywords: Network analysis, psychometrics, latent variables, centrality, clustering, personality traits, HEXACO

State of the aRt personality research: a tutorial on network analysis of personality data in R

A network is an abstract model composed of a set of nodes or vertices (*V*), a set of edges, links or ties (*E*) that connect the nodes, together with information concerning the nature of the nodes and edges (e.g., de Nooy, Mrvar, & Batagelj, 2011). Figure 1 reports the example of a simple network, with six nodes and seven edges. The nodes usually represent entities and the edges represent their relations. This simple model can be used to describe many kinds of phenomena, such as social relations, technological and biological structures, and information networks (e.g., Newman, 2010, Chapters 2–5). Recently networks of relations among thoughts, feelings and behaviors have been proposed as models of personality and of psychopathology: in this framework, traits have been conceived of as emerging phenomena that arise from such networks (Borsboom & Cramer, 2013; Cramer et al., 2012; Schmittmann et al., 2013). An R package, *qgraph*, has been developed for the specific purpose of analyzing personality and psychopathology data (Epskamp, Cramer, Waldorp, Schmittman, & Borsboom, 2012).

The aim of this contribution is to provide the reader with the necessary theoretical and methodological tools to analyze personality data using network analysis, by presenting key network concepts, instructions for applying them in R (R Core Team, 2013), and examples based on simulated and on real data. First, we show how a network can be defined from personality data. Second, we present a brief overview of important network concepts. Then, we discuss how network concepts can be applied to personality data using R. In the last part of the paper, we outline how network-based simulations can be performed that are specifically relevant for personality psychology. Both the data and the R code are available for the reader to replicate our analyses and to perform similar analyses on his/her own data.

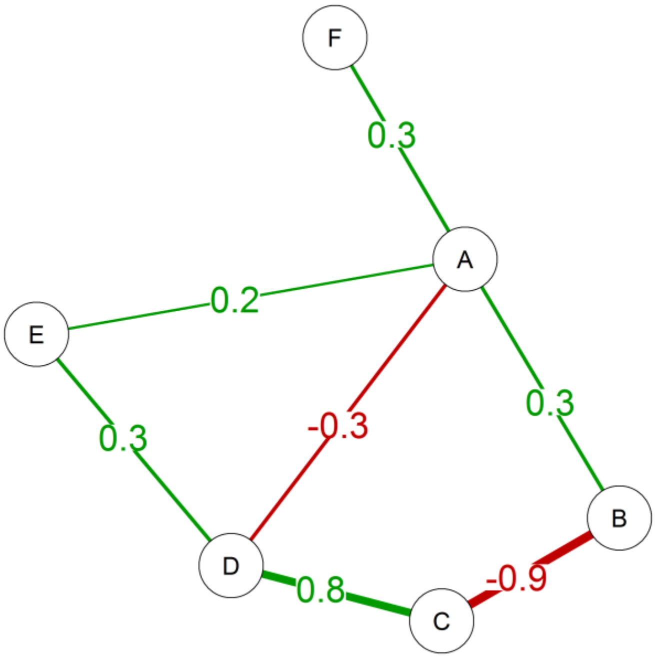

Figure 1. A network with six nodes and seven edges. Positive edges are green and negative edges are red. The letters identify the nodes, the numbers represent weights associated to the edges.

Constructing personality networks

A typical personality data set consists of cross-sectional measures of multiple subjects on a set of items designed to measure several facets of personality. In standard approaches in personality, such data are used in factor analysis to search for an underlying set of latent variables that can explain the structure of the correlations in the data. In a causal interpretation of latent variables (Borsboom, Mellenbergh, and van Heerden, 2003) that appears widespread in the personality literature, responses to items such as “I like to go to parties” and “I have many friends” are viewed as being causally dependent on a latent variable (i.e., extraversion). This approach has culminated in currently influential models such as the Five Factor Model of personality (McCrae & Costa, 2008), in which five dominant latent variables are ultimately held responsible for most of the structural covariation between responses to personality items (additional latent factors such as facets may cause some of the covariation).

Recently, however, this perspective has been challenged in the literature (Cramer et al., 2012). In particular, it has been put forward that the default reliance on latent variable models in personality may be inappropriate, because it may well be that the bulk of the structural covariation in personality scales may result from direct interactions between the variables measured through personality items. For instance, one may suppose that people who like to go to parties gain more friends because they meet more people, and people who have more friends get invited to good parties more often. In this way, one can achieve an explanation of the relevant pattern of covariation without having to posit latent variables.

Thus, in this scheme of thinking, one may suppose that, instead of reflecting the pervasive influence of personality factors, the structural covariance in personality is actually due to local interactions between the variables measured. In this way of thinking, personality resembles an ecosystem in which some characteristics and behaviors stimulate each other, while others have inhibitory relations. Under this assumption, the proper way to analyze personality data is not through the a priori imposition of a latent variable structure, but through the construction of a network that represents the most important relations between variables: this way, one may get a hold of the structure of the ecosystem of personality.

The current section explains how such a network structure can be estimated and visualized in R based on typical personality research data. We discuss the most important kinds of networks, explain how networks are encoded in adjacency matrices, and show how to extract important networks from data.

Directed and undirected networks

First, it is important to note that there are different types of networks, which yield different kinds of information and are useful in different situations. In a *directed* network, relationships between nodes are asymmetrical. Research on directed networks has seen extensive developments in recent years since the work of Pearl (2000) and others on causal systems. Methodology based on directed networks is most useful if one is willing to accept that the network under consideration is *acyclic*, which means that there are no feedback loops in the system (if A influences B, then B cannot influence A). A directed network without feedback loops is called a Directed Acyclic Graph (DAG). In contrast, in an *undirected* network, all relationships are symmetrical. These networks are most useful in situations where (a) one cannot make the strong assumption that the data generating model is a DAG, (b) one suspects that some of the relations between elements in the network are reciprocal, (c) one's research is of an exploratory character and is mainly oriented to visualizing the salient relations between nodes, and (d) one wants to study a weighted network, in which connections can have different strengths. Since the latter situation appears more realistic for personality research, the current paper focuses primarily on undirected networks.

Encoding a network in a weight matrix

The structure of a network depends on the relations between its elements; these are encoded in a *weight matrix*, which is a square matrix in which each row and column indicates a node in the graph and the elements of the matrix indicate the strength of connection of the edge between two nodes. A zero in row i and column j indicates that there is no edge between node *i* and node *j*. For example, the graph of Figure 1 can be represented with the following weight matrix:

	A	B	C	D	E	F
A	0	0.3	0	-0.3	0.2	0.3
B	0.3	0	-0.9	0	0	0
C	0	-0.9	0	0.8	0	0
D	-0.3	0	0.8	0	0.3	0
E	0.2	0	0	0.3	0	0
F	0.3	0	0	0	0	0

In this graph there are positive connections, for instance between nodes A and B, and negative connections, for instance between nodes A and D. The zeros in the matrix indicate that there is no connection between two nodes, such as between nodes A and C. Furthermore, we may note that the matrix is symmetric and that the diagonal values are not used in the graph.

The *qgraph* package (Epskamp et al, 2012) can be used to visualize such a weight matrix as a network:

```
mat <- matrix(c(
       0,  0.3,    0, -0.3,  0.2, 0.3,
     0.3,    0, -0.9,    0,    0,   0,
       0, -0.9,    0,  0.8,    0,   0,
    -0.3,    0,  0.8,    0,  0.3,   0,
     0.2,    0,    0,  0.3,    0,   0,
     0.3,    0,    0,    0,    0,   0), ncol = 6, nrow = 6, byrow = TRUE)
library("qgraph")
qgraph(mat, layout = "spring", edge.labels = TRUE, labels = LETTERS[1:6],
     fade = FALSE)
```

Here, the first argument in the `qgraph` function — the (`mat`) argument — calls the weight matrix to plot. The other arguments specify graphical layout.

Correlation networks, partial correlation networks, and LASSO networks

To illustrate network analysis on personality data we made public a dataset in which nine-hundred-sixty-four participants (704 female and 256 male, M age=21.1, SD=4.9, plus four participants who did not indicate gender and age) were administered the HEXACO-60 (Ashton & Lee, 2009). The HEXACO-60 is a short 60-items inventory that assesses six major dimensions of personality: honesty-humility, emotionality, extraversion, agreeableness vs. anger, conscientiousness and openness to experience (Ashton & Lee, 2007). Each of the major dimensions subsumes four facets, which can be computed as the average of two or three items. Participants indicated their agreement with each statement on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). An example of an item (of trait emotionality) is “When I suffer from a painful experience, I need someone to make me feel comfortable”.

We can load the HEXACO dataset into R as follows:

```
Data <- read.csv("HEXACOfacet.csv")
```

The reader may use `str(Data)` to get a good idea of what the data look like.

Correlation networks. A correlation matrix describes pairwise associations between the facets the HEXACO data and therefore can be used as a good starting point for estimating a graph structure. Therefore, we first compute Pearson correlations on this dataset using the `cor` function:

```
cor(Data)
```

Notice that a correlation matrix is symmetric and that a value of zero indicates no connection. Thus, a correlation matrix, by default, has properties that allow it to be used as a weights matrix to encode a network. Using this connection opens up the possibility to investigate correlation matrices visually as networks. To do so, we can use the *qgraph* package and ask it to plot the correlation matrix as a network; in the remainder, we will indicate this network as a *correlation network*. To facilitate interpretation, we color nodes according to the assignment of facets to traits as specified in the HEXACO manual:

```
groups <- factor(c(
    rep("Honesty Humility", 4),
    rep("Emotionality", 4),
    rep("Extraversion", 4),
    rep("Agreeableness vs. Anger", 4), rep("Conscientiousness", 4),
    rep("Openness to experience", 4)))
```

```
qgraph(cor(Data), layout = "spring", labels = colnames(Data),
      groups = groups)
```

Figure 2A represents the correlation structure of the facets of the HEXACO dataset. Green lines represent positive correlations, while red lines negative correlations. The wider and more saturated an edge is drawn, the stronger the correlation. As the reader may expect, the Figure shows that the correlations of facets within traits are generally higher than the correlations of facets between traits, which is likely to reflect the fact that in psychometric practice items are typically grouped and selected on the basis of convergent and discriminant validity (Campbell & Fiske, 1959).

Partial correlation networks. Correlation networks are highly useful to visualize interesting patterns in the data that might otherwise be very hard to spot. However, they are not necessarily optimal for the application of network analysis if the goal is to extract the structure of a data generating network. The reason is that correlations between nodes in the graph may be spurious, rather than being due to a genuine interaction between two nodes. For instance, spurious correlations may arise as the consequence of shared connections with a third, possibly latent, node (in this sense, factor analysis assumes all the correlations between personality items to be spurious). Often, therefore, instead of the correlation matrix, the inverse of the correlation matrix is used to construct a graph. This matrix is also called a precision matrix. Standardizing the precision matrix results in a matrix that contains partial correlations between variables, in which each entry represents the partial correlation between the respective row and column variables, with the remainder of the variables in the data partialled out. Networks constructed on this basis are called "partial correlation networks" or "concentration graphs" (Cox & Wermuth, 1993), and the statistical data generating structures that they encode are known as Markov random fields (Kindermann & Snell, 1980).

The partial correlation network can be obtained in qgraph by using the argument `graph = "concentration"`:

```
qgraph(cor(Data), layout = "spring", labels = colnames(Data),
        groups = groups, graph = "concentration")
```

The partial correlation network is shown in Figure 2B. In a partial correlation network, two nodes have an edge if and only if the partial correlation is nonzero; if a partial correlation is zero, there is no connection (Lauritzen, 1996). Partial correlation networks can be related to correlation networks in the following way. A correlation between two nodes means that there must be a path between these two nodes, direct or through other nodes. This means that we can estimate the structure of the graph by checking which connections are zero. Default significance tests can be used for this purpose (Drton & Perlman, 2004). However, significance tests require an arbitrary choice of significance level; different choices yield different results, with more stringent significance levels resulting in sparser graphs. If one ignores this issue, one has a multiple testing problem, whereas if one deals with it in standard ways (e.g., through a Bonferroni correction), one faces a loss of power.

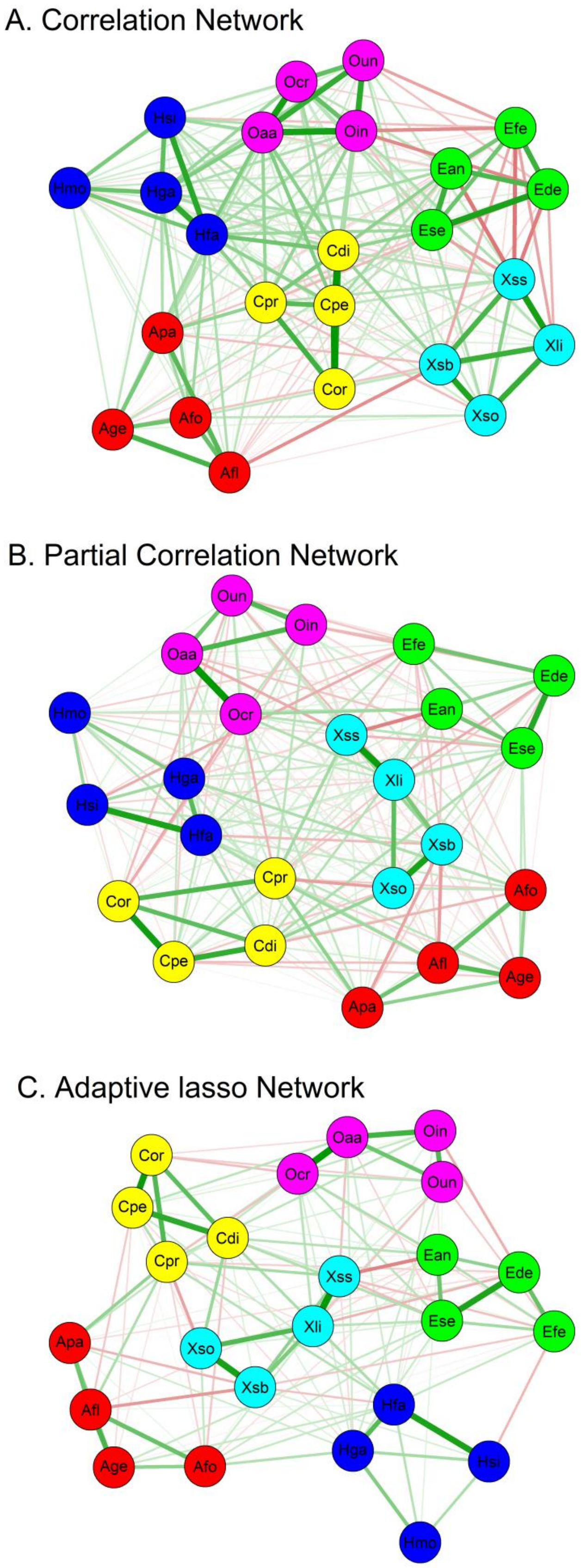

Figure 2. Networks of the HEXACO-60. Nodes represent personality facets (a description of each facet is provided in the Appendix), green lines represent positive connections and red lines represent negative connections. Thicker lines represent stronger connections and thinner lines represent weaker connections.

Adaptive LASSO networks. A practical way to deal with the issue of arbitrary choices is to construct networks based on different choices and to see how stable the main results are; however, a more principled alternative is to use a LASSO penalty (Friedman, Hastie, & Tibshirani, 2008) in estimating the partial correlation graph. This causes small connections to automatically shrink to be exactly zero and results in a parsimonious network. If the data indeed arise from a sparse network with pairwise interactions only, such a procedure will in fact converge on the generating network (Foygel & Drton, 2011). The adaptive LASSO is a generalization of the LASSO that assigns different penalty weights for different coefficients (Zou, 2006) and outperforms the LASSO in the estimation of partial correlation networks, especially if the underlying network is sparse (Fan, Feng, & Wu, 2009; Krämer, Schäfer & Boulesteix, 2009). The penalty weights can be chosen in a data-dependent manner, relying on the lasso regression coefficients (Krämer, et al., 2009). The adaptive LASSO is also convenient practically, as it is implemented in the R-package *parcor* (Krämer et al., 2009). Since the adaptive LASSO, as implemented in package *parcor*, relies on k-fold validation, `set.seed` can be used to ensure the exact replicability of the results, which might be slightly different otherwise. To estimate the graph structure of the HEXACO dataset according to the adaptive LASSO, the following code can be used:

```
library("parcor")
    library("Matrix")
    set.seed(100)
    adls <- adalasso.net(Data)
    network <- as.matrix(forceSymmetric(adls$pcor.adalasso))
    qgraph(network, layout = "spring", labels = colnames(Data),
    groups = groups)
```

The adaptive LASSO network is shown in Figure 2C. One can see that, compared to the partial correlation network, the adaptive LASSO yields a more parsimonious graph (fewer connections) that encodes the most important relations in the data. In our experience, the likelihood of false positives using this method is extremely small, so if an edge is present in the LASSO network one can trust that there is a structural relation between the variables in question (of course, the network does not specify the exact nature of the relation, which may for instance be due to a direct causal effect, a logical relation pertaining to item content, a reciprocal effect, or the common effect of an unmodeled latent variable).

Analyzing the structure of personality networks

Once a network is estimated, several indices can be computed that convey information about network structure[2]. Two types of structure are important. First, one is typically interested in the *global* structure of the network: how large is it? Does it feature strong clusters? Does it reveal a specific type of structure, like a small-world (Watts & Strogatz, 1998)? Second, one may be interested in *local* patterns, i.e., one may want to know how nodes differ in various characteristics: which nodes are most central? Which nodes are specifically strongly connected? What is the shortest path from node A to node B? Here we discuss a limited selection of indices that we regard as relevant to personality research, focusing especially on centrality and clustering coefficients. More extensive reviews of network indices may be found in Boccaletti, Latora, Moreno, Chavez, and Hwang (2006); Butts, (2008a); de Nooy and colleagues (2011); Kolaczyk (2009); and Newman (2010).

Descriptive statistics

Before the computation of centrality measures, a number of preparatory computations on the data are in order. First, the function `upper.tri` can be used to extract the unique edge weights in an adjacency matrix and save them in a vector:

[2] The adaptive LASSO networks, the correlation and the partial correlation networks are characterized by the presence of both positive and negative edges. The importance of signed networks is apparent not only in the study of social phenomena, in which it is important to make a distinction between liking and disliking relationships (e.g., Leskovec, Huttenlocher, & Kleinberg, 2010), but also in the study of personality psychology (e.g., Costantini & Perugini, in press). Some network indices have been generalized to the signed case (e.g., Costantini & Perugini, in press; Kunegis, Lommatzsch, & Bauckhage, 2009), however most indices are designed to unsigned networks. For the computation of the latter kind of indices, we will consider the edge weights in absolute value.

```
ew <- network[upper.tri(network)]
```

To compute the number of edges in the network, it is sufficient to define a logical vector that has value `TRUE` (= 1) if the edge is different from zero and `FALSE` (= 0) if the edge is exactly zero (i.e., absent). The sum of this vector gives the number of nonzero edges. With a similar procedure, it is possible to count the positive and the negative edges: it is sufficient to replace “ ! =” with “>” or “<”.

```
sum(ew != 0) # the number of edges
sum(ew > 0) # the number of positive edges
sum(ew < 0) # the number of negative edges
```

The network has 138 edges, of which 98 are positive and 40 are negative. The function `t.test` can be used to compare the absolute weights of the positive versus the negative edges:

```
t.test(abs (ew [ew > 0]), abs(ew [ew < 0]) , var.equal = TRUE)
```

In our network, positive edges are generally associated to larger weights (M = .11, SD = .09) than the negative edges (M = .07, SD = .03), and the t-test indicates that this difference is significant, $t(136) = 2.98$, $p = .0035$.

Centrality measures

Not all nodes in a network are equally important in determining the network’s structure and, if processes run on the network, in determining its dynamic characteristics (Kolaczyk, 2009). Centrality indices can be conceived of as operationalizations of a node’s importance, which are based on the pattern of the connections in which the node of interest plays a role. In network analysis, centrality indices are used to model or predict several network processes, such as the amount of flow that traverses a node or the tolerance of the network to the removal of selected nodes (Borgatti, 2005; Crucitti, Latora, Marchiori, & Rapisarda, 2004; Jeong, Mason, Barabási, & Oltvai, 2001) and can constitute a guide for network interventions (Valente, 2012). Several indices of centrality have been proposed, based on different models of the processes that characterize the network and on a different conception of what makes a node important (Borgatti & Everett, 2006; Borgatti, 2005). The following gives a succinct overview of the most often used centrality measures[3].

Degree and strength. First, *degree centrality* is arguably the most common centrality index and it is defined as the number of connections incident to the node of interest[4] (Freeman, 1978). The degree centrality of node C in Figure 1 is 2 because it has two connections, with nodes B and D. Degree can be straightforwardly generalized to weighted networks by considering the sum of the weights of the connections (in absolute value), instead of their number. This generalization is called *strength* (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004; Newman, 2004). For instance, strength of node C in Figure 1 is 1.7, which is the highest in the network. Degree and strength focus only on the paths of unitary length (Borgatti, 2005). In networks that represent causal relations, they can be interpreted as the number of nodes that are causally connected to the focal one (degree), the strength taking into account also the magnitude of each causal influence.

Closeness and betweenness. Several other measures exist that, differently from degree centrality and the related indices, consider edges beyond those incident to the focal node. An important class of these indices rely on the concepts of *distance* and of *geodesics* (Brandes, 2001; Dijkstra, 1959). The distance between two nodes is defined as the length of the shortest path between them. Since, in typical applications in personality psychology, weights represent the importance of an edge, weights are first converted to lengths, usually by replacing them with the

[3] The functions to implement centrality indices, clustering coefficients and small-worldness are implemented in the R package *qgraph* (Epskamp et al., 2012). Some of the functions rely on procedures originally implemented in packages *igraph* (Csárdi & Nepusz, 2006), *sna* (Butts, 2008b), and *WGCNA* (Langfelder & Horvath, 2008, 2012). These packages are in our experience among the most useful for network analysis.

[4] Degree centrality can be also easily extended to directed networks, by considering separately the number of incoming ties (*indegree*) and the number of outgoing ties (*outdegree*).

inverse value (Brandes, 2008; Opsahl, Agneessens, & Skvoretz, 2010). The *geodesics* between two nodes are the paths that connect them that have the shortest distance. *Closeness centrality* (Freeman, 1978; Sabidussi, 1966) is defined as the inverse of the sum of the distances of the focal node from all the other nodes in the network[5] . In terms of network flow, closeness can be interpreted as the expected time until arrival of something flowing through the network (Borgatti, 2005). In the network in Figure 1, node D has the highest closeness. To compute the exact value of closeness, one should first compute the distances between D and all the other nodes: A (1/0.3), B (1/0.8+1/0.9), C(1/0.8), E (1/.3) and F (1/.3+1/.3). The sum of all the distances is 16.94 and the inverse, 0.59, is the closeness centrality of D.

Betweenness centrality is defined as the number of the geodesics between any two nodes that pass through the focal one. The index is divided by the number of geodesics between any two nodes, independent of the fact that they traverse the focal node. The denominator is included to account for the possibility of several geodesics between two nodes: if two geodesics exist, each one is counted as a half path and similarly for three or more (Brandes, 2001; Freeman, 1978). Betweenness centrality assumes that shortest paths are particularly important (Borgatti, 2005): if a node high in betweenness centrality is removed, the distances among other nodes will generally increase. Both closeness and betweenness centrality can be applied to weighted and directed networks, as long as the weights and/or the directions of the edges are taken into account when computing the shortest paths (e.g., Opsahl et al., 2010). The betweenness centrality of node A in Figure 1 is 4 and is the highest in the network. The four shortest paths that pass through A are those between F and the nodes B, C, D, and E. Betweenness centrality can also be extended to evaluate the centrality of edges instead of nodes, by considering the geodesics that pass through an edge: this generalization is called *edge betweenness centrality* (Brandes, 2008; Newman & Girvan, 2004; Newman, 2004). For instance, the edge-betweenness centrality of the edge (D,E) is 3 and the three shortest paths that

[5] The computation of closeness assumes that the network is connected (i.e., a path exists between any two nodes), otherwise, being the distance of disconnected nodes infinite, the index will result to zero for all the nodes. Variations of closeness centrality that address this issue have been proposed (e.g., Kolaczyk, 2009, p. 89; Opsahl et al., 2010, n. 1). Alternatively it can be computed only for the largest component of the network (Opsahl et al., 2010).

pass through (D,E) are the one between D and E, the one between C and E (through D), and the between B and E (through C and D).

Several other variants of the shortest-paths betweenness are discussed in Brandes (2008), some of which are implemented in package *sna* (Butts, 2008b). Generalizations of betweenness centrality that account for paths other than the shortest ones have been also proposed (Brandes & Fleischer, 2005; Freeman, Borgatti, & White, 1991; Newman, 2005). In addition, Opsahl and colleagues (2010) proposed generalizations of degree, closeness, and betweenness centralities by combining in the formula both the number and the weights of the edges. They introduced a tuning parameter that allows setting their relative importance: a higher value of the tuning parameter emphasizes the importance of the weights over the mere presence of the ties and vice versa. Another important family of centrality indices defines the centrality of a node as recursively dependent on the centralities of their neighbors. Among the most prominent of those indices are *eigenvector centrality* (Bonacich, 1972, 2007), *Bonacich power* (Bonacich, 1987) and *alpha centrality* (Bonacich & Lloyd, 2001).

Clustering coefficients

Besides centrality, other network properties have been investigated that are relevant also for personality networks. The local *clustering coefficient* is a node property defined as the number of connections among the neighbors of a focal node over the maximum possible number of such connections (Watts & Strogatz, 1998). If we define a triangle as a triple of nodes all connected to each other, the clustering coefficient can be equally defined as the number of triangles in the neighborhood of a focal node, normalized by the maximum possible number of such triangles. The clustering coefficient is high for a node *i* if most of *i*'s neighbors are also connected to each other and it is important to assess the small-world property (Humphries & Gurney, 2008; Watts & Strogatz, 1998). The clustering coefficient can be also interpreted as a measure of how much a node is redundant (Latora, Nicosia, & Panzarasa, 2013; Newman, 2010): if most of a node's neighbors are also connected with each other, removing that node will not make it harder for its neighbors to reach or influence each other. Consider for instance the node D in Figure 1, which has three neighbors, A C, and E. Of the three possible connections among its neighbors, only one is present (the one between A and E), therefore its clustering coefficient is 1/3.

While in its original formulation the clustering coefficient can be applied only to unweighted networks (or to weighted networks, disregarding the information about weights), it has been recently generalized to consider positive edge weights (Saramäki, Kivelä, Onnela, Kaski, & Kertész, 2007). The first of such generalizations was proposed by Barrat and colleagues (2004) and has been already discussed in the context of personality psychology and psychopathology (Borsboom & Cramer, 2013). Onnela and colleagues (2005) proposed a generalization that is based on the geometric averages of edge weights of each triangle centered on the focal node. A different generalization has been proposed in the context of gene co-expression network analysis by Zhang and Horvath, which is particularly suited for networks based on correlations (Kalna & Higham, 2007; Zhang & Horvath, 2005). All of these generalizations coincide with the unweighted clustering coefficient when edge weights become binary (Saramäki et al., 2007). Recently three formulations of clustering, the unweighted clustering coefficient (Watts & Strogatz, 1998), the index proposed by Onnela (2005) and the one proposed by Zhang and Horvath (2005) have been generalized to signed networks and the properties of such indices have been discussed in the context of personality networks (Costantini & Perugini, in press).

Transitivity (or global clustering coefficient) is a concept closely connected to clustering coefficient that considers the tendency for two nodes that share a neighbor to be connected themselves for the entire network, instead than for the neighborhood of each node separately. It is defined as three times the number of triangles, over the number of connected triples in the network, where a connected triple is a node with two edges that connect it to an unordered pair of other nodes (Newman, 2003). Differently from the local clustering coefficient, transitivity is a property of the network and not of the single nodes. For instance, the network in Figure 1 has one triangle (A, D, E) and 12 connected triples, therefore its transitivity is (3*1)/12 = 1/4. Transitivity has been extended by Opsahl and Panzarasa (2009) to take into account edge weights and directions, and by Kunegis and collaborators to signed networks (Kunegis et al., 2009).

Small worlds

The transitivity and clustering coefficient can be used to assess the network *small-world property*. The small-world property was initially observed in social networks as the tendency for any two people to be connected by a very short chain of acquaintances (Milgram, 1967). The small-world property is formally defined as the tendency of a network to have both high clustering coefficient and short average path lengths (Watts & Strogatz, 1998). Small-world networks are therefore characterized by both the presence of dense local connections among the nodes and of links that connect portions of the network otherwise far away from each other. An index of *small-worldness* for unweighted and undirected networks has been proposed as the ratio of transitivity to the average distance between two nodes. Both transitivity and path length are standardized before the computation of small-worldness, by comparing them to the corresponding values obtained in equivalent random networks (with the same N and the same degree distribution). Alternatively, the index can be computed using the average of local clustering coefficients instead of transitivity. A network with a small-worldness value higher than three can be considered as having the small-world property, while a small-worldness between one and three is considered a borderline value (Humphries & Gurney, 2008). Because the assessment of small-worldness relies on shortest paths between all the pairs of nodes, it can be computed only for a connected network or the giant component of a disconnected network.

Application to the HEXACO data

Centrality analyses. The function `centrality_auto` allows to quickly compute several centrality indices. It requires the adjacency matrix or the weights matrix as input. The function automatically detects the type of network and can handle both unweighted and weighted networks, and both directed and undirected networks. For a weighted and undirected network, the function gives as output the node strength, the weighted betweenness and the weighted closeness centralities. The edge betweenness centrality is also computed.

```
centrality <- centrality_auto(network)
nc <- centrality$node.centrality
ebc <- centrality$edge.betweenness.centrality
```

The centrality values are computed and stored in variable centrality. Node centralities are then saved in the variable `nc` while edge betweenness centralities are saved in the variable `ebc`. The values of centrality for each node are reported in the Appendix. Table 1 reports the correlations among the three indices of node centrality.

Table 1.

Correlation of node centralities.

	1	2	3
1. Betweenness	1	.66**	.71**
2. Closeness	.63**	1	.76**
3. Strength	.71**	.82**	1

Note. **p<.01, ***p<.001. Pearson correlations are reported below the diagonal, Spearman correlations are reported above the diagonal.

The three indices of centrality converge in indicating that node Cpr (prudence) is among the four most central nodes in this network. Cpr is also the more closeness central node and owes its high centrality to the very short paths that connect it to other traits. For instance, facets Apa (patience), Xso (sociability), and Xss (social self-esteem) are even closer to Cpr than other conscientiousness facets. This suggests that in the personality network it is very easy that a change in some portion of the network will eventually make a person either more reckless or more prudent.

On the other hand, if a person becomes more reckless or more prudent, we can expect important changes in the overall network.

Hfa (fairness) is the most betweenness-central and strength-central node, but it is not particularly closeness-central (it is ranked 10th in closeness centrality). Figure 3 highlights the edges lying on the shortest paths that travel through node Hfa, in a convenient layout (the code for producing this figure is in the supplemental materials). The high betweenness centrality of Hfa is due the role that Hfa plays in transmitting the influence of other honesty-humility facets to different traits, and vice versa. One could speculate that if a person is greedy (Hga) immodest (Hmo) and insincere (Hsi), but nonetheless fair (Hfa), the influence of his/her lack of honesty-humility will not propagate so easily to the rest of his/her personality. The edge between nodes Hsi (sincerity) and Hfa is also the most betweenness-central in the whole network.

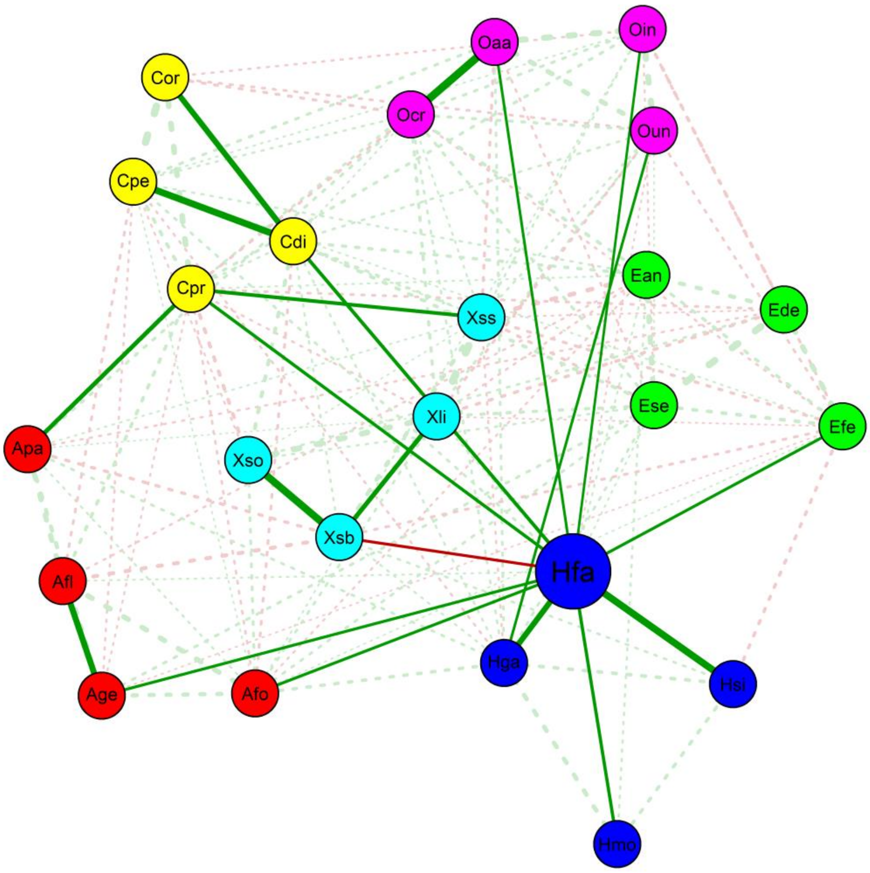

Figure 3. Shortest paths that pass through node Hfa (fairness). The edges belonging to the shortest-paths are full, while the other edges are dashed.

Clustering coefficients. Many indices of clustering coefficient can be easily computed using function `clustcoef_auto`. The function requires the same input as `centrality_auto` and is similarly programmed to recognize the kind of data given as input and to choose an appropriate network representation for the data. By applying the function, we can immediately collect the results:

```
clustcoef <- clustcoef_auto(network)
```

Table 2 reports the correlation among several clustering coefficients. The unsigned indices are computed using the absolute values of the weights. Considering the very strong correlation among the signed indices, we will use the signed version of the Onnela's (2005) clustering coefficient for the following analyses[6].

Table 2.

Correlation among indices of local clustering coefficient.

	1	2	3	4	5	6	7
1. Watts and Strogatz (1998)	1	0.11	0.69**	0.51*	0.89**	0.53**	.92**
2. Watts and Strogatz, signed (Costantini & Perugini, in press)	0.16	1	0.27	0.51*	0.33	0.75**	.14
3. Zhang and Horvath (2005)	.53**	0.71**	1	0.86**	0.70**	0.60**	.77**
4. Zhang and Horvath, signed (Costantini & Perugini, in press)	.43*	0.8**	0.94**	1	0.57**	0.82**	.54**
5 .Onnela et al. (2005)	.81**	0.41*	0.76**	0.66**	1	0.68**	.89**
6 Onnela et al., signed (Costantini & Perugini, in press)	.41*	0.92**	0.87**	0.93**	0.70**	1	.52**
7. Barrat et al. (2004)	.95**	.23	.65**	.50*	.89**	.49*	1

Note. *p<.05, **p<.01. Pearson correlations are reported below the diagonal, Spearman correlations are reported above the diagonal.

Combining clustering coefficients and centrality. The signed clustering coefficient can be interpreted as an index of a node's redundancy in a node's neighborhood (Costantini & Perugini, in press): the importance of the unique causal role of highly clustered nodes is strongly reduced by the presence of strong connections among their neighbors. In general, it is interesting to inspect whether there is a relation between centrality indices and clustering coefficients: in our experience, we found that the centrality indices were often inflated by the high clustering in correlation networks.

[6] Using the signed version of Zhang and Horvath's (2005) index would yield very similar results.

However this might be not true for networks defined with adaptive LASSO, which promotes sparsity (Krämer et al., 2009).

The following plots can be used to visualize both the centrality and the clustering coefficient of each node. The code reported here is for betweenness centrality, but it is easy to extend it to other indices by just replacing "`Betweenness`" with the index of interest. First the plot is created and then the node labels are added in the right positions, using the command text. Command `abline` can be used to trace lines in the plot. A horizontal line is created to visually identify the median value of betweenness and a vertical line to identify the median value of the clustering coefficient.

```
plot(clustcoef$signed_clustOnnela, nc$Betweenness, col = "white")
text(clustcoef$signed_clustOnnela, nc$Betweenness, rownames(nc))
abline(h = median(nc$Betweenness), col = "grey")
abline(v = median(clustcoef$signed_clustOnnela), col = "grey")
```

The resulting plots are showed in Figure 4. It is apparent that the most central nodes do not have a particularly high clustering coefficient in this case and this is especially true for nodes Hfa and Cpr, which are among the most central in this network. The clustering coefficient correlates negatively with closeness centrality (r = -.60, p = .002), with strength (r = -.75, p < .001), and with betweenness centrality (r = -.37, p = .08).

One node, Hmo (modesty), emerges as both particularly high in clustering coefficient and low in all the centrality measures. A closer exam of its connections reveals that Hmo has only four neighbors, the three other facets of honesty-humility (His, Hfa, and Hga) and facet anxiety of emotionality (Ean). All of its neighbors are connected to each other, with just two exceptions: a connection is missing among Ean and two nodes of honesty-humility, Hsi and Hga. Even if the edges incident in node Hmo were blocked, its neighbors would be nonetheless connected to each other directly or by a short path. Modesty therefore does not seem to play a very important unique role in the overall personality network.

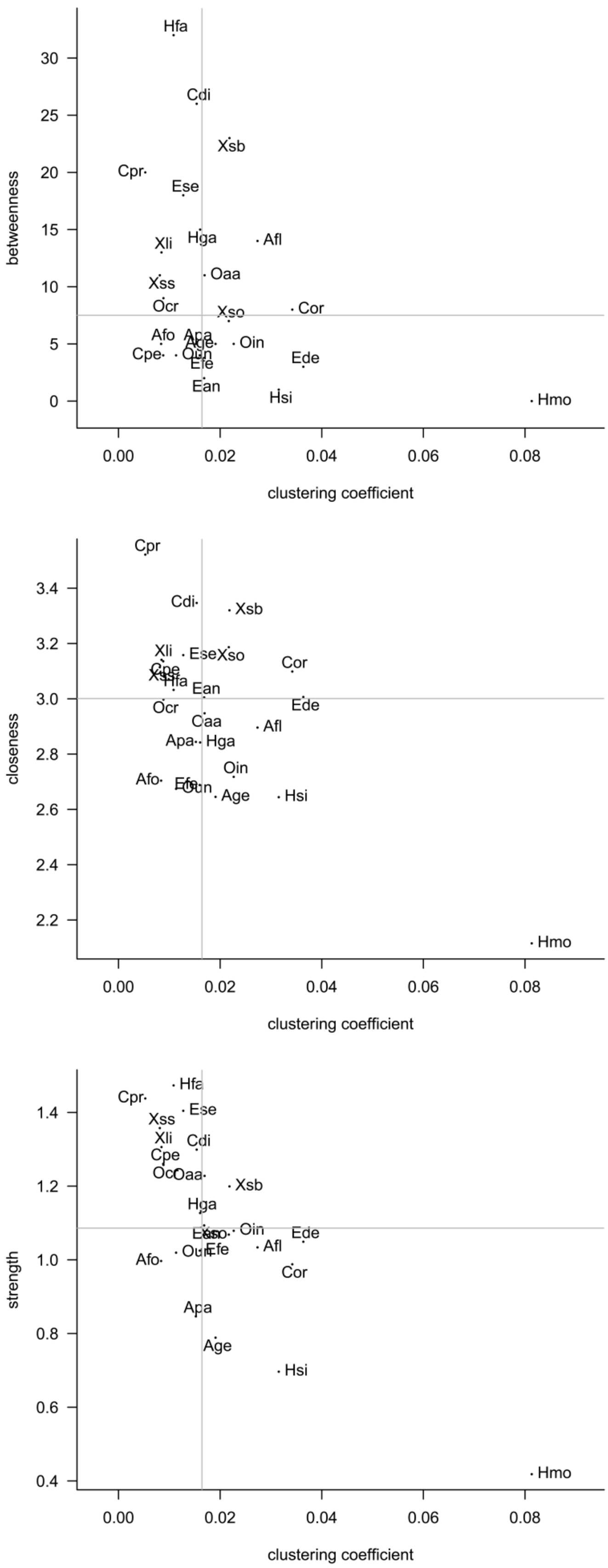

Figure 4. Centrality and clustering coefficient. The horizontal and the vertical lines represent the median values of centrality and clustering coefficient respectively. The closeness values are multiplied by 1000.

Transitivity and small-world-ness. The function `smallworldness` computes the small-worldness index (Humphries & Gurney, 2008). First the function converts the network to an unweighted one, which considers only the presence or the absence of an edge. Then the average path length and the global transitivity of the network are computed and the same indices are calculated on B=1000 random networks, with the same degree distribution of the focal network. The resulting values are entered in the computation of the small-worldness index. The output includes the small-worldness index, the transitivity of the network, and its average path length. It also returns summaries of the same indices computed on the random networks: the mean value and the .005 and .995 quantiles of the distribution. Function `set.seed` can be used to ensure the exact replicability of the results. The function requires the network as input and it is optionally possible to set the values of three parameters, `B`, `up` and `lo`, which are respectively the number of random networks and the upper and lower probabilities for the computation of the quantiles

```
set.seed(100)
smallworldness(network)
```

The small-worldness value for our network is 1.02. An inspection of the values of transitivity and of average path length shows that they are not significantly different from those emerged from similar random networks. Therefore we may conclude that this personality network does not show a clear small-world topology.

Emerging insights. In this section, we showed how it is possible to perform a network analysis on a real personality dataset. We identified the most central nodes and edges, discussed centrality in the light of clustering coefficient and investigated some basic topological properties of the network, such as the small-world property. Two nodes resulted particularly central in the network and were the facet prudence of conscientiousness (Cpr) and the facet fairness of honesty-humility (Hfa).

Our network did not show the small-world property. We consider this result as particularly important. The absence of a strong transitivity means that the connection of two nodes with a common neighbor does not increase the probability of a connection between themselves. The absence of a particularly short path length implies that it is not generally possible for any node to influence any other node using a short path. This can be interpreted as a protective mechanism of personality: a dysfunction in one component does not hamper or modify the overall functionality of the person. This result is not in line with the small-worldness property that emerged in the DSM-IV network reported by Borsboom, Cramer, Schmittmann, Epskamp, and Waldorp (2011). This difference might be attributable to the strategies that were used for defining this network and the DSM-IV network. Moreover, it has been hypothesized that the small-world property might be at the basis of phenomena connected to the comorbidity that arise in psychopathology (Cramer et al., 2010); this also may simply not be a property of normal personality. Future research may be directed towards the question of what network structure characterizes normal versus abnormal personality.

Simulating personality networks

In addition to the analysis of empirical data, network modeling offers extensive possibilities in the area of theory development. This is because, in contrast to purely data analytic models like factor analysis, networks are naturally coupled to dynamics (e.g., see Kolaczyk, 2009): they can evolve, grow, and change over time, with direct consequences for their dynamic behavior. This makes it possible to start thinking about questions like: How do personality networks form in development? Do they grow and, if so, how do they change in structure over time? Do different people have different network structures, and how would such differences relate to growth and dynamics?

Because networks have been so extensively studied in other fields, one can use existing analytical insights on the relevant processes (e.g., Newman, 2008; Kolaczyk, 2009; Grimmett, 2010). When applicable, existing analytical approaches can be very powerful. However, in order to use such analytical approaches, one often has to consider assumptions that are unlikely to be met in personality (e.g., many theorems require one to assume that nodes are exchangeable save for their position in the network, or work only for unweighted networks). In such cases, a specifically tailored simulation methodology can be an extremely versatile tool to study the behavior of networks. This can both enlighten one's data analytic results (e.g., by checking how a given dynamical process would pan on a network extracted from data; e.g., see Borsboom et al., 2011) and help in theory development (e.g., by working out what a hypothesized network would imply theoretically).

In particular, simulation work can be used to design some hypothetical data and see how these data “behave” in appropriate analyses. Here, designing data refers to simulating data according to some pre-specified rules. The obvious strength of testing analytical procedures or concepts with simulated data is that the mechanisms by which the data arose are known—a luxury researchers almost never have when working with real data. Therefore, it is possible to see if the focal theoretical concept can, in principle, result in the expected kind of observed data or co-exist with other concepts, or whether the analytical procedure of interest can yield accurate conclusions. Obviously, designed data can provide no empirical proof for a theoretical concept—but they can guide thinking and this is almost as good.

One can attempt to simulate personality network data to exactly the same two ends. For example, some relevant questions can be the following. Is it possible to generate data that look similar to what personality psychologists commonly work with, starting from network principles? And if so, how do available network analyses tools behave when applied to these data? This section describes only one possible way of simulating personality data from the network perspective. In particular, we demonstrate how the coalescence of observable variables into traits can be simulated.

One possible way to start. We can start off with creating observable variables that will constitute the nodes of the network. For the purpose at hand, we assume that the nodes can be uncorrelated at the outset and that their clustering results from direct causal connections among them. Therefore, the initial value of each node is drawn separately from standard normal distribution.

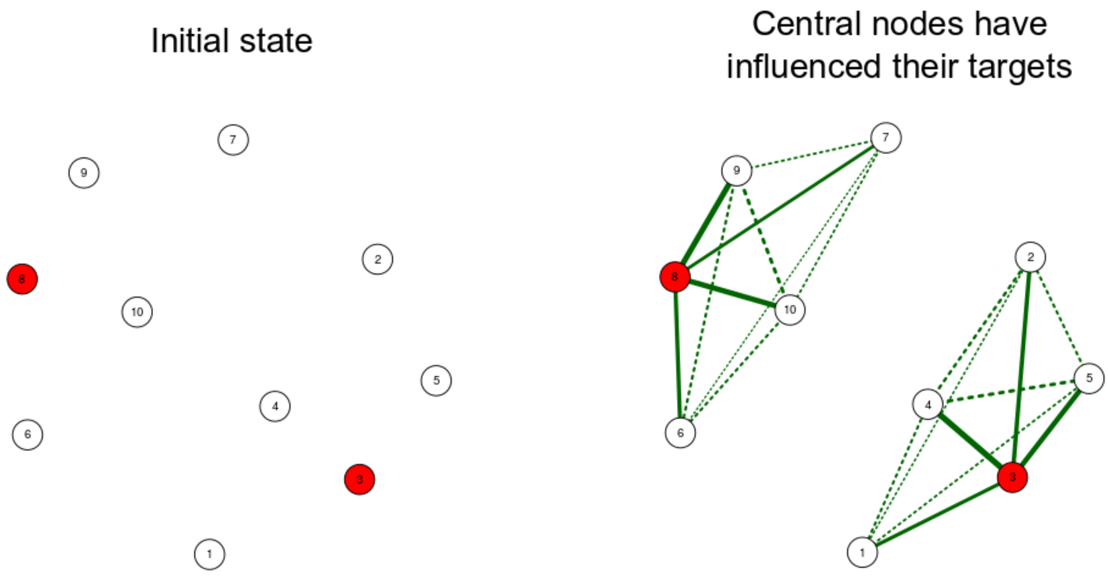

Figure 5. A network of 10 nodes. At the initial stage (left panel), no influences have been spread around and therefore nodes are uncorrelated. At a later stage (right panel), two central nodes (red) have sent direct influences (solid lines) to nodes close to them. Dashed lines represent indirect connections.

The next step is to allow the nodes to influence each other. We do not expect that every node should be able to influence every other node: this could eventually result in a tightly intercorrelated lump of variables, which is hardly what characterizes real personality data. Instead, we assume that some nodes are more likely to influence each other and this is what leads to the structuredness of personality network (i.e., a multi-trait structure). Here, we rely on the concept of centrality and assume that some (central) nodes are likely to influence a number of other nodes (targets). Specifically, the central nodes may have the strongest influence on the nodes around them or their direct influence may be inversely proportional to the distance between them and their target nodes, or the two possibilities can be combined.

Essentially, this may be all that is necessary for a structured personality-like network to appear: we assign some nodes the central role and let them send little influences to other nodes either around them or all over the network (see Figure 5), including other central nodes if these happen to be around or otherwise among the targets. This is exactly what the R function `simulator` (see supplemental materials for details) does. It takes in the following four arguments. First `data`, a matrix or a `data.frame` containing uncorrelated variables, where the number of columns is the desired number of nodes in the network and the number of rows is the desired 'sample' size. Second `centrals`, a vector indicating the positions of the central nodes. Third `n.targets`, a number specifying the number of target nodes that can be directly influenced by the respective central node, with the influence waning with distance. If this number equals or exceeds the total number of nodes in the network, then the central node can influence all variables in the network but, again, the strength of the influence decreases as the distance between the central and target nodes increases. Fourth, `circular`, a logical specifying if the target nodes are systematically selected from among nodes at shortest distance from them (if `TRUE`, default) or randomly from among all nodes. If `TRUE`, the network will look circular. Fifth, `weights`, a vector specifying the mean and standard deviation of the normal distribution from which the strengths of the influences are sampled. The function returns the input dataset of uncorrelated variables on which the specified network influences have been applied.

It could be expected that, by being influenced by a common node, the levels of the respective target nodes become correlated and thereby a trait-like cluster appears; in factor analysis (FA) or principal component analysis (PCA), the central node would appear as having the highest factor loading. Note that in this case the central node essentially serves the role of the latent variable in factor analysis only that it is not really latent—it is one of the observed variables. If this idea works in the simulation—and it is really so trivial that it must work—then it suggest an interesting theoretical possibility: perhaps one of the indicators (e.g., item or facet) of a personality factor is the cause of other indicators rather than there being an underlying direct cause for all of them (that is, there may be an underlying cause but its effect on nodes other than the central one is indirect, mediated by the central node). Of course, if the central node does not happen to be observed because, for example, the relevant items were not included in the questionnaire, a trait may still appear and then there is indeed an unobserved direct common cause for all of the measured variables.

A simplest possible simulation. To illustrate the principles by which networks can produce the appearance of statistical factors in the data, we run `simulator` by feeding in 10 uncorrelated variables of length 10000, specifying the 5th node as central and allowing it to influence all other nodes (the influence generally wanes with increasing distance from the central variable, as said above) with strengths that are drawn from a normal distribution with a mean of .3 and standard deviation of .1. These weights are completely arbitrary. The relevant code is:

```
items <- simulator(
     data = replicate(10, rnorm(10000)),
     centrals = 5,
     n.targets = 10,
     weights = c(.3, .1))
```

Subjecting the resulting data (i.e., items) to PCA (function `principal` from the package *psych*; Revelle, 2013) results in a one-component solution that accounts for about 15% to 20% of variance in the ten variables that had initially been uncorrelated. The fifth variable has the highest correlation with the component and the further away from it the smaller the loadings become. Centrality analysis based on *qgraph* shows that the fifth node has the highest betweenness and closeness centralities. Fitting an unidimensional reflective confirmatory analyses (CFA) model (one latent trait defined by the ten variables without residual correlations allowed) on the data tends to yield good model fit, but the residual covariance matrix may not always be positive definite because the latent trait is almost perfectly defined by the single central node. CFA models can be fitted with the `cfa` function from *lavaan* package (Rosseel, 2012).

Obviously, there may be more than one central node responsible for a trait-like cluster. If they can influence (i.e., are close to) each other, they become correlated, so become their target nodes and thereby a single trait-like cluster forms. For example, there may be, say, two interrelated central nodes among those that coalesce into neuroticism: anxiety and low mood (nodes 5 and 6 in the below code).

```
items <- simulator(
     data = replicate(10, rnorm(10000)),
     centrals = c(5,6),
     n.targets = 10,
     weights = c(.3, .1))
```

Subjecting the resulting data (items) to PCA results in a one-component solution that accounts for about 20% to 25% of variance in the 10 variables. Variables 5 and 6 have the highest factor loadings and they also outperform others in centrality. Fitting a unidimensional CFA model on these data tends to yield somewhat poorer fit than can be observed for data with a single central node, but chances of ending up with a not positive definite residual covariance matrix are small. The slightly poorer model fit may be caused by the stronger associations around the central nodes than elsewhere, which lead to the violation of the local independence assumption of the latent trait model. Interestingly, the local independence assumption also tends to be violated in real personality data (e.g., Gignac, Bates, & Jang, 2007) and such simulations may hint at one possible reason for this relatively ubiquitous phenomenon. Causes of poor model fit can be easily detected using the `modinidces` function of package *lavaan*.

This simulation was very minimalist in all possible ways because its idea was to provide the first feel for what could be done to simulate personality networks. For example, there was just one trait-like cluster, whereas we know that personality consists of many. Likewise, there were no individual differences in the network architecture but, eventually, we may want to model these too.

A slightly more complex simulation. Using the same `simulator` function, we can also simulate the coalescence of nodes into multiple trait-like clusters. For this purpose, we can place central nodes or groups of them further away from each other, meaning that some central nodes become more or less independent of each other and therefore form at least somewhat independent clusters. The exact degree of independence of the clusters depends on how much their targets overlap.

Likewise, we can introduce individual differences in network architecture: what nodes are assigned the central role and which are their exact targets can vary across simulated individuals (agents). Also, connections strengths can systematically vary. However, although the exact nodes that become central may vary, there has to be at least some level of consistency across agents in this respect. If the network of every agent was completely different, then there would be no between-agent consistency and therefore no trait-like clusters would be likely to appear *across* agents. This points to an interesting and possible substantive question: how much consistency in the centrality of nodes is needed for realistic-looking data to appear. One way to test this is to allow central nodes to be chosen separately for each agent in a quasi-random manner by specifying the probabilities for some nodes to be selected somewhat higher than the respective probabilities for other nodes. The interesting question then may be: how much imbalance in these centrality-probabilities is necessary for traits to appear?

In this example, we simulate a network consisting of 21 nodes that are expected to coalesce into three traits despite every agent having somewhat different network architecture. We introduce some consistency in the locations of central nodes via a vector of probabilities for each of the 21 nodes to be selected as central. In particular, nodes 3, 4, 5, 11, 17, 18, 19 have somewhat higher probability (`p`) of being selected among the central nodes (`p` = .06, .09, or .12) than other nodes (`p` = .03). The exact number of nodes selected as central varies from 0 to 14 with a median of about 7. The number of targets as well as average connection strengths also varies across agents.

First, we assign the probabilities:

```
probs <- c(.03,.03,.06,.09,.06,.03,.03,.03,.03,.03,
     .12,.03,.03,.03,.03,.03,.06,.09,.06,.03,.03)
```

Next, we replicate `simulator` 10000 times with each simulation having unique start values and producing scores for only one agent. This way we obtain data for 10000 uniquely simulated agents.

```
people <- replicate(10000,
     simulator(
          data = t(replicate(21, rnorm(1))),
          centrals = sample(21, sample(0:14,1), prob = probs),
          n.targets = sample(21,1),
          weights = c(rnorm(1,.3,.1), rnorm(1,.3,.05))))
```

In this simulation there is a fair level of idiosyncrasy across agents in how their networks of 21 nodes operated. It may now be interesting to ask whether the imbalanced probabilities for central node selection were sufficient for a consistent three-trait structure to appear in between-agent level analysis. Submitting these data (i.e., `people`) to PCA, about five components tend to be suggested by parallel analyses but a screeplot tends to point three dominant components and two smaller components slightly beating the chance (if the network is not specified as circular, then three components tend to be suggested). The three-component solution, typically accounting for more than 30% of variance in observables, tends to yield a relatively well-interpretable loading matrix. A corresponding three-factor CFA model can also be fit on these data but the fit indices are poor and modification indices tend to suggest several ways of improving the model. This, of course, is very similar to what happens with empirical personality data.

Another interesting question may be the correspondence between the probabilities for particular nodes to be central (`probs`) and the centrality estimates derived post hoc. Indeed, betweenness and closeness centralities of particular nodes derived from the simulated data tend to have moderate to high correlations with probabilities of these nodes being selected as central at the outset. That is, nodes that are designed to have a higher chance of being central appear as central indeed.

Obviously, other methods of analyzing networks that have been described above can also be applied on the simulated data. To get a feel for these analytical tools, the reader is encouraged to fiddle with `simulator` arguments and see how the changes reflect in the results of the network analyses of interest.

Extensions to more complicated cases. This section demonstrated only one way of simulating personality network data; there are likely to be other approaches that start from very different conceptual mechanisms and may or may not end up with similar results. Likewise, the demonstrated simulations were conceptually very simple and only addressed the coalescence of nodes into trait-like clusters. To the extent that the network perspective correctly reflects human personality, however, such networks are likely to function as dynamic systems that grow, obtain relative stability and interact with environment. Such networks can also be simulated using R (Mõttus, Penke, Murray, Booth, & Allerhand, 2013) but this is beyond the scope of this section.

Discussion

Network approaches offer a rich trove of novel insights into the organization, emergence, and dynamics of personality. By integrating theoretical considerations (Cramer et al., 2010), simulation models (Mõttus et al., 2013; Van der Maas et al., 2006), and flexible yet user-friendly data-analytic techniques (Epskamp et al., 2012), network approaches have potential to achieve a tighter fit between theory and data analysis than has previously been achieved in personality research. At the present time, the basic machinery for generating, analyzing, and simulating networks is in place. Importantly, the R platform offers an impressive array of packages and techniques for the researcher to combine, and most of the important analyses are currently implemented. We hope that, in the present paper, we have successfully communicated the most important concepts and strategies that characterize the approach, and have done so in such a way that personality researchers may benefit from using network modeling in the analysis of their own theories and datasets.

In the present paper, we have applied network modeling to an illustrative dataset, with several intriguing results that may warrant further investigation. However, we do stress that many of our results are preliminary in nature. The primary reason for this is that current personality questionnaires are built according to psychometric methodology that is tightly coupled to factor analysis and classical test theory (Borsboom, 2005). This makes their behavior predictable from pure design specifications, which in turn limits their evidential value. That is, if one makes the a priori decision to have, say, 10 items per subscale, and selects items on the basis of their conformity to such a structure, many of the correlations found in subsequent research are simply built into the questionnaire. Therefore, it is hardly possible to tell to what extent results reflect a genuine structure, or are an artifact of the way personality tests are constructed. Trait perspectives are not immune to this problem, as in some cases the factors of personality may simply appear from questionnaire data because they have been carefully placed there.

An interesting question is whether all individuals are scalable on all items, as current methodology presumes. It is entirely possible, if not overwhelmingly likely, that certain items assess variables that simply do not apply to a given individual. Current psychometric methods have never come to grip with the "n.a." answer category, and in practice researchers simply force all individuals to answer all items. In networks, it is easier to deal with the n.a.-phenomenon, as nodes deemed to be inapplicable to a given person could simply be omitted from that person's network. This would yield personality networks that may differ in both structure and in size across individuals, an idea that resonates well with the notion that different people's personalities might in fact be also understood in terms of distinct theoretical structures (Borsboom et al., 2003; Cervone, 2005; Lykken, 1991). The application of experience sampling methodology and of other ways to gather information on dynamical processes personality may also offer an inroad into this issue (Fleeson, 2001; Hamaker, Dolan & Molenaar, 2005; Bringmann et al., 2013).

The notion that network structures may differ over individuals, and that these differences may in fact be the key for understanding both idiosyncrasies and communalities in behavior, was illustrated in the simulation work reported in the present paper. Future research might be profitably oriented to questions such as (a) what kind of structural differences in networks could be expected based on substantive theory, (b) how such differences relate to well-established findings in personality research (see also Mõttus et al., 2013), (c) which network growth processes are theoretically supported by developmental perspectives. Of course, ultimately, such theoretical models would have to be related back to empirical data of the kind discussed in the data-analysis part of this paper; therefore, a final highly important question is to derive testable implications from such perspectives. This includes the investigation of how we can experimentally or quasi-experimentally distinguish between explanations based on latent variables, and explanations based on network theory.

Ideally, these future developments are coupled with parallel developments in statistical and technical respects. Several important extensions of network models are called for. First, as noted in this paper, many network analytics were originally designed for unweighted networks. Although some of the relevant analyses have now been extended to the weighted case (see Boccaletti et al., 2006; Opsahl et al., 2010; Constantini et al., in press), several other techniques still wait such generalization. One important such set of techniques, which were also illustrated in the present work, deals with the determination of network structure. Both the theoretical definition of global structures (e.g., in terms of small-worlds, scale-free networks, and random networks) and the practical determination of these structures (e.g., through coefficients such as small-worldness or through fitting functions on the degree distribution) are based on unweighted networks. It would be highly useful if these notions, and the accompanying techniques, would be extended to the weighted network case.

A second technical improvement that should be within reach is how to deal with data that likely reflect mixtures of distinct networks (as in the second simulation in the current paper). In the case of time series data, such approaches have already been formulated through the application of mixture modeling (Bringmann et al., 2013); however, statistical techniques suited to this problem may also be developed for the case of cross-sectional data. The issue is important in terms of modeling idiosyncrasies in behavior, but may also be key in terms of relating normal personality to psychopathology (Cramer et al., 2010). Naturally, this includes the question of how we should think about the relation between normal personality and personality disorders.

References

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. doi:10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: a short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. doi:10.1080/00223890902935878

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 101, pp. 3747–3752). doi:10.1073/pnas.0400087101

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: structure and dynamics. *Physics Reports*, *424*(4-5), 175–308. doi:10.1016/j.physrep.2005.10.009

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*, 113–120. doi: 10.1080/0022250X.1972.9989806

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*(5), 1170–1182. doi:10.1086/228631

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564. doi:10.1016/j.socnet.2007.04.002

Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, *23*(3), 191–201. doi:10.1016/S0378-8733(01)00038-7

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. doi:10.1016/j.socnet.2004.11.008

Borgatti, S. P., & Everett, M. G. (2006). A Graph-theoretic perspective on centrality. *Social Networks*, *28*(4), 466–484. doi:10.1016/j.socnet.2005.11.005

Borsboom, D. (2005). *Measuring the mind: conceptual issues in contemporary psychometrics*. Cambridge: Cambridge university press.

Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. doi:10.1146/annurev-clinpsy-050212-185608

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. doi:10.1371/journal.pone.0027407

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219. doi:10.1037/0033-295X.110.2.203

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, *25*(1994), 163–177. doi:10.1080/0022250X.2001.9990249

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, *30*(2), 136–145. doi:10.1016/j.socnet.2007.11.001

Brandes, U., & Fleischer, D. (2005). Centrality Measures Based on Current Flow. In V. Diekert & B. Durand (Eds.), *STACS 2005 SE - 44* (Vol. 3404, pp. 533–544). Berlin: Springer. doi:10.1007/978-3-540-31856-9_44

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS ONE*, *8*(4), e60188. doi:10.1371/journal.pone.0060188

Butts, C. T. (2008a). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology*, *11*(1), 13–41. doi:10.1111/j.1467-839X.2007.00241.x

Butts, C. T. (2008b). Social network analysis with sna. *Journal of Statistical Software*, *24*(6), 1-51.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81. doi:10.1037/h0046016

Cervone, D. (2005). Personality architecture: within-person structures and processes. *Annual Review of Psychology*, *56*, 423–452. doi:10.1146/annurev.psych.56.091103.070133

Costantini, G., & Perugini, M. (in press). Generalization of clustering coefficients to signed correlation networks. *PLoS One*.

Cox, D. R., & Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statistical Science*, *8*(3), 204–218.

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., … Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: you can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414–431. doi:10.1002/per.1866

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: a network perspective. *The Behavioral and Brain Sciences*, *33*(2-3), 137–193. doi:10.1017/S0140525X09991567

Csárdi G., Nepusz T. (2006) The igraph software package for complex network research, *InterJournal, Complex Systems 1695*(5).

Crucitti, P., Latora, V., Marchiori, M., & Rapisarda, A. (2004). Error and attack tolerance of complex networks. *Physica A*, *340*(1-3), 388–394. doi:10.1016/j.physa.2004.04.031

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (2nd ed.). Cambridge: Cambridge University Press.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271. doi:10.1007/BF01386390

Drton, M., & Perlman, M. D. (2004). Model selection for Gaussian concentration graphs. *Biometrika, 91*(3), 591-602.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18.

Fan, J., Feng, Y., & Wu, Y. (2009). Network exploration via the adaptive LASSO and SCAD penalties. *The Annals of Applied Statistics*, *3*(2), 521–541. doi:10.1214/08-AOAS215

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. doi:10.1037//0022-3514.80.6.1011

Foygel, R., & Drton, M. (2011). Bayesian model choice and information criteria in sparse generalized linear models. *arXiv Preprint arXiv:1112.5635*.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. doi:10.1016/0378-8733(78)90021-7

Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, *13*(2), 141–154. doi:10.1016/0378-8733(91)90017-N

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics, 9*(3), 432–441. doi:10.1093/biostatistics/kxm045

Gignac, G. E., Bates, T. C., & Jang, K. L. (2007). Implications relevant to CFA model misfit, reliability, and the five-factor model as measured by the NEO-FFI. *Personality and Individual Differences*, *43*(5), 1051–1-062. doi:10.1016/j.paid.2007.02.024

Grimmett, G. (2010). *Probability on graphs: random processes on graphs and lattices*. Cambridge: Cambridge University Press.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical Modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*(2), 207–233. doi:10.1207/s15327906mbr4002_3

Humphries, M. D., & Gurney, K. (2008). Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PloS One*, *3*(4), e0002051. doi:10.1371/journal.pone.0002051

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42. doi:10.1038/35075138

Kalna, G., & Higham, D. J. (2007). A clustering coefficient for weighted networks, with application to gene expression data. *AI Communications*, *20*(4), 263–271.

Kindermann, R., & Snell, J. (1980). *Markov random fields and their applications*. Providence: American Mathematical Sociology. doi:10.1090/conm/001

Kolaczyk, E. D. (2009). *Statistical analysis of network data: methods and models*. New York, NY: Springer. doi:10.1007/978-0-387-88146-1

Krämer, N., Schäfer, J., & Boulesteix, A.-L. (2009). Regularized estimation of large-scale gene association networks using graphical Gaussian models. *BMC Bioinformatics*, *10*, 384. doi:10.1186/1471-2105-10-384

Kunegis, J., Lommatzsch, A., & Bauckhage, C. (2009). The Slashdot Zoo: mining a social network with negative edges. In *Proceedings of the 18th international conference on world wide web* (pp. 741–750). doi:10.1145/1526709.1526809

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. doi:10.1186/1471-2105-9-559

Langfelder, P., & Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software, 46*(11), 1–17.

Latora, V., Nicosia, V., & Panzarasa, P. (2013). Social cohesion, structural holes, and a tale of two measures. *Journal of Statistical Physics*, *151*(3-4), 745–764. doi:10.1007/s10955-013-0722-z

Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press.

Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Signed networks in social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1361–1370. doi:10.1145/1753326.1753532

Lykken, D. (1991). What's wrong with Psychology, anyway? In D. Cicchetti & W. M. Grove (Eds.), *Thinking Clearly About Psychology, Vol. 1*. Minneapolis: University of Minnesota Press

McCrae, R. R., & Costa, P. T. J. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *Sage handbook of personality theory and assessment, Vol. 1* (pp. 273–294). Los Angeles: Sage.

Milgram, S. (1967). The small world problem. *Psychology Today, 1*(1), 61-67.

Mõttus, R., Penke, L., Murray, A. L., Booth, T., & Allerhand, M. (2013) *Personality differences without common-cause latent factors are possible and can explain key findings in personality psychology*. Manuscript submitted for publication.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*(3), 167–256. doi:10.1137/S003614450342480

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E, 70*(5), 056131. doi:10.1103/PhysRevE.70.056131

Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, *27*(1), 39–54. doi:10.1016/j.socnet.2004.11.009

Newman, M. E. J. (2008). The physics of networks, *Physics Today 61*(11), 33-38. doi:10.1063/1.3027989

Newman, M. E. J. (2010). *Networks: an introduction*. New York: Oxford University Press.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Phisical Review E, 69*, 026113. doi:10.1103/PhysRevE.69.026113

Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E, 71*(6), 065103. doi:10.1103/PhysRevE.71.065103

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. doi:10.1016/j.socnet.2010.03.006

Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks, 31*(2), 155–163. doi:10.1016/j.socnet.2009.02.002

Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge: MIT Press.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Revelle, W. (2013) *psych: procedures for personality and psychological research.* R package version 1.4.1

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603. doi:10.1007/BF02289527

Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., & Kertész, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, *75*(2), 027105. doi:10.1103/PhysRevE.75.027105

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: a network perspective on psychological phenomena. *New Ideas in Psychology*, *31*(1), 43–53. doi:10.1016/j.newideapsych.2011.02.007

Valente, T. W. (2012). Network interventions. *Science*, *337*(6090), 49–53. doi:10.1126/science.1217330

Van der Maas, H. L. J., Dolan, C. V, Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. doi:10.1037/0033-295X.113.4.842

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. doi:10.1038/30918

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology, 4*(1). doi:10.2202/1544-6115.1128

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. doi:10.1198/016214506000000735

Appendix, Centrality Indices

Node	Dimension	Facet	Betweenness	Closeness	Strength
Hsi	Honesty-Humility	Sincerity	1	2,64	0,70
Hfa	Honesty-Humility	Fairness	**32**	3,03	**1,47**
Hga	Honesty-Humility	Greed-avoidance	15	2,84	1,12
Hmo	Honesty-Humility	Modesty	0	2,11	0,42
Efe	Emotionality	Fearfulness	4	2,69	1,02
Ean	Emotionality	Anxiety	2	3,01	1,09
Ede	Emotionality	Dependence	4	3,01	1,05
Ese	Emotionality	Sentimentality	18	3,16	**1,41**
Xss	Extraversion	Social self-esteem	10	3,11	**1,36**
Xsb	Extraversion	Social boldness	**22**	**3,32**	1,20
Xso	Extraversion	Sociability	7	**3,19**	1,07
Xli	Extraversion	Liveliness	14	3,14	1,31
Afo	Agreeableness vs. anger	Forgiveness	5	2,70	1,00
Age	Agreeableness vs. anger	Gentleness	5	2,65	0,79
Afl	Agreeableness vs. anger	Flexibility	14	2,90	1,03
Apa	Agreeableness vs. anger	Patience	5	2,85	0,85
Cor	Conscientiousness	Organization	8	3,10	0,99
Cdi	Conscientiousness	Diligence	**26**	**3,34**	1,30
Cpe	Conscientiousness	Perfectionism	4	3,14	1,26
Cpr	Conscientiousness	Prudence	**20**	**3,52**	**1,44**
Oaa	Openness to experience	Aesthetic appreciation	11	2,95	1,23
Oin	Openness to experience	Inquisitiveness	5	2,72	1,08
Oun	Openness to experience	Creativity	4	2,68	1,02
Ocr	Openness to experience	Unconventionality	9	2,99	1,25

Note. The four most central nodes according to each index are reported in **bold**.