

Research Article

Expanding Network Analysis Tools in Psychological Networks: Minimal Spanning Trees, Participation Coefficients, and Motif Analysis Applied to a Network of 26 Psychological Attributes

Srebrenka Letina ^{1,2}, Tessa F. Blanken,³ Marie K. Deserno,^{4,5} and Denny Borsboom⁴

¹Department of Network and Data Science, Central European University, Hungary

²HAS Centre for Social Sciences “Lendület” Research Centre for Educational and Network Studies (RECENS), Hungary

³Department of Sleep and Cognition, Netherlands Institute for Neuroscience, Amsterdam, Netherlands

⁴Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

⁵Dr. Leo Kannerhuis and REACH-AUT, Doorwerth, Netherlands

Correspondence should be addressed to Srebrenka Letina; sreb.letina@gmail.com

Received 27 June 2018; Revised 16 October 2018; Accepted 29 November 2018; Published 21 February 2019

Guest Editor: Nicole Beckage

Copyright © 2019 Srebrenka Letina et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The analysis of psychological networks in previous research has been limited to the inspection of centrality measures and the quantification of specific global network features. The main idea of this paper is that a psychological network entails more potentially useful and interesting information that can be reaped by other methods widely used in network science. Specifically, we suggest methods that provide clearer picture about hierarchical arrangement of nodes in the network, address heterogeneity of nodes in the network, and look more closely at network's local structure. We explore the potential value of minimum spanning trees, participation coefficients, and motif analyses and demonstrate the relevant analyses using a network of 26 psychological attributes. Using these techniques, we investigate how the network of different psychological concepts is organized, which attribute is most central, and what the role of intelligence in the network is relative to other psychological variables. Applying the three methods, we arrive at several tentative conclusions. Trait Empathy is the most “central” attribute in the network. Intelligence, although peripheral, is weakly but equally related to different kinds of attributes present in the network. Analysis of triadic configurations additionally shows that the network is characterized by relatively strong open triads and an unusually frequent occurrence of negative triangles. We discuss these and other findings in the light of possible theoretical explanations, methodological limitations, and future research.

1. Introduction

In the last decade, network approaches have been increasingly used in psychological science for the investigation of psychological constructs and their interrelations in psychological science, as complementary or alternative to typically used and well-established methods (e.g., confirmatory factor analysis and structural equation modelling). This approach has introduced a different perspective on psychological constructs and has found its application in many subfields of psychology: intelligence [1], psychopathology [2], personality psychology [3], and social psychology [4]. One specific asset of the network approach is that it defines psychological constructs as constituents of a complex system of direct interactions enabling us to ask detailed questions about

relationships of mutual influence among these constructs [5–8]. Specifically, Gaussian graphical models (GGM [9]) for continuous variables and Ising models for binary variables [10] have been used for network estimation with the aim to describe conditional independence relationships between variables, operationalized as partial correlations or conditional associations between variables [7, 11]. In this approach, a psychological network consists of nodes, psychological variables, and connections between nodes that represent the degree (and direction) of associations between each pair of variables, when the influence of every other variable in the network is controlled for.

After the construction of psychological networks, the quantitative analysis often proceeded with the computation

of a centrality analysis to answer which variable is most “dominant” or “important.” Also, some global features have been of interest, such as network connectivity [12]. However, besides centrality measures and global measures of network structure, which focus on microscopic and macroscopic level of network, respectively, other analytical tools have been mostly ignored and rarely used in the study of psychological networks. This limited focus results in a limited set of questions that can be answered. We argue that, in order to answer research questions using psychological networks, researchers should go beyond the measures commonly used in psychology. The field of network science offers many alternative metrics that are worth considering when translating one’s research question into quantifiable network properties. The main idea of this paper is to apply such techniques, which are already widely used in network science, to provide deeper understanding of psychological networks.

The structure of the paper is as follows. Firstly, we will describe some of the challenges in the analysis of psychological networks and link them with the three methods we propose in this paper, following with the general overview of the network that will be used for the demonstration of these methods. Next, we describe an illustrative dataset and apply the methods typically used in network analysis. Subsequently, we explain three methods that can be used to shed light on the network topology: minimum spanning trees (MSTs), the participation coefficient (PC), and motif analysis. For each method, we will explain specific procedures and modifications and conclude with results and discussion. Finally, in the general discussion, we summarize the benefits and possibilities of including the proposed methodologies in the field of network psychometrics and highlight interesting hypotheses that we arrived at using these analytical tools.

1.1. Identifying Challenges in the Analysis of Psychological Networks. In this paper, we propose three methods that not only provide novel insights into the network, but also circumvent some prominent methodological issues in the field of psychological networks: finding a way to operationalize the importance of all variables included in the network in a more general way, dealing with network of variables that are not of the same kind, and how to investigate the intermediate network level.

(1) *Finding the hierarchical arrangement of nodes in the network:* The main purpose of centrality analysis used in the analysis of psychological networks so far was to determine how entities in the network may be ordered regarding their connections with other variables (e.g., using the number and strength of connections) and regarding their overall position in the network, that is, to find out which entity is the most “dominant.” The answers that arise from the application of different measures (typically strength, betweenness, and closeness) are likely to be different, as all of them capture different notions of what centrality means. However, the selection of the “right” measure is not the only challenge. Due to the small and dense nature of psychological networks, centrality measures may not meaningfully differentiate among specific nodes.

As a solution to those issues we suggest the use of the minimum spanning tree (MST), applied firstly on economics in the stocks analysis of time-series data [13]. The MST is a reduced subnetwork that connects all nodes based on the identification of the minimal set of edges needed. Besides providing a topological and hierarchically arranged skeleton of all nodes in the network, it additionally provides an insight into groupings of nodes based on their content similarity.

(2) *The implicit assumption about the homogeneity of nodes in the network:* Most commonly used centrality measures are based on a node’s relation to every other node in the network. Thereby, these techniques implicitly assume that all nodes are a priori equally likely to be connected with any other node. This assumption is often untenable, as psychological networks may include one or more entities, or groups of entities, which differ in nature and/or measurement and therefore constitute a cluster (referred to as community or module, e.g. see [14]). In psychology, such a community may arise in part because of preexisting differences between the variables in, for example, nature of the variables (e.g., cognitive, behavioral, and emotional), kind of measurement (e.g., subjective vs. objective), or some methodological aspect of data collection.

In the estimated network, variables that are more similar regarding these preexisting differences (i.e., that belong to the same community in this sense) are more likely to be associated than variables belonging to different communities. Thus, these variables may show stronger associations among themselves and will *by construction* rank higher on common centrality measures like degree and strength. Note that this effect is especially pronounced when the size of different communities is not equal, as nodes belonging to the biggest community will by default have higher degree and strength. On the other hand, if some variables are different in some of the aforementioned ways from other variables included in the network, they may by default be expected to have less strong connections with other variables in the system. As a result, we might wrongly identify some node as central while, at the same time, a variable with a truly important role might be missed. This is important because psychological networks are increasingly starting to include psychological entities of different kinds. For example, recently some researchers [15] called for inclusion of other variables besides symptoms when analyzing psychopathological systems.

To circumvent the issue of nodes’ heterogeneity, we propose the participation coefficient (PC [16]) to be used as a corrective in the procedure of estimating the most central node, because it addresses the uniformity of the edges a node has to different groups of nodes in the network.

(3) *The network’s mesolevel (or local structure):* Visualization of small networks, such as psychological networks, provides immediate insights into the dyadic relationships between nodes, at the network as a whole, and even can provide some notion of the grouping of nodes. Similarly, measures typically used in the analysis of networks of this kind cover analyses at the microscopic network level. Macroscopic (global) properties of a network are easily computed, although their usefulness is less clear in psychological networks due to their small size and the impossibility to

claim that all relevant nodes are included in the network. The interpretation of commonly used centrality measures and global measures of network structure (e.g., average shortest path and clustering coefficient) as reflecting the importance of nodes in the system implicitly assume that the network contains all factors that are relevant to the system. However, one inherent characteristic of psychological networks is that it rarely models all factors that are relevant to the system [15]. In these cases, computing centrality measures based on indirect ties (betweenness and closeness) and global network measures may not capture all relevant information. While this is a problem when analyzing the entire system, much can be learned from shifting the focus to structural patterns on a more fine-grained level (i.e., mesoscopic level, “local” network structure). Methods for investigation of small configurations in network have been first developed in social network analysis [17] and have been redefined when applied to different types of (usually large) networks (e.g., neuronal networks, transcriptional networks, and the structure of the Internet) at the beginning of the century and have become known as “motif analysis.” Motif analysis enables researchers to systematically investigate smaller configurations of nodes. It can help us determine, among other things, whether certain patterns, that is, subgraphs, represent interesting relations between constructs or methodological artefacts.

Moreover, this method addresses one of the basic questions in modeling networks: how global properties of networks can be understood from its local properties and how local topology is related to function [18]. For example, in psychological networks, different measures of intelligence are known to correlate positively; they show a positive manifold. In the language of network mesolevel analysis, this means that the system of different intelligence measures is characterized by smaller local structures that display positive relationships with each other. Van der Maas et al. [1] proposed a dynamical model of intelligence in which these patterns are interpreted as indicating that reciprocal causation or mutualism plays the most important process in that system. In other words, if a network expresses certain pattern of relationships in “high” degree, it may inform us about underlying process(es) driving the system that is represented as the network.

Each of the three methods, and especially the last two (the participation coefficient and motif analysis), give a clearer picture of *all* nodes in the network. It could be argued that they provide a *more “egalitarian” approach* to nodes that constitute a network, in a sense that they allow finding that noncentral (in terms of strength, betweenness, or closeness) nodes can be equally important for different parts of network or have an interesting role in a smaller part of network. That information can be easily overlooked when using only most basic network analytics. Given that psychological networks are usually relatively small, it is plausible that researchers will be interested to learn more about each node in the network, whether it is central or peripheral. Moreover, sometimes nodes that are peripheral can be of special interest and/or relevance (e.g., suicidal ideation in the network of depression symptoms and intelligence in the network of psychological traits).

1.2. Applying Three Methods in the Investigation of the Network of Different Psychological Attributes. Network analysis has been used mostly for looking more closely at one (or several related) psychological concepts, where nodes represent psychometric items that are part of a self-report measure (e.g., a questionnaire). In the current study, as an illustrative dataset for the proposed methods, we look at a network in which nodes are aggregated scores on self-report measures (also known as “parcels” of a questionnaire) that operationalize different psychological concepts (e.g., latent variables), most of which are not highly related, and among which direct causal relations may not be assumed. The variables in our network are supposed to measure relatively stable individual differences whose development “proceeds along mutually causal lines” [19, p.239]. Moreover, the conditional associations between those constructs are likely to be small, as most of them are assumed to be independent. To the best of our knowledge, this is the first research that looks at the network of different psychological attributes presented as aggregated items. We use network approaches to gain new insight into how different parts of that psychological system are connected, and which attributes have the most prominent role.

In the network of psychological constructs measured by self-reports we included cognitive ability (a proxy of g-factor [20]) measured with ability test (sometimes referred to in psychology as subjective and objective tests, respectively). The reason for including this substantially different variable in the network is twofold. First, we aim to demonstrate network methods that can provide more nuanced descriptions of all nodes, whatever their centrality in the network is. Including a variable, a node, which is known to be conceptually and methodologically different from others in the network, and at best only modestly associated with just some of nodes in the network, will set the stage for demonstrating added value of proposed methods. Second, we use the opportunity to address the old question of how cognitive ability and personality are related [21], to see how this question can be formulated and answered within the network approach.

Theoretically, intelligence is not expected to correlate with personality domain. For decades, researchers dealing with personality–intelligence connection have been using correlational studies to identify if significant relationship(s) exist(s). Yet, as Eysenck [22] in his review of the topic concludes, the research showed a striking lack of significant correlations, with few exceptions. For example, small associations have been found between intelligence and psychopathological profile [23], and introversion–extraversion related differences in style of intellectual performance (speed/accuracy ratio; [24]). Seeing that this approach failed to find any substantial relationship, Salovey and Mayer [25] suggest that question should be asked in a more complex way, for example, looking at the difference in the factorial structure of intelligence for groups with different personality profiles, and vice versa. Analytically, this suggestion is very much in line with network approach, because it looks at the whole set of variables at once, and is not as much focused on the size of specific effects. From a theoretical perspective, several attempts of an

integrative approach to both personality and intelligence with a wider theoretical framework for understanding their interrelations can be found in the literature, for example, social intelligence theory within cognitive theory of personality [26] and Motivational Systems Theory [27]. They are closely related to Smirnov's [28] view of intelligence as thinking, and personality as inherent component of all thought processes, while the link between the two is goals and problems in daily life.

2. Methods

2.1. Data and Measures. The dataset used in the current study has been collected within the context of the *myPersonality* project [29, 30]. In this project, participants self-administered one or more psychological questionnaires online, through a Facebook application (active from 2007 until 2012). Participation was voluntary and completely anonymous, and participants provided consent. In total, more than 20 different questionnaires were offered, and participants completed a self-chosen, variable number of questionnaires at a self-chosen place and time.

Of the available questionnaires, we selected 11 questionnaires, covering 31 psychological attributes, guided by three criteria: we wanted to include psychological concepts that (i) have a clear theoretical background and were measured with validated instruments with good psychometric properties; (ii) are considered to have high temporal reliability and stability; and (iii) had relatively high number ($N > 1000$) of participants who also self-administered other questionnaires. To prevent including concepts that are too similar, we excluded concepts that correlated very highly to other concepts (around 0.60 in absolute correlations) and that had a clear theoretical overlap. This resulted in the inclusion of 26 psychological concepts. To facilitate interpretation, we reversed the scores of the negatively framed variables (Neuroticism, Depression, Militaristic values, and Violent occult interests) such that all variables can be interpreted as higher scores representing more favorable outcomes, except for Schwartz's values, where such rationale was not possible since having or not having high scores on certain value should be evaluation-free, meaning not positive or negative by default. The interpretation of the variables after recoding is listed in Table 1. More information on data processing, sample description, description of missing data, and descriptive statistics of 26 psychological variables is offered in the Supplementary Materials (SM, Sections 1-4).

We included 1,166,923 participants with a score on at least two of the psychological attributes (hereafter: variables). Of a subsample of participants, demographic information was available on gender (44.6%, of which 64.8% female and 35.2% male) and age (20.8%; $M \pm SD = 26.1 \pm 6.7$, range: 14-89 years). The sample consisted of participants from 220 different countries, and 35.7% of participants were from the US, UK, Canada, Australia, and India, respectively. A concise description of the included constructs and the instruments used is given in Table 1.

2.2. Network Estimation. We used partial correlations to estimate (For network estimation, visualization, and centrality analysis following R packages were used: *BDgraph* [42], *qgraph* [43], and *networktools* [44]. MST, PC, and motif analysis is done in *NetworkX* Python module [45]. Code used can be provided from the first author upon request.) the network. Partial correlation networks do not contain spurious correlations that are generated by common cause and chain structures within the network and can encode a basic data-generating network structure [46]. To estimate the network, we used a nonregularized method recently proposed by Williams and Rast (in press) [47] because, given our large sample size, relatively small number of variables, and our interest to detect weak ties, it is not advised to use regularization techniques like the LASSO that are often used ([47, 48], *in press*). More details about the process of determining the optimal estimation method for our data, and about the nonregularization method used, can be found in SM, Section 5.

To prevent the inclusion of spurious edges because of our overall large sample size, we artificially reduced the sample size by setting the N parameter in the estimation to $N=4131$ (i.e., the median number of completed pairwise observations, for more details see SM, Section 3) instead of the total sample size of $N=1166923$. The estimated network is shown in Figure 1. The included edges were significant at alpha level of 0.001. Note that partial correlations are usually *smaller* than first-order correlations when interpreting the edge weights.

At first glance (Figure 1) at the network it can be seen that most of the nodes from the same group (questionnaire) cluster together in the network, except for Big Five traits that are more scattered across the network, especially Openness.

2.3. Robustness Analysis. To check robustness of our results, we tested it in two ways. First, we randomly split the sample in half 100 times and estimate a network on each half separately. Subsequently, we compare the two estimated networks on a metric of interest. If the network estimation is reliable, then the networks should be similar for both halves of the data, and, hence, the metrics should show high correspondence. This procedure is similar to that of Forbes et al. [49]. It should be noted, however, that, by using only half of the data to estimate a network, the statistical power drops considerably which will especially affect the estimation of small edges. Therefore, we conducted a second robustness analysis in which we randomly selected 100 sets containing 80% of the original sample and compared the network estimated on this subsample to the network estimated on the complete dataset. We computed the average correlation of the pairs of matrices estimated for the split halves (robustness analysis I) and between the whole sample and the random (80%) fractions (robustness analysis II). For the split halves, the average correlation was 0.82, indicating a high level of reliability. However, if we only evaluate the edges that are present in both estimates, on average, the reliability drops to 0.59 (similarity index). The average difference in the number of edges is 6.35, which is around 2% of all possible edges. For the random (80%) fractions, the similarity index increased to 0.85. The results are presented in more depth in SM, Section 6.

TABLE 1: Description of 26 psychological attributes included in the network.

Psychological attribute (or <i>Group of attributes (number of attributes in the group)</i>), <i>Questionnaire (author(s))</i>
Short description of measured attribute (number of items)
<i>Values - based on Schwartz Theory of Basic Values (6), Schwartz Value Survey – SVS (Schwartz, 1992) [31]</i>
Achievement - personal success through demonstrating competence according to social standards. (4 items)
Hedonism - pleasure or sensuous gratification for oneself. (3 items)
Power - social status and prestige, control or dominance over people and resources. (4 items)
Self-direction - independent thought and action—choosing, creating, exploring. (5 items)
Tradition - respect, commitment, and acceptance of the customs and ideas that one's culture or religion provides. (6 items)
Universalism - understanding, appreciation, tolerance, and protection for the welfare of all people and for nature. (8 items)
<i>Big Five Traits (5), 20–100-item IPIP questionnaire (8 length versions), also included data on 336-item IPIP Personality Facets questionnaire (Goldberg et al., 2006). Both questionnaires are proxies for Costa and McCrae's NEO-PI-R facets (Five Factor Model) [32]</i>
Emotional Stability (reversed Neuroticism) - the tendency not to experience negative emotions, such as anger, anxiety, or depression.
Extroversion - characterized by positive emotions, surgency, and the tendency to seek out stimulation and the company of others.
Openness to experience - a general appreciation for art, emotion, adventure, unusual ideas, imagination, curiosity, and variety of experience.
Agreeableness - tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others.
Conscientiousness - tendency to show self-discipline, act dutifully, and aim for achievement.
<i>Interests (4), The Sensational Interests Questionnaire – SIQ (Egan et al., 1999) [33]</i>
Low militaristic interests (reversed Militaristic interests) – an individual with low active interest in militaristic activities (e.g. guns and shooting). (10 items)
Low violent-occult interests (reversed Violent-occult interests)– an individual with low active interest in violent or occult activities (e.g. black magic). (7 items)
Intellectual interests – an individual's active interest in cerebral activities (e.g. philosophy). (7 items)
Interest in wholesome activities – an individual's active interest in active recreation (e.g. camping, hill walking). (5 items)
<i>Body Consciousness (3), Body Consciousness Questionnaire –BCQ (Miller, Murphy, & Buss, 1981) [34]</i>
Private body - awareness of internal sensations. (5 items)
Public body - awareness of observable aspects of body. (6 items)
Body competence – self-confidence in the body's performance. (4 items)
<i>Integrity assessment (2), Rust's Sense of Fairness and Impression Management, Orpheus (Rust & Golombok, 1989), 36 items. [35]</i>
Fair-mindedness (or <i>Sense-of-fairness</i>) – measures how balanced and impartial person is in her decision making.
Self-Disclosure – measures to what extent a person conducts her life transparently. Reversed values are used as a measure of Impression Management and Social desirability (Lie scale).
<i>“Stand-alone” traits – six psychological attributes which are not part of a group of constructs, each is measured with separate questionnaire</i>
Awareness of physical symptoms and sensations , <i>Pennebaker's Inventory of Limbic Languidness - PILL (Pennebaker, 1982) [36]</i>
Scale measures how often a person notices and reports a broad array of physical symptoms and sensations (e.g. chest pain, heart racing, dizziness). (54 items)
<i>Self-monitoring, Snyder's Self-Monitoring Scale, (Snyder, 1974) [37]</i>
Scale measures how much person monitors her self-presentations, expressive behavior, and nonverbal affective displays. (25 items)
Low Depression , <i>Center for Epidemiologic Studies Depression Scale (CES-D), NIMH, (Radloff, 1977) [38]</i>

TABLE 1: Continued.

Reversed Depression, measures lack of symptoms of depression in nine different groups as defined by the American Psychiatric Association Diagnostic and Statistical Manual, fifth edition. (20 items)
Empathy , <i>Empathy Quotient - EQ</i> , (Baron-Cohen & Wheelwright, 2004) [39] Scale measures self-reported ability to tune into how others are feeling, and to understand what they may be thinking. It measures both the affective and the cognitive components of empathy. (60 items)
Life satisfaction , <i>Satisfaction With Life Scale- SWLS</i> (Diener, Emmons, Larsen, & Griffin, 1985) [40] Scale measures general wellbeing and satisfaction with one's life. (5 items)
Intelligence , <i>MyIQ test, myPersonality's</i> 20-item proxy for Raven's Standard Progressive Matrices (Raven, 2008) [41], University of Cambridge's Psychometrics Centre (Chan & Kosinski) Ability test measures cognitive skills and clear-thinking ability, and pattern recognition abilities known to have the highest correlation with the general intelligence factor. (20 items)

TABLE 2: Description of ties in partial correlation network.

	Signed ties	Absolute weights	Positive ties	Negative ties
Mean	0,01	0,13	0,13	-0,13
SD	0,158	0,090	0,092	0,088
Min.	-0,39	0,05	0,05	-0,39
25%	-0,09	0,06	0,07	-0,16
Mdn	0,06	0,10	0,10	-0,10
75%	0,11	0,16	0,15	-0,06
Max.	0,53	0,53	0,53	-0,05
N. of ties	144	144	80	64

3. Illustrative Results: Network Description

3.1. Edge Weights in the Network. The current estimated network has 144 edges out of 325 possible edges, showing a good balance between sparsity and density (Figure 2). The distribution of the edges is summarized in Table 2; 64 edges (44%) are negative and 80 edges (56%) are positive. The number of negative edges is higher than usually observed psychological networks. Note that this is dependent on the network under consideration. If a network includes variables that all come from the same questionnaire (e.g., 10 depression items), then it would be expected that many (or all) edges are positive. In the current network, variables from various psychological questionnaires are included; they are not expected to correlate highly or/and positively by definition. Figure 2 also shows that, due to artificially decreasing statistical power and due to setting alpha to 0.001, edges around 0 are eliminated (< 0.05 in absolute value). For more details on the correlation network and estimated partial correlation network, and detailed analysis of ties, see Sections 7 and 8 in SM.

3.2. Centrality of Nodes. In addition to centrality measures that are typically used in psychological networks, we include more recently developed measures of node's expected influence ([50], for short explanation see Section 9 in SM).

Centrality measures can roughly be categorized into two groups, measures that look only at the local surroundings of a node (i.e., only the edges adjacent to the node) and measures that try to quantify the position of a node in the network by

also taking into account nodes that are not directly adjacent to the node. Figure 3 shows centrality measures of the first category—considering only adjacent nodes. Figure 4 shows centrality measures of the second category—considering nodes beyond those directly adjacent to the node of interest. Comparing the different centrality measures, both within the same category or across categories, clearly shows that the measures diverge. Thus, different centrality measures indicate different nodes to be the most central. Although this follows logically from the way the different measures are computed, as each measure captures different aspects of centrality, it highlights the need to carefully consider the metrics used as it can greatly influence the answer to the question that is posed.

As Figure 3 shows, based on a node's direct ties, the most central node varies across measures. Based on strength, the value Tradition is the most central node, followed by Empathy, Extraversion, and another value, Universalism. Among the least central nodes are Agreeableness, Body Competence, and Awareness of Physical Symptoms.

Alternatively, when centrality measures consider more than the local environment of the node, a different arrangement of centrality emerges (Figure 4), with less agreement between different measures. Here, Empathy is the most central node, followed by Extraversion and Emotional Stability, while Tradition drops to the fourth place. The least central nodes are Self-Disclosure, Intelligence, and Awareness of Physical Symptoms.

Robustness analysis of all centrality measures used in this study is presented in Section 6 of SM.

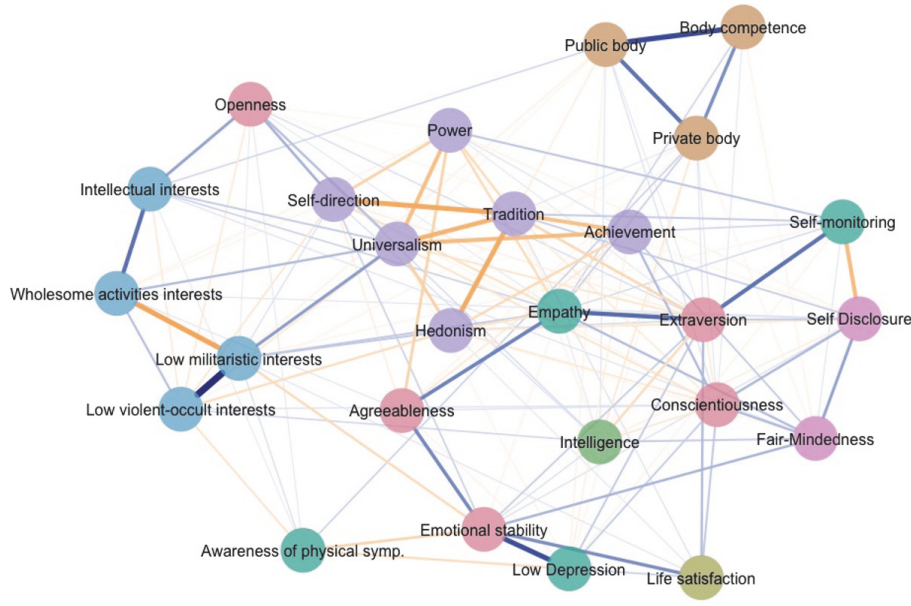


FIGURE 1: Nonregularized partial correlation network (set $N=4131$, true $N=1066921$, layout spring, cut = 0). Blue ties signify positive relations and orange ties signify negative relations. The thickness of a tie is proportional to its absolute weight.

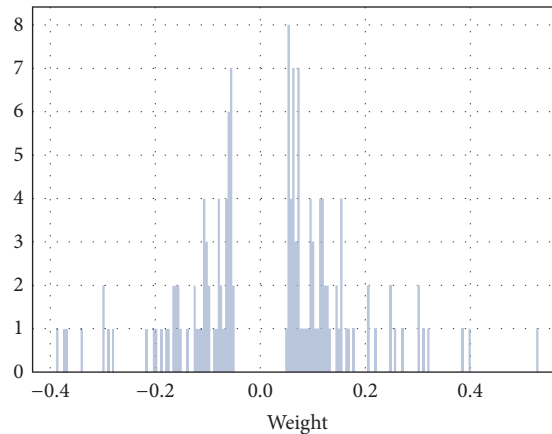


FIGURE 2: Distribution of weights in partial correlation network ($L = 144$).

4. Introducing Three Network Methods for the Analysis of Psychological Networks

4.1. The Minimum Spanning Tree. As demonstrated in previous section, different centrality measures capture different aspects of a node's position in the network, and the centrality of a node will differ depending on the centrality measure used. For that reason, we propose a way to look at the question about centrality differently, in a more general way. To be clear, we are not stating that centrality measures used so far in the research are inadequate, but we are merely trying to ensure a more general perspective to centrality. An alternative way to characterize relationship between all nodes in a network is by computing the minimum spanning tree (MST) [13]. The MST detects the hierarchical organization of the nodes and reduces the number of edges to those that carry the most

information on the similarity of the nodes. Specifically, the MST is based on the distance between the nodes and selects the subset of edges (*number of nodes* - 1) without cycles, and with minimal total distance possible. This "skeleton" structure of the filtered network may be used if we want to get the answer to the general question which node is the most central, by not looking at the specific centrality aspects, but instead focusing on the network's most essential and local ties.

To compute the MST of our current network, first the distances among the nodes must be computed. An appropriate function for converting correlation to distances when negative correlations are present is as follows:

$$d(i, j) = \sqrt{2(1 - r_{ij})} \quad (1)$$

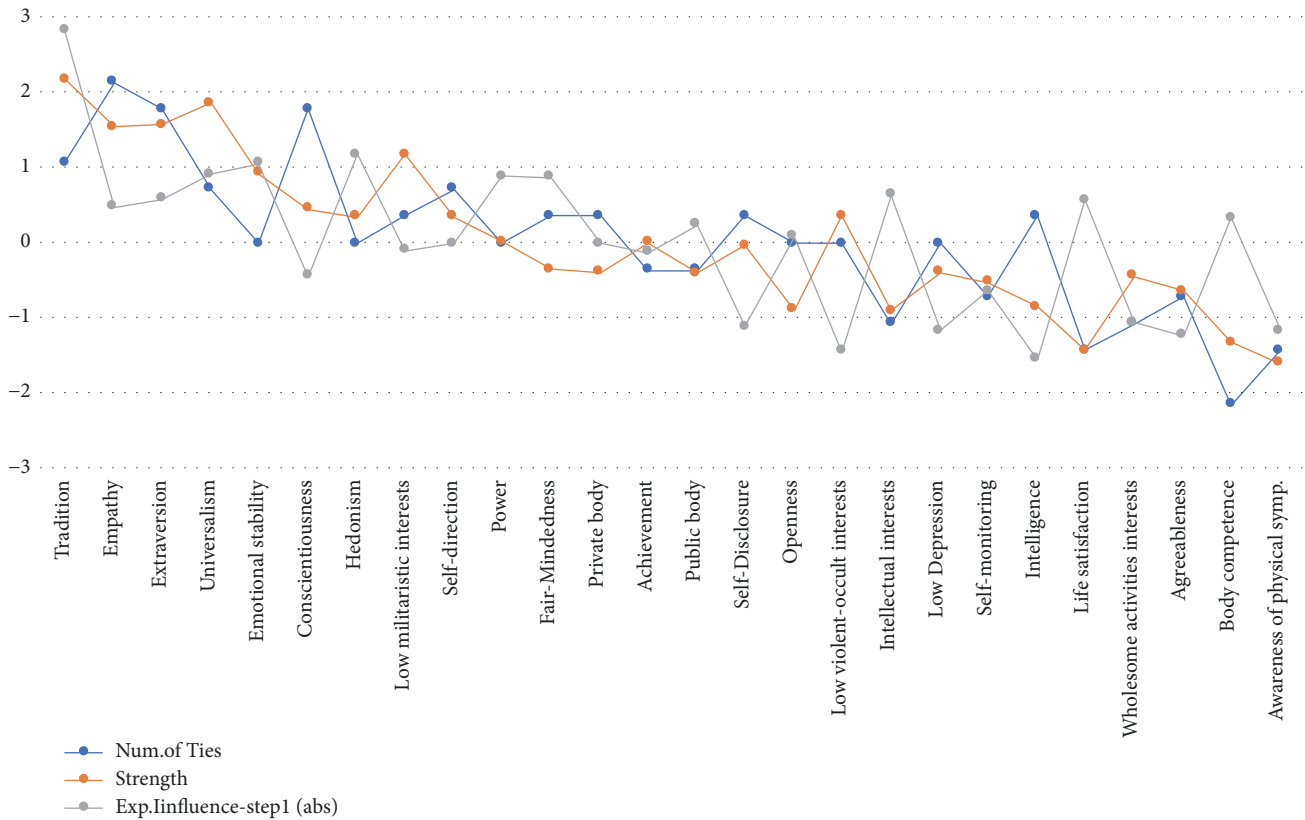


FIGURE 3: Centrality measures 1: based on node's direct ties (standardized values).

Equation (1) (Gower's distance measure [13, 51]) takes the direction of the correlation into account by assigning the largest distance to a perfect negative correlation, and the smallest distance to a perfect positive correlation. According to this equation the distances range from 0 to 2, where an intermediate distance of 1.4 is assigned to variables that are uncorrelated. The relationship between the (partial) correlation coefficient and the distance measure is shown in Figure 5.

Equation (1) is the preferred distance measure to distance inversely proportional to shared variance ($d(i, j) = 1 - pr_{ij}^2$). From the mathematical point of view, it is a more rigorous definition of distance and it gives monotonic transformation of coefficients. Most importantly, (1) gives more differentiated measure of distance than distance based on the shared variance, because in the latter the loss of information occurs since it translates partial correlations of the opposite sign and the same absolute values to the same distance. If negative ties are not present in the network, both measures will produce the same MSTs; otherwise the output will most likely differ (MST based on the shared variance is shown in SM, Section 11, Figure 15). Given the mentioned advantages and since almost half the ties in our network are negative, we have chosen to use it for MST construction. However, as it will be discussed in Section 5 and analysed in SM, Section 12, this measure is sensitive to reverse coding of variables included in the network.

Note that taking partial correlations instead of correlations when calculating distances means that, for each pair of nodes, it indicates how distant they are after the similarity based on covariance with other nodes in the network is excluded.

The MST of 26 psychological attributes is shown in Figure 6. The information about "centrality" of a node is very clear from the hierarchical structure, although centrality measures can provide more detailed picture (see SM, Section 10). The nodes with more direct edges and closer to the middle (centre) of the tree are most central.

Empathy is the most central node in the MST in the sense that it features the smallest distance to all other attributes. From Empathy, four branches emerge with only Sensational Interests and Body Consciousness being on the same branch as all other attributes from the same group. All branches are heterogeneous regarding the group of attributes they consist of, but they can be interpreted as having some commonalities in meaning. The branch with three Body Consciousness constructs along with Awareness of Physical Symptoms captures attributes related with body perception in general. The branch starting with Low Militaristic Interests can be interpreted as representing interests, values, and openness, which are related to what is often referred to as "lifestyle." The branch that starts with Extraversion relates to the attributes that describe one's agency and control in social world. Finally, the biggest and most heterogeneous

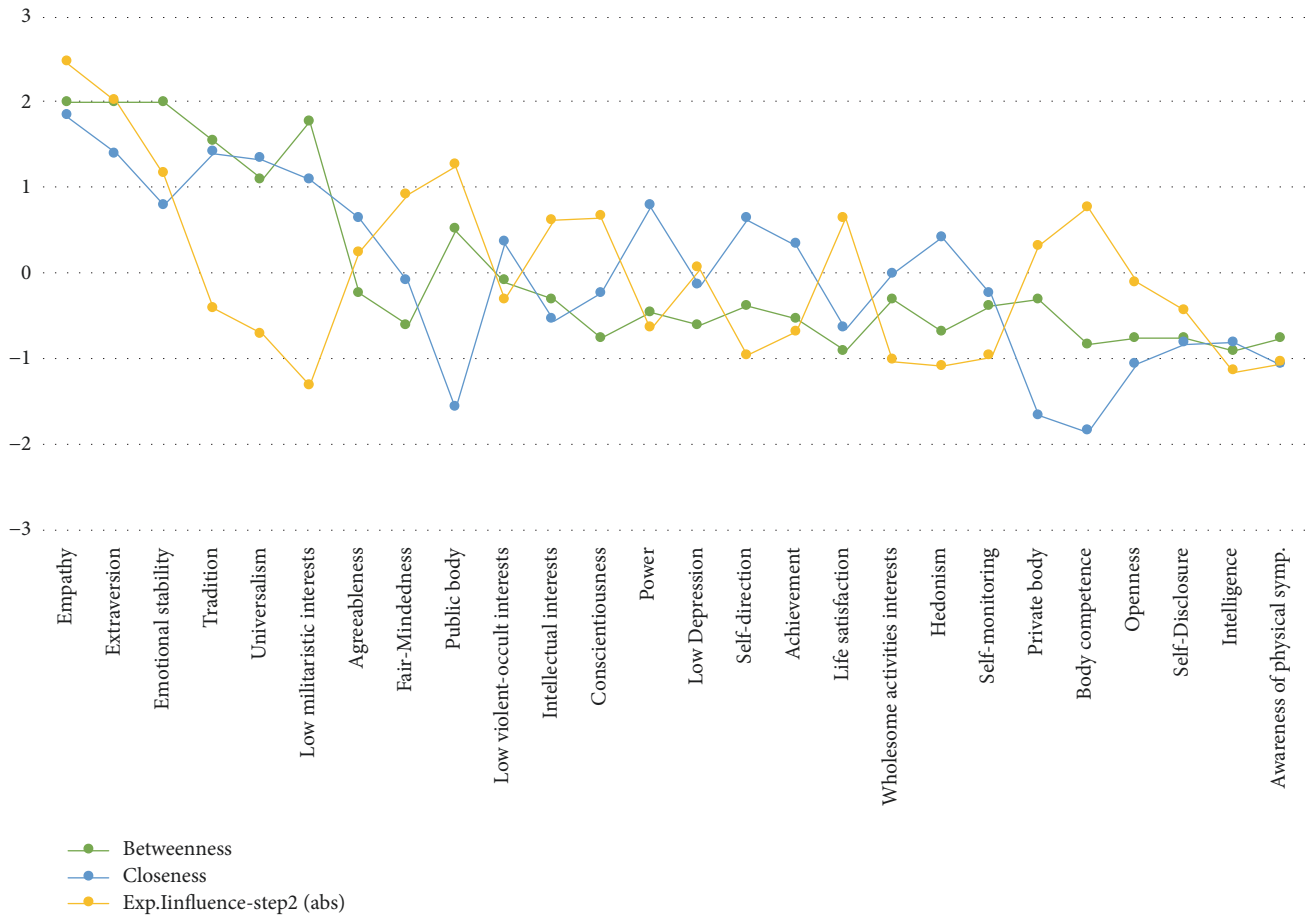


FIGURE 4: Centrality measures 2: based on links more than one distance away from the node (standardized values).

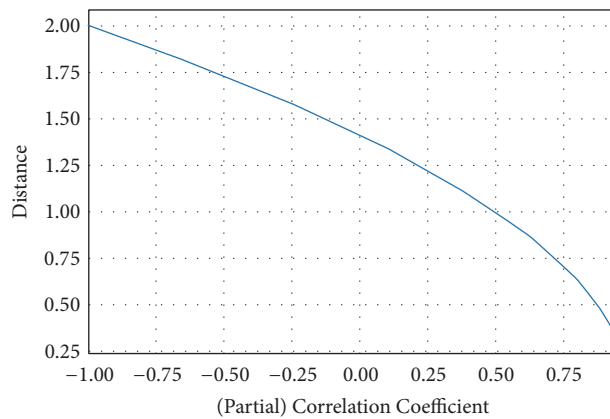


FIGURE 5: The relationship between partial or correlation coefficients and distance measure.

branch starting with Agreeableness is made of attributes that are highly socially esteemed and describe one’s “relation” to others, oneself, and life in general. It is interesting to observe that Intelligence is placed on that branch and it branches out from Fair-Mindedness. This visual inspection shows another useful feature of MST; it gives indirect information on the hierarchical and overlapping, data-driven, clusters in the network. For example, in Figure 6, we can see two

pairs of branches, or clusters, which *overlap* in Empathy. Alternatively, taking Empathy as the origin, there are four branches, or clusters, that overlap in that node.

According to the MST based on the distance defined in (1), two nodes are more distant in terms of steps (ties) between them in the filtered network (tree) if they are negatively associated than if they are not associated at all. That is why, for example, Tradition and Self-Direction (*pr*

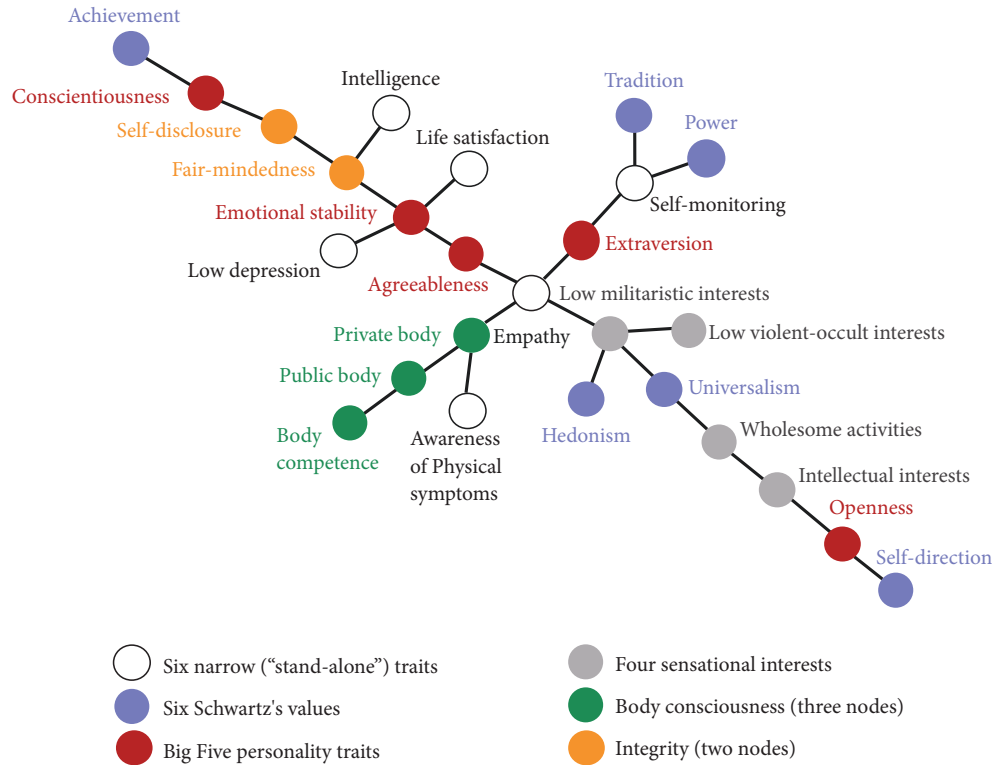


FIGURE 6: Minimum spanning tree (MST) based partial correlation.

$= -0,37$) are placed on different branches and are more distant than Emotional Stability and Conscientiousness ($pr = 0$) that lie on the same branch. From the perspective of psychological networks, the MST preserves the specific content and meaning of the variable. More importantly, since its construction was affected by signs of weights, not only their absolute value, this filtered network can be a useful tool in testing whether two networks made of the same nodes really differ. Two networks estimated on two different samples will not usually be identical. However, if their MST is the same or very similar, this may indicate that their differences are not important. Similarity indexes of MSTs converge with the similarity indexes of whole networks. Nevertheless, reliability based on MST correlations seems to be lower than that based on network correlations in smaller samples (split halves), indicating that in fact the most informative ties are differently estimated (for details see Section 7 in SM).

4.2. The Participation Coefficient. In psychological networks, nodes (variables) may differ in their nature. Some may come from the same framework, while some may be stand-alone nodes. In network parlance, some nodes are part of a community and some nodes form a community of one or few. Note that these communities are not derived from data, but rather they are based on preexisting differences.

In the current dataset, for example, we had 26 psychological concepts, measured by 11 questionnaires. As such, there are groups of variables, varying in size, that belong to the

same questionnaire and that are part of the same theoretical framework (e.g., three concepts on body consciousness) or measure the same kind of trait (e.g., measures of different "values"). Moreover, the psychological concepts that are part of the same questionnaire will likely be completed at the same time, while different questionnaires may have been taken days, months, or even years apart. Therefore, it is important to take these preexisting differences into account, if we want to explore which of the variables play an important role in the network.

One way to deal with these theoretically defined, preexisting communities is by employing measures that take this community structure into account and specifically evaluate connections a node has with nodes in different communities. One such method is the Participation Coefficient (PC), first introduced in the field of biological networks [16]. The PC takes the community structure into account, as it specifically quantifies how the edges a node has are distributed to different communities (similar in logic to Shannon entropy measure.). The important departure in our application of the PC is that it is not used on an empirical community structure, but rather on "communities," that is, groups of nodes and "stand-alone" nodes that were considered to exist in the network (a kind of "ground truth"). Framed as a hypothesis, the null hypothesis in the use of PC would state that preexisting groups of constructs (or data-driven communities) do not influence centrality scores of nodes. Showing that the rank order of nodes according to given centrality measure changes once the measure is corrected

with PC can be interpreted as supporting the rejection of the null hypothesis.

The calculation of the PC measure follows

$$PC_i = 1 - \sum_{m=1}^G \left(\frac{k_{i,m}}{k_i} \right)^2 \quad (2)$$

where PC_i signifies the PC score for a node i , while G , m , $k_{i,m}$ and k_i denote the network, each module in the network, number of ties of node i with nodes in that module, and number of all node's ties, respectively. The expression $k_{i,m}/k_i$ is simply the ratio of all node's ties that go to the specific module. In a version for weighted networks the number of links ($k_{i,m}/k_i$) in (2) is replaced with the sum of strengths which means that the expression $s_{i,m}/s_i$ signifies proportion of total strength of node i , invested in a single module:

$$PC_i = 1 - \sum_{m=1}^G \left(\frac{s_{i,m}}{s_i} \right)^2 \quad (3)$$

This difference means that if a node has the same number of links to every module, but they differ in strength, it will not achieve a maximum PC value. Here, strength is defined as the sum of absolute weights of all links involving node i , which means we disregard the sign of ties.

If a node has an equal number of edges to all the communities in the network (i.e., a uniform distribution of edges to all communities), the PC is closer to 1 (the highest possible value depends on the number of modules in the network; therefore average PCs of different networks can be compared only if PCs are normalized by theoretical maximum value, which is 0.50 for 2-module community, 0.80 for a network containing 5 communities, and, in our network containing 11 communities, the maximum PC value is 0.96). Alternatively, if a node has edges only to nodes within its own community, the PC is 0. It is important to note that the PC is not simply the number of links a node has to other communities in the network, but it rather quantifies the *equality of the distribution* of edges a node has to the other communities. In weighted networks, the PC is maximized if a node is connected equally to all the communities in the network: equal in both the number and the strength of edges to the other communities (i.e., a uniform distribution of edges and edge weights to all communities). More uniform distributions of nodes to all other communities correspond to higher PC values. For example, a node with one tie to each module will have the same PC as a node with two ties to each module. Similarly, a node with just one link to *each* module will have a higher PC than a node who has many links to some, but no links to other modules. A node with a high PC can influence all parts of the network *equally*, meaning that the node is equally important to every defined community. Such a node can be seen as a common denominator in terms of its potential influence on all communities in the network and can therefore help us understand the network as a whole. Note that PC considers only the node's direct ties, displaying the local perspective as MST. Moreover, that feature makes it very suitable for the analysis of a network where some elements of the network may not be included, and where therefore

measures relying on the whole network (e.g., betweenness and closeness) may not be appropriate. However, since the PC solely quantifies the equality of the distribution of ties (or strength of those ties, in version for weighted networks) and disregards number (sum of strengths) of ties, we propose to use it in combination with a measure that considers both the number and the strength of the connections a node has and disregards the information about communities (preexisting or otherwise). One such measure is the Participation Ratio (PR [52]). Participation Ratio is defined [2] with the following formula:

$$C_D^{w\alpha}(i) = k_i \times \left(\frac{s_i}{k_i} \right)^\alpha = k_i^{(1-\alpha)} \times s_i^\alpha \quad (4)$$

where $C_D^{w\alpha}(i)$ is Participation Ratio of node (i), k_i is number of ties of node (i), s_i is the strength of the node, while α is a positive tuning parameter. If its value is set between 0 and 1, having a high number of ties (degree) increases $C_D^{w\alpha}(i)$, if $\alpha = 0$, it is equal to the node's strength, whereas, if α is set above 1, the number of ties decreases the value of $C_D^{w\alpha}(i)$, in such a way that a node with a greater concentration of its strength on only a few nodes and low degree has higher value than a node with the same strength but more ties. In our analysis, the α is set to 0.5, so that, for example, if a node A has a higher number of links and the same total strength as node B, the node A will have higher value of $C_D^{w\alpha}$. In this way both having high total strength and having more ties is favoured.

In short, PR is a single measure that quantifies both the number of edges a node has and the strength of these edges and weighs both equally (i.e., corresponding to an alpha of 0.5), and, as other measures defined so far in this paper, focuses only on node's direct links.

We transformed both measures to the same scale (range 0-1), visualized in Figure 7. Subsequently, for each node we computed the geometric mean of both measures. We opted for the geometric mean as it rewards consistency in scores on the two different measures. For each node, the PC, PR, and its geometric mean are shown in Figure 8.

Interestingly, as can be seen from Figures 7 and 8, the PC and PR can diverge for some nodes. For example, if we only focus on the number of edges and their strength, which is summarized in the PR, Tradition is highly central. However, Tradition has a relatively low PC, indicating that while it has relatively many and strong edges, these are not equally distributed throughout the network. Inspecting the estimated network in Figure 1, it can be seen that, indeed, the strongest edges of Tradition are mainly within its own community. Alternatively, Intelligence is not considered central based on the number and strength of its edges, but, taking the distribution of edges into account, we see that the connections of Intelligence are equally distributed to the other communities in the network (see Figure 16 in SM). This information would have been lost, if we had only focused on the number and strength of the edges (and other centrality measures related to these aspects).

In short, this example clearly illustrates that, when the objective is to find out which nodes play an important role in the network as connectors, it is important to consider whether there might be preexisting communities that should

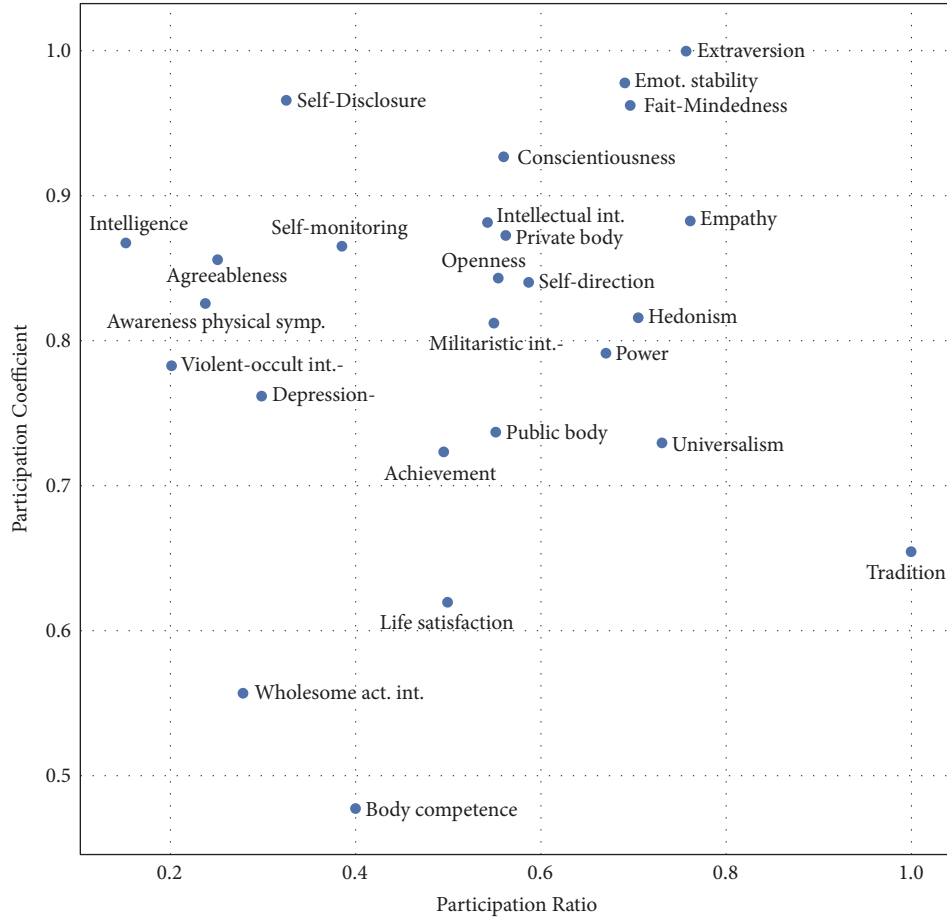


FIGURE 7: Scatterplot of standardized values of participation coefficient and participation ratio (min-max scale from 0 to 1) for all nodes in the network.

be taken into account. Not taking these preexisting communities into account might obscure the importance nodes belonging to small communities and “stand-alone” nodes that are not part of any community.

4.3. Analysis of Triadic Motifs. In this section, first we will explain the rationale behind the selection of motifs to be investigated, and the analysis of motif frequency, intensity, and coherence, followed by results and discussion, where the identification of specific motifs (and interpretation) is also included.

(i) *Selection of motifs:* Motifs usually represent subgraphs of three to five nodes for which different patterns of absent and present ties are examined. Many analyses of mesoscopic structures include or focus on triads, all possible configurations of three nodes. This is a sensible choice, because a triad is the smallest and the most basic network unit that defines the clustering of a network (transitivity) and can be characterized as the “simplest nontrivial motif” [53, p.2]. For undirected, unweighted, and unsigned networks, four types of triads exist: (1) triads without ties/edges (*empty triads*); (2) triads with one tie present, and two ties absent (*one edge triads*); (3) triads with one edge absent, and two edges present, referred

to in the literature as *two-path*, *two-star*, or *open triads* (or *forbidden triads* in weighted networks when present edges are strong); and (4) triads with all edges present (*triangles*, *closed triads*) (Triads should not be confused with triplets. Triplets are like triads, but they are defined only by the presence of the edges and do not by the absence of edges. For example, both triangles and open triads are triplets of two edges.). Usually, the first two types of triads are not considered in the analysis, and some researchers define triads more strictly as systems of three nodes with at least two ties among them (e.g., [54]). The number of possible triads increases when the sign and weights of the edges are considered (e.g., [55]), as will be done in our analysis. Depending on the research question, some motif configurations may be of special interest and should be investigated, while others can be excluded from the analysis.

(ii) *Analysis of motif occurrence, intensity, and coherence (including the identification of specific motifs):* Once the motifs of interest are defined, the next step is to determine the frequency of each motif in the empirical network (each unique combination of three nodes is counted once). This yields a first insight into the network patterns at the mesolevel. The most frequent motif describes the most dominant pattern of connectivity in the given network among the motifs that are examined. However, the frequency alone yields

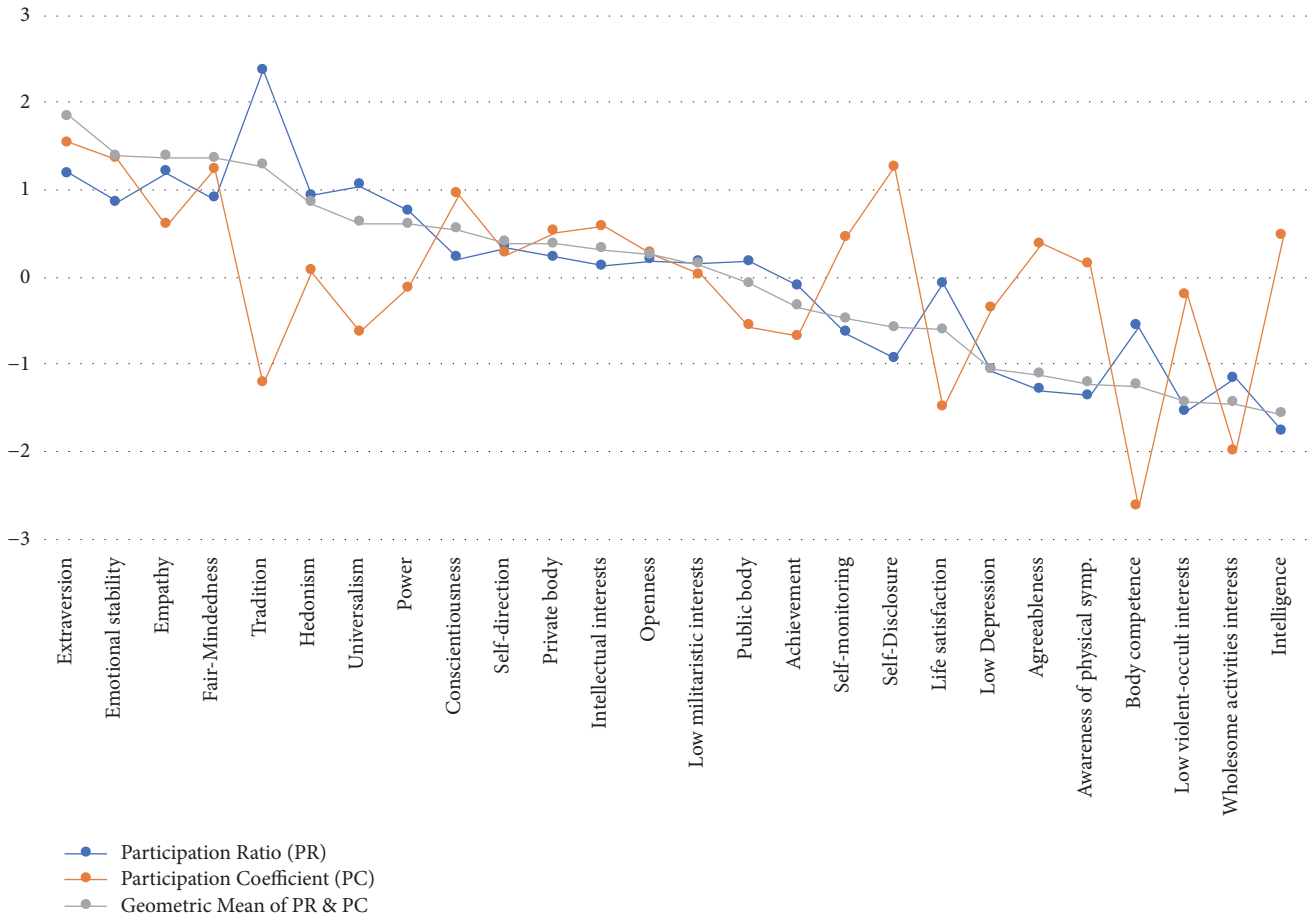


FIGURE 8: Centrality measures 3: participation ratio (the values of geometric means for Empathy and Extraversion are higher than both PR and PC. This is due to standardization of each measure. The plots with raw scores are shown in SM, Section 9, Figure 10.) ($\alpha = 0.5$), participation coefficient, and their geometric mean (standardized values).

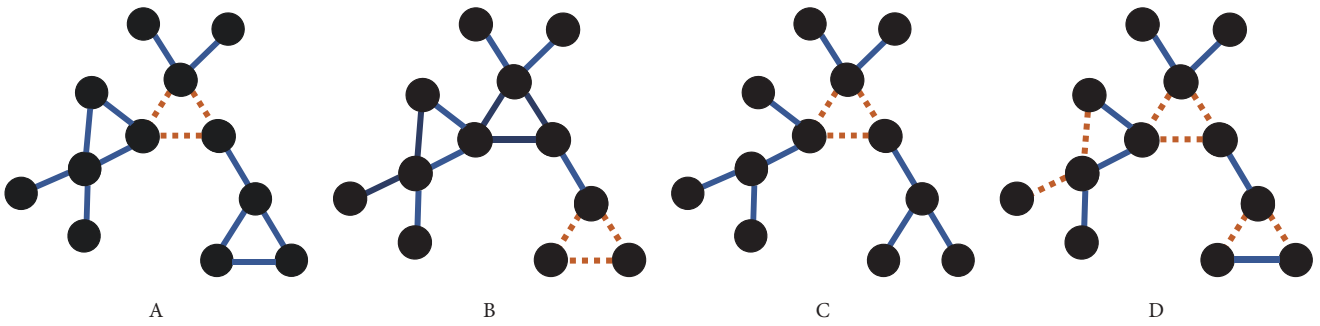


FIGURE 9: Four networks with 12 nodes and one negative triad (NNN). Networks A, B, and C have only three negative edges, while D has the same structure and density (number of edges) as A and B, but more negative edges.

limited information, because certain motifs might occur more frequently simply because of the network structure (in the context of describing the reference (null) model, the terms: network structure, topology, and degree sequence, are used interchangeably in this paper) and weight distribution. For example, imagine a hypothetical network of twelve nodes (variables) in which we observe predominantly positive edges, representing partial correlations between pairs

of variables, except for three negative edges (described in Figure 9).

If we find one negative triad in a network, based on frequency alone, we could treat that finding as somewhat interesting but not especially informative about the network as a whole. However, when we consider what the chances are of observing three nodes connected with three negative edges in that system, that finding is of greater importance

for understanding the whole network as a system. Figure 9 describes extreme (and unlikely) examples of psychological networks which are used to illustrate why it is useful to additionally look at the chance of certain motifs occurring in the system. The weight distributions of networks A and B in Figure 9 are the same, while network C has a different structure compared to A, B, and D, because just one closed triad (triangle) is present. Since the structure is different, the weight distribution of C is also different. The chance of a NNN occurring in a network with the same structure and weight distribution is smallest in C, followed by A and B, where it is equal. The highest chance of observing such a triad is in D, because it has more negative edges and triads than other three networks. If networks are representing symptoms (behavioral, emotional, cognitive, or physical) of a disorder, three negatively associated symptoms in A, B, and *especially* in C are more important characteristic of the system than in network D. They are less likely to occur by chance in these three networks, and therefore more likely to describe a process which is important for understanding the network. For example, a triad NNN in A could be interpreted as a process of negative feedback which is central for the network (it “drives” the network). In B, NNN is equally important but it describes occurrence of a negative “loop” in a peripheral part of the network, among symptoms that are less central. In C, NNN is even more essential for understanding the network than in A, as it could be described as the sole driving force of the network, each of the negatively connected nodes in the triad relates to a different set of nodes. Note that motif analysis per se does not differentiate between A and B as the centrality of configurations is not accounted for. Finally, NNN in D is a central configuration which shows an interesting pattern of association between three symptoms, worth of attention in the interpretation of the network. However, it is not as important for describing the process underlying the network formation since other negative associations between nodes and within triads are present. The same reasoning applies if nodes are representing other nonpathological tendencies, like personality traits, values, etc. In these networks the difference will be in the average weights of edges, which is likely to be smaller than in case of networks featuring psychopathological symptoms or other more correlated variables.

Therefore, for each motif, we establish whether it occurs more or less frequently than would be expected by a null model. In weighted networks, the appropriate null model is a random network (to be precise, it is not a random graph model, but a configuration model (for more details see [56])) with fixed topology (degree sequence) and randomized weights from the same distribution of weights as observed in the empirical network (for more details on general null models see [57]). The quantification of occurrence of a specific configuration in a network is usually done by comparing it with the occurrence of the same motif in a reference model (for introduction see [18]). Distribution of motif frequencies is obtained by generating a sample of random networks. The empirical frequency of a motif is compared against that distribution and if it appears significantly (this significance should not be confused with significance of ties in the motif) more (less) often than it would be expected by reference model

it signifies the motif is indeed “a motif.” (Sometimes the term “motif” is used only for these configurations for which this step of analysis shows that they are significantly over- or underrepresented. In this article, we do not make such distinction, as we refer to every investigated configuration as a motif, and after the analysis is done, we describe it as significant or not.) It describes an important characteristic of the investigated network. Motifs that occur more frequently describe a common configuration of nodes and therefore provide information about the network connectivity. Moreover, these motifs could have some important functional roles in the system. For example, closed triads are usually overrepresented in social networks, because they represent a process of social (triadic) closure, while in a network of intelligence measures they may indicate the process of mutualism [1].

However, in weighted networks the analysis of motif frequency omits the information about the weights (unless it is in some ways included in the definition of the motif). For example, if two motifs have the same occurrence in a network (let us assume for the sake of the argument that both have equal distribution of frequencies based on appropriate random models), but the first is (on average) made of stronger ties than the second, we cannot treat them as equally describing the local structure of the network, that is, to be equally likely to describe some important process in the network. Although they are equally present in the network, the first is expressed more strongly and is therefore more likely to describe some important process.

To address this issue, Onnela et al. [53] introduced the Intensity measure (the geometric mean of all the weights (in the case of absent ties in the motif, these are treated as zero weights) in a motif (5), where l_g stands for number of ties in the motif), which looks at the motifs not as discrete objects who are either present or not (expressed or not expressed) in the network, but rather as objects existing on a continuum, where zero or low Intensity values imply that motif is present in low degree. As such, the Intensity I can be used to identify high and low Intensity motifs in the system:

$$I_{(g)} = \left(\prod_{(i,j) \in l_g} w_{ij} \right)^{1/|l_g|} \quad (5)$$

In addition to Intensity (I), a Coherence ($Q_{(g)}$) ratio can be computed that quantifies how internally coherent the weights in motifs are by computing the ratio between the geometric and the arithmetic mean. It ranges from 0 to 1, with higher scores indicating less difference between the weights (in absolute terms). As was the case with the analysis of occurrence of motifs, the significance of both Intensity and Coherence is estimated in comparison with the distribution of their values for a given motif in reference model.

A motif that is underrepresented in the network, in terms of occurrence or intensity, describes a pattern of relationships which, for some reason, is unlikely to happen in a network. In other words, when we exclude the hypothesis that a given occurrence or intensity of a certain configuration does not come from a reference system, it points out that there may

be an additional origin for the effect, possibly the function of the system [18]. In case of psychological networks, the occurrence and significance of a motif which is not easily interpretable may also happen as an artefact (e.g., due to the sample on which the network is estimated, problems in the network estimation procedure, or measurement error). For that reason, a motif analysis can be useful in the analysis of psychological networks, forasmuch as it can help quantify and identify presence of unexpected configurations in the network as well.

In the next section, the motif analysis on illustrative data is described in detail and results are presented and discussed.

4.3.1. Selection of Motifs and Analysis of Motif Occurrence.

When the sign of an edge is considered, seven configurations of triads are possible (disregarding empty triads and triads with only one edge, see Figure 9). Four of them fall under “closed” triads or triangles: triads with either only positive (positive triad, PPP) or only negative weights (negative triad, NNN) and triads consisting of two positive and one negative weight (PPN) or two negative and one positive weight (NNP). NNN and PPN are also known as imbalanced triads (NNN is also sometimes considered as imbalanced triad in social networks, but some debate exists over whether it is truly imbalanced or not. Not to confuse with too many similarly named triad, we will use the term “imbalanced” in this article only when referring to triad with one positive and two negative ties and to triads that do not satisfy the triangle inequality principle (the latter is explained in the following text)), in social balance theory [58, 59] because they signify configurations of affective ties between persons which is not likely to appear in social networks (or if it appears it is not likely to persist; that is, it is likely to change). The remaining three triads are open triads (2paths) consisting of two ties: with only positive weights (2path pos., POP where “0” stands for the absent weight), only negative weights (2path neg., NON), or with one positive and one negative weight (2path mixed, PON or NOP).

Networks, especially social networks, tend to show transitivity; if person A is connected with (friend of) person B, who is connected with (friend of) person C, A and C are likely to be connected (friends). Although, in recent years, we have witnessed a surge of research on psychological networks, we still do not know enough about their general properties. Correlations, and especially partial correlations, do not have to be transitive, but it is often the case that if a trait A positively correlates with trait B, which is also correlated positively with trait C, then we expect traits A and C to correlate positively as well. If that is the case, POP motifs should appear less often than expected by the reference model. Likewise, according to the social balance theory, closed triads with one or three negative edges (i.e., PPN and NNN) are less likely to occur in social networks [58–60]. We hypothesize that, in psychological networks too, NNN and PPN triads represent configurations which are not expected to occur frequently because of two reasons. First, it is challenging to explain how three psychological attributes feature negatively partial correlations. One possibility is that a process of negative

feedback among attributes exists. A second possibility is that the three nodes positively contribute to a common effect, which has been implicitly or explicitly conditioned on. A third possibility is that the variables are measured with error, and the partial correlation picks up negative correlations between the error terms.

On the other hand, positive associations between A and B, and B and C, render a possible negative association between A and C difficult to interpret (PPN triad). The importance of detecting such configurations in psychological networks lies in the fact that they either describe unusual finding(s) or may point to the existence of methodological artefacts. In both cases, we benefit from knowing about the presence of such configurations. It should be noted that, while it is more straightforward to predict that such configurations could be less frequent in a correlation network, in the case of partial correlation network they could be more likely to occur. To the best of our knowledge no analysis of this kind has been performed on a network representing (partial) correlations. The summary of hypotheses is shown in Table 4, in Section 5.

Among the motifs (Figure 10, third row), the only significant motif is the negative triad (percentile 99.7). In other words, the negative triad appears more frequently than would be expected by chance, given the same degree sequence and weight distribution. Path2 with positive ties (POP), indicating high presence of nodes which are bridges, is overrepresented, and the imbalanced triad (PPN) is underrepresented, but neither reaches the level of significance.

To identify only the strongest motifs, we looked at signed motifs with an added threshold (see Figure 11). To end up with a similar number of examples for each motif, we selected a threshold of 0.15 (around 75 percentile of edge weights, see Table 1) for closed triads and a threshold of 0.20 for 2path motifs. Among the motifs that meet this threshold, one specific motif may be of relevance for psychological networks. This is the last motif in Figure 11, which we called imbalanced triplets II T., based on the work of Toivonen et al. [61] (hence the T. in the name, for definition see Figure 11). Toivonen and colleagues investigated a correlation network of emotion concepts and argued that this motif describes patterns that cannot be depicted in any dimensional space without being distorted. This “imbalanced triplet” describes a pattern which is contrainuitive, although not necessary unreal, and it is similar in logic to NNP triad. If A, B, and C represent three psychological dimensions (e.g., emotions and traits), and positive correlations between A and B, and B and C exist, depending on the strength of r_{AB} and r_{BC} , A and C ought to correlate at least as the half of either of the two (r_{AB} or r_{BC}) which is the weaker correlation. Otherwise the ABC triad does not satisfy the triangle inequality principle; that is, it cannot be described by dimensional techniques (in Euclidean space), while a network representation can be used for detecting their presence.

As mentioned for the NNN and PPN motifs, while we can expect low occurrence of imbalanced triplets II T. in a correlation network, in a partial correlation network this is quite different. An imbalanced triad in a partial correlation network implies that the partial correlation between A and B is small given C, which means that A and B approach

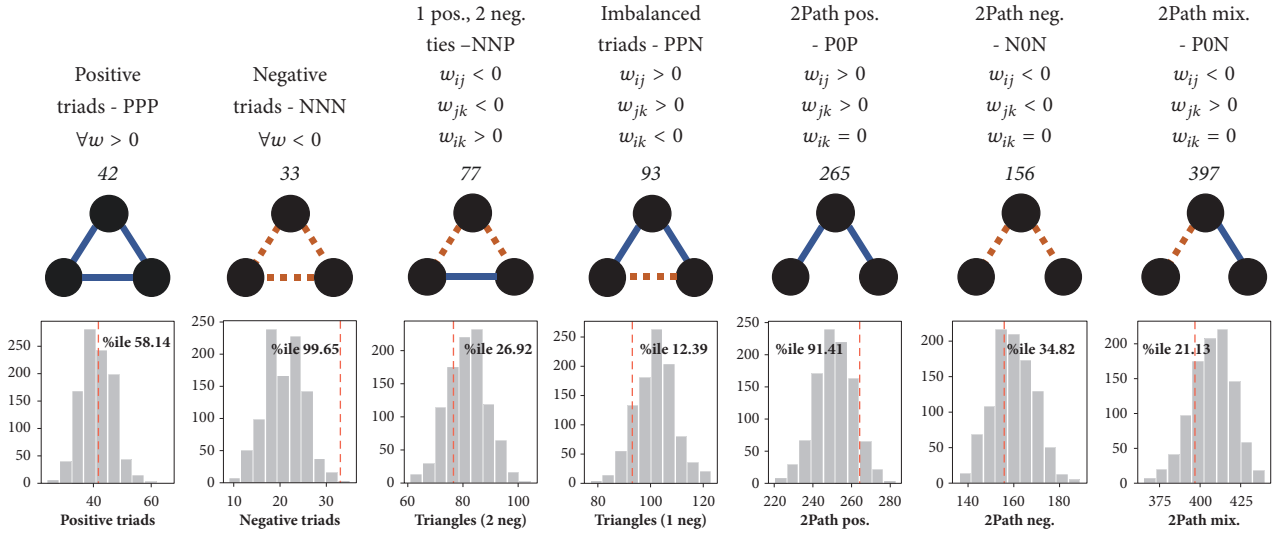


FIGURE 10: Signed motifs, name used in this study, the definition, schematic figure, and the figure showing distribution of motif frequencies in 1000 random networks with the same degree sequence and weight distribution and percentile value of the frequency of empirical network in that distribution.

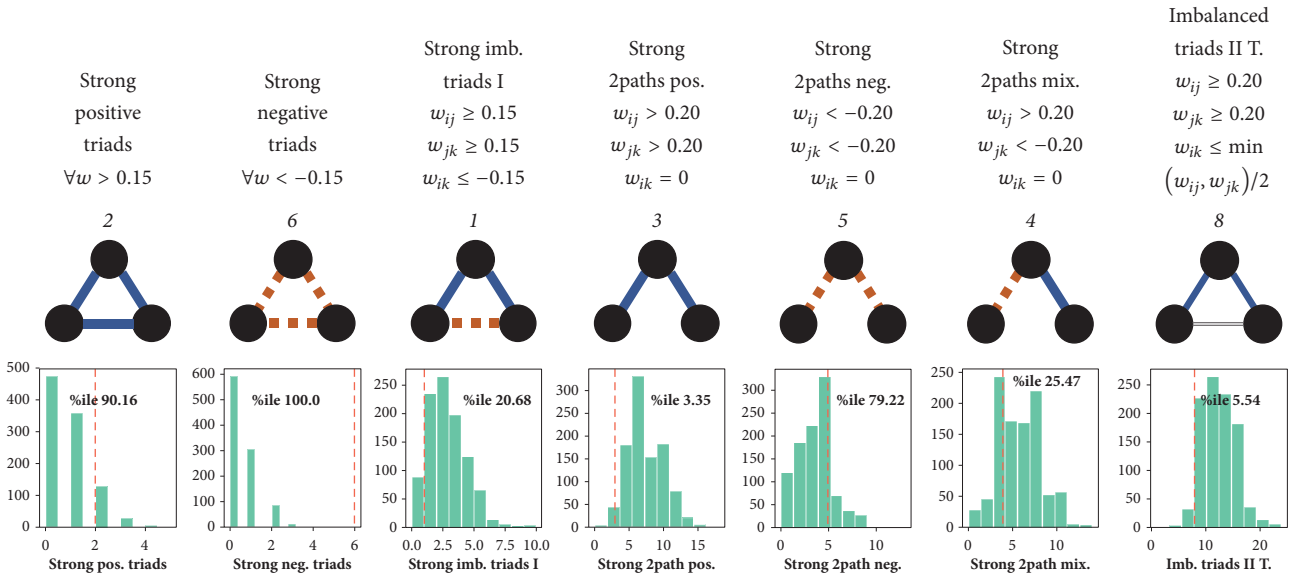


FIGURE 11: Weighted and signed motifs, name used in this study, the definition, schematic figure, and the figure showing distribution of motif frequencies in 1000 random networks with the same degree sequence and weight distribution and percentile value of the frequency of empirical network in that distribution.









conditional independence given C. This in turn is consistent with a chain (A->B->C or A<-B<-C) or a fork (A<-B->C). Both of these may yield indirect, but important clues to the causal structure within the triad. Those triads are good candidates for more focused analytical approaches that allow for causal inference (e.g., mediation or path analysis). Thus, regardless of frequency, the imbalanced triplets II T. represent a configuration that describes possibly interesting phenomena which would go unnoticed with dimensional methods [61].

Results show that, even when we “focus” just on motifs of relatively strong ties (Figure 11 (third row), all of them

identified in Table 3) again only the NNN triad occurs significantly and more than expected by chance. The cardinality (a term used in network analysis to address the significance of a motif) of the motifs in this network is thus not dependent on the strength of the weights. However, the strong imbalanced triads, 2paths with positive weights, and imbalanced triplets II T. have the tendency to be underrepresented. This pattern is expected in social networks, where imbalanced triads and “forbidden triads” (2paths) are generally less expressed, and this network shows similar tendencies.

All motifs defined in Figure 11 are identified and described in more detail in Table 3.

TABLE 3: Weighted and signed motifs identified.

Motif	 $i, j, k; (pr_{ij}, pr_{jk}, pr_{ik}), [r_{ij}, r_{jk}, r_{ik}]$
	Private body, Public body, Body competence; (.30, .40, .25), [.52, .58, .49] Conscientiousness, Fair-Mindedness, Self-Disclosure; (.16, .21, .16), [.34, .37, .37]
	Tradition, Universalism, Power; (-.34, -.29, -.16), [-.20, -.46, -.11] Tradition, Universalism, Hedonism; (-.34, -.21, -.36), [-.20, -.12, -.38] Tradition, Universalism, Achievement; (-.34, -.30, -.28), [-.20, -.34, -.24] Tradition, Self-direction, Power; (-.37, -.20, -.16), [-.47, -.12, -.11] Tradition, Hedonism, Achievement; (-.36, -.17, -.28), [-.38, .01, -.24] Universalism, Hedonism, Achievement; (-.21, -.17, -.30), [-.38, -.01, -.24]
	Militaristic int.-, Universalism, Wholesome act. int.; (.22, .16, -.39), [.20, .19, -.37]
	Agreeableness, Empathy, Extraversion; (.27, .32, .0), [.45, .39, .16] Life satisfaction, Emotional stability, Agreeableness; (.25, .26, .0), [.48, .35, .25] Wholesome act. int., Intellectual int., Openness; (.31, .21, .0), [.41, .44, .17]
	Self-direction, Tradition, Universalism; (-.37, -.34, .0), [-.47, -.20, .21] Hedonism, Tradition, Self-direction; (-.36, -.37, .0), [-.38, -.47, .16] Achievement, Tradition, Self-direction; (-.28, -.37, .0), [-.24, -.47, .09] Self-direction, Power, Universalism; (-.20, -.29, .0), [-.12, -.46, .21] Hedonism, Universalism, Power; (-.21, -.29, .0), [-.12, -.46, .21]
	Militaristic int.-, Universalism, Tradition; (.22, -.34, .0), [.20, -.20, -.10] Intellectual int., Wholesome act. int., Militaristic int.-; (.31, -.39, .0), [.41, -.37, -.11] Militaristic int.-, Universalism, Power; (.22, -.29, .0), [.20, -.46, -.10] Militaristic int.-, Universalism, Achievement; (.22, -.30, .0), [.20, -.34, -.10]
	Self-monitoring, Extraversion, Empathy; (.30, .32, .09), [.31, .39, .12] Depression-, Emot. stability, Agreeableness; (.38, .26, -.06), [.55, .35, .17] Emot. stability, Agreeableness, Empathy; (.26, .27, -.06), [.35, .45, .15] Violent-occult int.-, Militaristic int.-, Universalism; (.53, .22, -.11), [.45, .20, -.07] Life satisfaction, Emot. stability, Depression-; (.25, .38, .11), [.48, .55, .40] Wholesome act. int., Intellectual int., Openness; (.31, .21, .0), [.41, .44, .17]* Agreeableness, Empathy, Extraversion; (.27, .32, .0), [.45, .39, .16]* Life satisfaction, Emot. stability, Agreeableness; (.25, .26, .0), [.48, .35, .25]*

* Identified also as a 2path pos. motif due to overlap in the motif definition with Imb. triad II T.

(-) after the name of a psychological attribute means that it has been reversed.

Strong PPP triads may indicate the presence of a common cause, for instance, because the three variables measure the same underlying psychological construct, which then acts as a latent variable. Unsurprisingly, the relationships among the three constructs measured by the Body Consciousness questionnaire represent one such case. Another such motif is made of Conscientiousness and two integrity measures, Fair-Mindedness and Self-Disclosure, pointing out that they are likely capturing similar psychological dimension. A second possibility that may underlay PPP triads is a positive feedback between the variables, as found in the mutualism model for intelligence.

All six NNN triads involve Schwartz's values, with Tradition being present in five of them. This configuration cannot emerge from a common cause and may suggest a

negative feedback loop between the attributes. Still, such an interpretation is formed on conclusions about intraindividual differences that are based on interindividual data, which may not necessarily hold. A second possible reason for observing NNN triads is that the variables have been conditioned on a common effect to which each of them positively contributes. The logic here is the following. Suppose that three variables A, B, and C increase the probability of common effect D. If we condition on D, we only consider the values of A, B, and C for a given value of D. Suppose we observe that the effect is present (or D has a high value), but A is not present (or has a low value). Then that information makes it more likely that B or C are present (or have a high value). Thus, conditioning on D, we expect A, B, and C to be negatively related so that they form an NNN triangle in the partial correlation network.

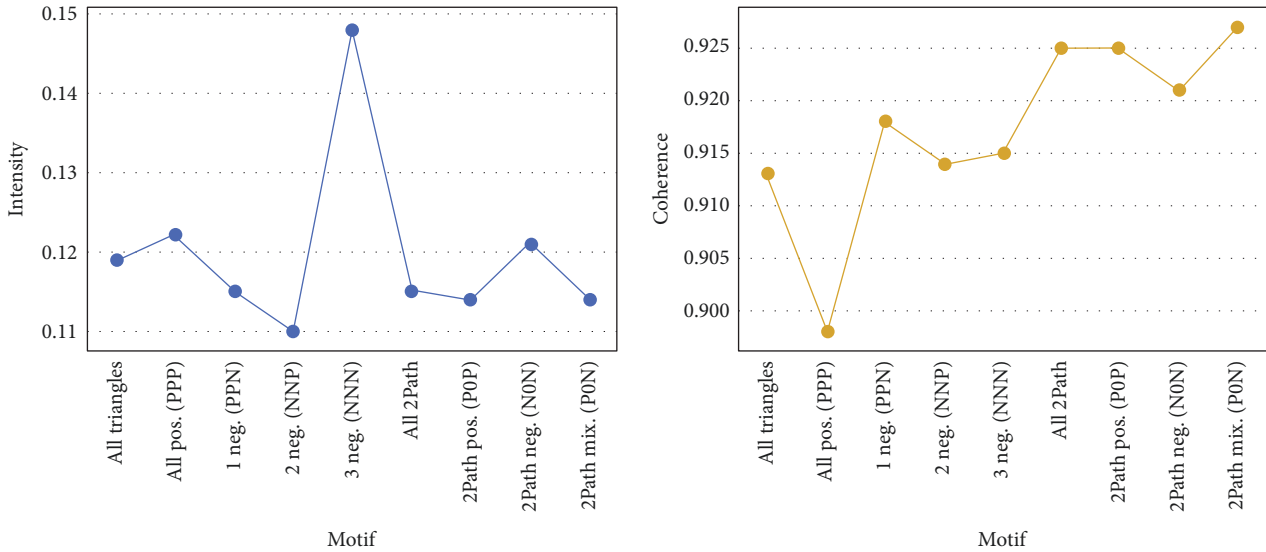


FIGURE 12: Means of intensity and coherence of all triads and signed motifs in the network.

One NNP triad consists of a negative association between Low Militaristic values and Interests in wholesome activities, while both variables are positively correlated with Universalism. This triad identifies a puzzling relationship that might suggest multidimensionality of the Universalism value. Positive 2paths show that Empathy, Emotional Stability, and Intellectual Interests may play the role of mediators. Negative and mixed 2paths similarly show the variable in central position (position “J” in Table 3) as bridging the remaining two attributes in the subgraph. Finally, eight configurations present the strongest imbalanced triplets II T. in the network, which are not possible to describe in the metric space. Three of them also fall under 2paths, due to the overlap in the motif definition. The variable in position “J” (see Table 3, first row) in this motif is likely to be a broad concept with multiple meanings.

4.3.2. Analysis of Motif Intensity. In previous research, the Intensity measure has been applied for triadic motifs consisting of positive weights only. Therefore, we modified the approach described by Onnela et al. [53] by calculating I and Q separately for triads with a different configuration of positive and negative ties to allow comparing the Intensities across different motifs. The average Intensity and Coherence for all investigated motifs are shown in Figure 12.

Visual inspection of Figure 12 reveals that the differences in Intensity and Coherence between the motifs are very small (y axes show range of 0.05 for I , and 0.025 for Q). When looking at the structural motifs concerned only about presence and absence of ties, and not their weights, all triads have a higher Intensity than 2paths, but the difference is very small. In psychological networks, it would be expected that triangles have a higher Intensity than 2paths, as triangles represent mutual connections between all three nodes, making it more likely that the nodes will

reinforce each other. Because of this reinforcement, it would be expected that the weights are of higher absolute value than in 2paths, where one edge is missing, making such effect less plausible.

The most intensive motif, that is, the motif with the highest average geometric mean of weights, is a triad made of three negative ties NNN, followed by positive triad (PPP) and 2path with two negative ties (N0N). The finding that a NNN motif is the most intensive is somewhat surprising for networks of this kind, but, before attempting interpretation, we will proceed first with analysis of Coherence, followed by significance testing.

Internal Coherence of 2paths (open triads) is somewhat higher than for closed triads (Figure 12, right panel), which is to be expected as 2paths consist of one weight less than triads. PPN seems to have relatively higher, while PPP relatively lower Q .

Having a high (low) average Intensity of a motif does not imply that the motif is highly (lowly) expressed in the network. Therefore, the next step is to check how significant the Intensities are. The same applies to the Q , where a high Q of a motif does not imply it is significantly more coherent. To answer those questions, the Intensities and Coherences of each motif are compared with the mean of I and Q of each motif in an ensemble of 1000 random networks. The results of the analysis are shown in Figure 13.

The only motif whose Intensity (percentile value > 97.5) is significantly high is a triad with three negative ties (NNN), which is in line with the results on the frequency and the descriptive analysis presented in Figure 12. Although the average Intensity is not high in absolute terms (slightly above 0.14), the frequency and Intensity analysis both suggest that the NNN motif is an important characteristic of the network. In Table 3, we saw that all NNNs involve only Schwartz’s values. NNN motifs show a tendency to be “nested” around few nodes; only the nodes that represent Schwartz’s values are “responsible” for the high frequency (and Intensity) of that

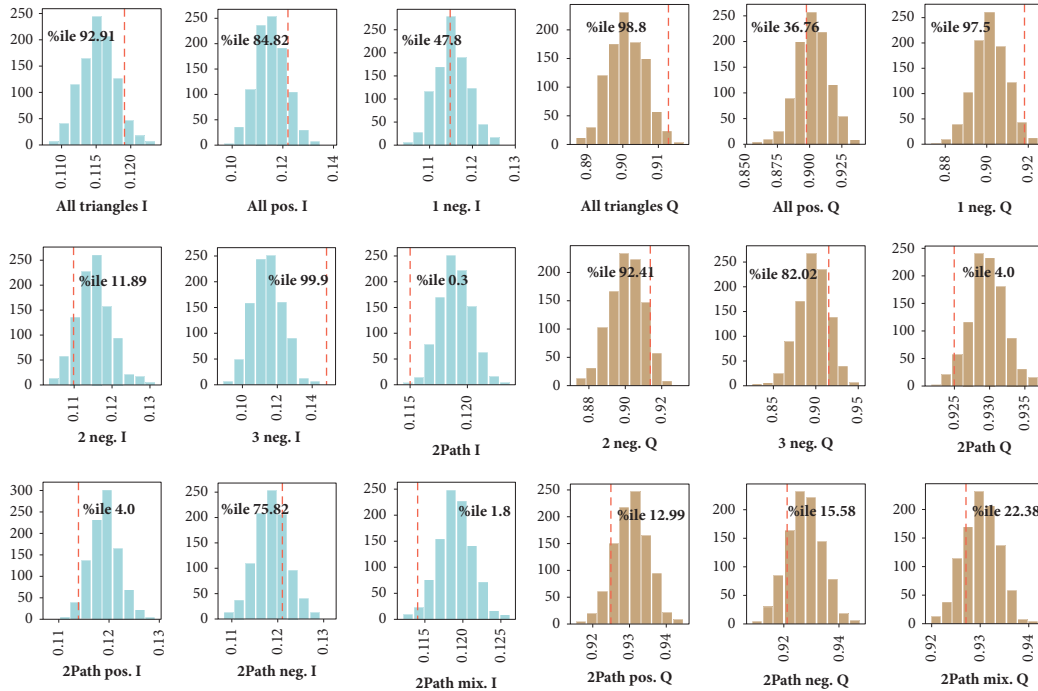


FIGURE 13: Significance of motifs' intensity and coherence: distributions of intensity (first three columns, colored blue) and coherence (last three columns, colored light brown) of all closed and open triads, 2paths, and signed motifs in 1000 random networks with same structure and weight distribution as empirical network, with its percentile values.

motif on a network level. Furthermore, from Figure 1 (and the centrality analyses) we observed that not all Schwartz's values are central. From that we may generate a hypothesis that the most prominent characteristic of the psychological system of 26 attributes is described by a negative feedback between values, although the cluster with such pattern is not central in the system. A second possibility is that some of the values are involved in a common effect with respect to one of them, which might for instance arise when, say, Tradition is caused by all other variables. Due to the conditioning on the common effect, the NNN pattern may arise for the causal variables in the partial correlation network. A final possibility would be that the high occurrence of NNN may be the result of estimating network on a sample which is self-selected (i.e., implicitly conditioned) on a variable that is a common effect of Schwartz's values.

Two motifs with significantly small Intensity (percentile value < 2.5) are all 2paths motif (structural, disregarding the signs of ties), and 2paths with one negative and one positive tie (with mixed ties, PON). The later finding is an example of the importance of comparison with the reference system. When we analyzed only the average Intensity, we have found that 2paths have a higher Intensity than other motifs. Comparing this to what may be expected given the network structure and weight distribution, we can see that, in fact, the Intensity of 2paths, although somewhat higher in absolute value than Intensity of other motifs, is significantly smaller than it would be expected by the null model. The "intuitive" expectation about smaller Intensity of 2paths due to the lack of third link is supported.

Closed triads (all triangles) display significantly high internal Coherence. From the tie's perspective, this may suggest that weights of similar strengths show the tendency to form triads, or, from a node perspective, that psychological attributes that form a triad tend to be connected with ties of similar strengths (in absolute values). Imbalanced triad (PPN, called "1 neg." in Figure 13) is also significantly more coherent, meaning that the weights within this triangle tend to be equally distributed (they do not show big variations). Interestingly, so-called imbalanced triads in this network consist of "balanced" edge weights. The overall pattern of results show that a significant I does not imply significance in Q , which highlights that they measure two different aspects of this system.

5. General Discussion and Conclusions

This paper has demonstrated how the use of three metrics taken from network science can enrich our understanding about psychological networks. Given the effort invested in estimating the network structure, it is a missed opportunity not to use the information it entails more fully to gain deeper understanding of estimated network. This "omission" may be understood and partly explained by researchers in the field being preoccupied primarily with network estimation methods [11, 47, 62] and replicability issues [49, 63, 64] that arise from the fact that network structures between variables are considerably more difficult to determine, relative to, for example, internet links or electricity nets; after all, conditional association between variables is not observable,

but must be estimated from data. Appropriately dealing with sampling error in estimating network structures, as well as assessing their robustness, has therefore been the priority in psychological network analysis.

The concise overview of the three methods in terms of hypothesis and research questions and procedure is given in Table 4.

We demonstrated on illustrative dataset how each of the methods proposed here adds new information about the network structure. First, the MST can help us in shedding light on the topological arrangement of psychological attributes in the network. Specifically, in the current example, the MST suggests that Empathy is the most similar to all other traits and plays the role of a “network connector;” it is the most central trait when centrality is based on the network filtered down to its most essential ties. In the network which also includes Big Five traits, it was somewhat surprising to see that Empathy has such important standing. This could be due to the questionnaire used for this trait (the Empathy Quotient) which captures affective and cognitive aspects (see Table 1). The authors [39] of the questionnaire state that the cognitive component of Empathy is closely related with an individual’s “Theory of Mind,” a cognitive process that allows people to understand others and oneself. It might, thus, be plausible that cognitive processes related to Theory of Mind serve as a central hub in the system. In addition, it is tempting to see the analogy and state that the trait which is seen by some to hold society together may also hold this network of different psychological attributes together. This finding is worth of further attention due to an implicit and misguided notion that Big Five traits are the best representative of psychological differences between individuals. If true, in network terms that would imply that they are expected to be in the top five most central nodes, which is the case only for some of them. In fact, Openness is among peripheral nodes. Nonetheless, further theoretical consideration and research is needed. The MST provided an additional insight into possible clusters of attributes and showed that clusters, that is, different branches of the tree, for the most part do not align with different kinds of psychological variables. For example, Big five traits and Schwartz’s values are placed on different branches, suggesting that the grouping of variables is based on specific content rather than “nature” of a psychological variable (e.g., whether it is a trait, value, or interest). Furthermore, we used the fact that MST preserves the information of edge signs to employ it for robustness test of network estimation.

Second, by including information about the participation coefficient based on predefined communities, which also included “communities of one,” we highlighted the specific role of some nodes based on their equal importance to the structure of different parts of the network. We found that Intelligence, although weakly connected to other traits, and by all centrality measures quite peripheral, does seem to have an interesting property of being relatively equally associated with all different kinds of nodes in the network. Based on this finding we can hypothesize that cognitive ability relates to personality: not in terms of substantial effect sizes but because it relates at a constant strength to most “parts” of psychological system. In other words, the question

about relation between cognitive ability and psychological individual differences could be better answered if instead of looking at the “size” of that influence (operationalized with some statistical measure), researchers refocus their attention on the “broadness” of that influence. This agrees with the suggestion of Salovey and Mayer [25] that, instead of looking at pairwise correlations, a more complex analysis that looks at many connections at once should be preferred. Likewise, network ties of Intelligence seem to imply a different relation with Big Five model than reported in the recent review [65]. When 24 other relevant individual differences (26 minus 2 variables whose connection is under consideration) are controlled for, the strongest tie is not with Openness, but with Agreeableness and Extraversion (both negative and around 0.10).

We used PC together with the Participation Ratio to arrive at more sensible centrality measure, which showed that different centrality indices converge to Extraversion, Emotional Stability, and Empathy as the three most central nodes in this network. Centrality of Extraversion and Emotional Stability would be expected since they are one of the traits that have been recognized as important psychological dimensions and systematically studied from early on in psychological science. Empathy taking the “third place” is somewhat surprising, but, as discussed before, could be related with this trait capturing cognitive processes that are essential and fundamental in many social interactions [66].

Finally, we used motif analysis to research possibly interesting three-node configurations and investigate whether this psychological network “behaves” as a social network regarding its balance of negative and positive ties within a triad, and the results showed this is not the case. We learned that some configurations that are challenging to interpret exist in the network at a higher frequency than would be expected in the reference system; most notably, this was the case for NNN triads. Identification of strong motifs revealed that these triads originate mostly from one group of nodes, Schwartz’s values, possibly revealing negative feedback or (implicit) conditioning on a common effect of some or all of the variables. NNN triads are also significantly stronger than expected, but otherwise intensity and coherence do not seem to be related with frequency of motifs.

Methodological Considerations Related with the Reverse Coding of Variables. An important issue related with network modeling of relationships between continuous variables which probably did not receive enough attention so far is the effect of reverse coding of variables on the results of network methods (we are grateful to a reviewer for pointing out this issue). It becomes an even more pressing issue when nodes are aggregations of more complex concepts, not easily described as positive or negative (e.g., some values), or when variables present dimensions which are interpretable on both ends (e.g., emotional stability–neuroticism, extraversion–introversion), and often coded arbitrary. This is the case for many continuous variables in psychology, and probably for all variables in our dataset to some extent. For example, Emotional Stability (ES) is often coded negatively as Neuroticism (N), begging the question what would happen

TABLE 4: An overview of the three methods.

MINIMUM SPANNING TREE	
Recommended for network	Dense networks and/or networks with many small edges
Important procedure steps	(1) Selecting a distance measure: (i) Gower's distance (ii) Distance inversely proportional to the shared variance
Output	(2) Centrality analysis (by inspection or/and with standard centrality measures computed)
Methodological considerations	MST – the filtered network
Other analytical possibilities	Distance measures (i) and (ii) will produce different MSTs if a network has negative ties* Looking at MST branches as communities
Effect of reverse-coding variables	Using MST to test the robustness of the network estimation of the most essential edges MST can include distance metric as weight for further analysis
Hypothesis/Research question	(i) affected* (ii) not affected RQ: Which node is the most central? Which (overlapping) communities exist in the network?
PARTICIPATION COEFFICIENT as a corrective	
Recommended for network	(1) Pre-existing differences in the kinds of nodes (2) Networks with communities
Important procedure steps	If (1) is true: (a) Defining the node groups (b) Calculating PC (c) Choosing the centrality measure to be corrected with PC (optional) If (2) is true: (a) Data-driven detection of communities (b) Calculating PC (c) Choosing the centrality measure to be corrected with PC (d) Comparing the rank order of chosen centrality measure before and after the correction
Output	PC values for each node (in (1)(b) and (2)(b)); The corrected centrality measure (for (1)(c) and (2)(c)) Communities should not overlap ((1) and (2))
Methodological considerations	(1) a Group affiliations may be ambiguous (2) a Decision about appropriate community detection algorithm
Other analytical possibilities	PC version that treats positive and negative edges separately
Effect of reverse-coding variables	Not affected if signs are not taken in the account when calculating PC
Hypothesis/Research question	$H_0 = A$ centrality measure is not affected by (pre-existing or data-driven) communities

TABLE 4: Continued.

	MOTIF ANALYSIS
Recommended for network	(1) Not for networks with small number of nodes and/or very low density (2) Has negative and positive ties (3) If weighted, additional steps in the procedure (a) Defining motifs of interest and the null model (b) Motif identification and frequency (c) Significance testing of motif frequency
Important procedure steps	If (3) is true: (d) Motif intensity (and coherence) (e) Significance testing of motif intensity (and coherence)
Output	Identified motifs; Motif Frequency; P-values for frequencies; Motif intensity (and coherence); P-values for the intensities (and coherence)
Methodological considerations	The definition of the null (reference) model
Other analytical possibilities	Other motif structures, e.g. that include more nodes
Effect of reverse-coding variables	Identified motifs and motif frequency will be different*, but conclusions about significance (of frequency, intensity, and coherence) will tend to converge
Hypothesis/Research question	Many research questions and hypotheses possible. In this study: <i>Signed edges will tend to cluster in the line with what is observed in social networks and correlational networks (balance theory, forbidden triads, imbalanced triplets);</i> $H_1 = \text{POP, NNN}$, and PPN will be less frequent than expected by chance. RQ1 = Do same pattern of results holds true when only relatively stronger motifs are considered? RQ2 = Are Intensity and Coherence measure following the same pattern of results as the frequency of motifs?

*This should be the case, but there is a possibility that distances related to negative edges are present in a network in such a way that will not affect the MST construction, e.g. a weak negative tie that exists among two peripheral nodes that have ties to other more central node.

with the results of analyses if we used N instead of ES? To find out we repeated most of the analyses reported in this paper with the network that had N instead of ES, and several other networks with some of the variables recoded. The results are presented in detail in SM (Section 12), while here we will highlight just the most important conclusions. The estimated network will have the same structure and absolute values of weights, but all the edges of reversed node will change their sign. Weight distribution of network is affected too, due to the changes in signs of some of the weights. The most affected are the results of MST, but only if the preferred distance measure is used. Otherwise, with the measure inversely proportional to shared variance, MST results are unaffected. This situation brings up the dilemma of which distance metric to use: the more rigorous one that is affected by variable coding, or the one which leads to a possibly substantial loss of information but is immune to reverse coding? We do not provide an answer, because, as usual, it will depend on the specific network, variables included, and the research question. Nevertheless, researchers need to be aware of this issue. In contrast with MST, PC that takes only absolute value of weights is not affected by reverse coding. Motif analysis will produce different motif frequencies, intensity, and coherence values, but the results of significance testing will not be affected to a greater extent and will tend to converge for the same network with differently coding some of the variables.

A logical conclusion following from previous section is that the three methods discussed in this paper require an effort to be applied to a psychological network, as some additional decisions need to be reached such that they are in accordance with research questions/goals (also explained in Sections 4.1, 4.2, and 4.3). Each decision has its repercussions. In case of MST, one needs to consider the presence of negative ties and what is achieved by deciding to look at two negatively associated nodes as more dissimilar than two nodes that are not connected at all. For PC, the nature of nodes included in the network needs to be carefully looked at, while for motif analysis some notion about which specific configurations may reveal interesting patterns in the network should be formed. The common ground of all three methods is that they look at direct, local ties, but in the contrast to the degree centrality they provide more fine-grained information. This presents a potential for a deeper understanding of any network but is also a very convenient feature for networks that do not have well-defined boundaries. By boundaries, we refer to two issues. The first issue is the possibility that some node(s) which are part of the system are not included in the network analysis. This is an issue for our network where selection of variables was atheoretical, since a “global” theory that describes all psychological attributes does not exist. The selection was further constrained by data availability. For example, we can think of some potentially important attributes that are not in the network, for example, self-efficacy, need for cognition, and narcissism. While acknowledging this, the limitation had its advantage in indirectly preselecting some of the currently most studied/used (and therefore, it could be argued, important) concepts. The second issue is related to the first one and refers to the nature of the investigated network.

Some networks are more easily influenced by “externalities;” for example, for a psychological network this may include some important life events that can bring about the change in the network by directly or indirectly influencing one or more nodes. Hence, global properties of such network, and measures relying on all ties in it, may be less useful. The fact that whole system is not represented and that it is an “open” system, as is the case in probably many psychological networks studied so far, was the motivation for introducing these three network methods that rely more heavily on local than global network structure.

To conclude, the added value of more information provided by more complex network tools comes at the price of less straightforward procedures, and making more decisions (hopefully informed by theory and previous research). However, we believe that those elements are just more salient when using these three methods, than when using typical centrality analysis based on different centrality indices, where many assumptions are implicit (e.g., that all nodes are equally likely to be connected to any other node). Therefore, we look at this requirement for higher deliberation as a good practice in general when applying any network analysis to psychometric data, as it challenges researchers to think more about nature of nodes, ties, and smaller network configurations in the network. Nevertheless, that is not an easy task. Understanding these “new” methods may be at first somewhat less straightforward and difficult for researchers not heavily involved in network analysis. This is especially true for motif analysis, which is by far the most complex of the three. Given that network approach is relatively new in psychology, it will take some time for network ideas and methods to “sink in.” Unfortunately, it also lacks strong theories. Be that as it may, better understanding of its analytical tools and exploratory (and that sometimes means undertheorized) potential will greatly facilitate the development of such theories. William James’s argument that “a degree of vagueness can be beneficial to science when attempting new research directions” [67, p.2] nicely captures the point we are trying to make. This holds true not only for network theories, but also for any kind of theories which aim to integrate many small (“local”) theories in psychology.

The methodology presented offers interesting possibilities for applications to other areas. For example, it would be informative to see how equally distributed ties are of depression symptoms among different groups of symptoms (e.g., thoughts, physical symptoms, behaviors, and feelings), and which symptoms are most central when that information is taken into account. We are not suggesting that all methods should be used in every analysis. The most appropriate methods and its specific procedure should be established based on careful consideration of the data at hand, research questions and theory behind it, and knowledge of existing network science tools. Our goal was to expand the latter.

Network approach is often compared to other multivariate methods more commonly used in the field of psychology, for example, structural equation modelling (SEM), confirmatory factor analysis (CFA), mediation analysis (MA), hierarchical clustering (HC), and multidimensional scaling (MDS). Although detailed comparison is out of scope of this

paper, we will proceed with a general overview with highlight on three most notable differences between network approach and most of multivariate methods used in psychology that are more closely related with three specific methods we introduced in this paper. Firstly, the network approach is less directly guided by researcher's assumptions about the connections between variables than most other methods (e.g., CFA), that is, except for the decision about the variables that will be included in the network. In reality, the decision about which variables will be included in the network is constrained with data availability. In this regard, using PC can help in indirectly controlling for some aspects of that constraint, acting as a corrective measure for possible bias in the selection of nodes that have been included in the network.

Secondly, in comparison with SEM, and MA, network analysis usually deals with a greater number of variables at once, implying that SEM and MA may be more appropriate for smaller set of variables, especially if clear theoretical expectations exist about relationships between the constructs.

Finally, other approaches are not trying to look at the set of investigated variables as a system and reveal the properties of that system; they rarely go beyond the microlevel of examining specific connections. In that sense, MST and motif analysis are valuable tools within network approach. MST can be used, among other reasons (mentioned in this paper), to filter the most important connections in the system and to provide answer about the most central variables/nodes on a more general level than specific centrality measures. One part of the output of motif analysis, the identification of motifs, can be viewed as a counterpart to MA (or SEM if configurations tested with motif analysis include more than three variables/nodes) among network methods. However, other outputs of motif analysis, significance of motif frequency, intensity/coherence analysis, and its corresponding significance testing aim at insights that use aggregated information about microlevel to inform about the properties of system as a whole.

In conclusion, at this rather early stage of its application in the field of psychology, network analysis is mostly an exploratory approach, but that is likely to change with the introduction of more sophisticated methods that may provide additional insights. In turn, this will enhance the development of specific network theories that can be explicitly tested, resulting in unique contributions to our knowledge about psychological phenomena.

If we view network approach as a different way of thinking about psychological constructs, then exploring networks more "deeply" may lead us to interesting and important findings that would otherwise be missed. Those findings can lead to new questions, generate new specific hypotheses, and help form truly progressive network theories of psychological phenomena.

Limitations of This Study. Our goal was to demonstrate three methods by applying them to an illustrative dataset. The dataset, however, has some limitations that are important to note. Although we had an atypically large sample (for psychological research), it featured considerable amount of

missing data, and how exactly to deal with this problem in network modelling is still an open question [68]. Another open issue in psychological networks is measurement error, which is not accounted for. On an interpretative level, since nodes in network are entities, it is not clear whether their associations can be interpreted as conceptual overlap. To the list of open questions that fall beyond the scope of this paper, we may add the common method variance, which could be responsible for observing some of the edges. However, given that we used partial correlations in the network construction, we believe that most of common method variance (except those unique to a pair of variables) is in that way excluded. Furthermore, one of the sources of common method variance, social desirability, is explicitly included in our network because Self-Disclosure is used as indicator of proclivity to give socially desirable answers (the higher the trait, the smaller the proclivity). Finally, although we had a relatively big sample (pairwise), we do not know how selection bias may influence the results. The trade-offs of "big data" in general is that, on the one hand, it provides more diverse and bigger samples, but, on the other hand, self-selection bias can affect results in many different and unexpected ways. This can play out at multiple levels. For instance, FB users may be unrepresentative regarding some of the traits or due to demographics [69], or FB users who used the *myPersonality* application could be, on average, psychologically different. For example, it could be argued that the sample consists of people who are more interested in psychological aspects of reality and in understanding themselves and others when compared to the general population. In line with this possibility, general self-selection may have influenced our findings about the important role of Empathy in the network. Lastly, individuals chose freely to fulfil certain questionnaire(s). Inasmuch as the choice was not random, there is always a possibility that individual psychological attributes influenced that choice (e.g., more depressed individuals could be less likely to fill in an intelligence test).

In the context of those limitations, the findings we arrived at while demonstrating three methods are presented as tentative and their value is in generating new and interesting hypotheses. Furthermore, in our tentative interpretations, due to our network made of well-studied and diverse psychological attributes and due to the scope of this article, we just scratched the surface of many more interesting "small" findings (e.g., each identified triad in Table 3 would be a good starting point for discussion and for generating further hypotheses). That being said, harvesting an already existing dataset, which contains information about many psychological attributes of big number of people, repurposing it to demonstrate "new" methods, and, while doing so, addressing some new and some old questions (network of psychological attributes and cognition-personality relationship) present potentially useful exploratory research.

Future Research. Regarding specific questions related to our dataset, future research would benefit from more theoretically guided inclusion of psychological attributes in the network, including different types of intelligence measures that capture more than g-factor. More objective (behavioral)

measures of attributes would enhance the validity of findings. Longitudinal data (within-subjects networks) and data on specific populations (e.g., regarding mental health, age, gender, and culture) would in addition enable answering questions about network dynamics and network structure. Future research can use simulation studies to determine how exactly each of the methods is affected by differences in network density, size, number of groups, structure, weight distribution, etc. This would be especially interesting for MST, as we explicitly mentioned that it could be used to check the robustness of network estimations. We used PC on what we called “predefined communities,” but when there are no differences between nature of psychological attributes PC might be used in typical way as well, which starts from empirically determined communities (such example is given in SM, section 13). Likewise, the PC measure can be extended in such a way that one could calculate it for positive and negative links separately. In the motif analysis, we looked only at triads; future work can include higher-order configurations, motifs that involve more than three nodes (e.g., bow tie).

Finally, we selected three network metrics for this article, but there are other measures and techniques that could be fruitfully used in the analysis of psychological networks (e.g., coefficient of intramodule activity, missing link prediction). The message is that network science methodology develops rapidly, and psychologists using network analysis would do well to embrace the possibilities these methods offer in both, analysis and stating new research questions, hypotheses, and even theories.

Data Availability

The data used to support the findings of this study were supplied by David Stillwell and Michal Kosinski under license and so cannot be made freely available. Requests for access to these data should be made to David Stillwell, contact@mypersonality.org.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Srebrenka Letina conceived the idea for the study, asked for the data access, did the data processing, analyses and visualizations, and wrote the paper. Tessa F. Blanken, Denny Borsboom, and Marie K. Deserno edited the text and its structure and provided feedback on the manuscript.

Acknowledgments

We thank David Stillwell and Michal Kosinski for allowing the access to the myPersonality database (myPersonality.org). The work on this paper was partially sponsored by Central European University Foundation, Budapest (CEUBPF). The

theses explained herein are representing the author's own ideas but do not necessarily reflect the opinion of CEUBPF. We acknowledge COSTNET (Cost Action CA15109) in funding the short scientific mission which resulted in this work. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement no. 648693). Denny Borsboom is supported by ERC Consolidator Grant no. 647209. We thank Donald Williams for the help in the estimation of nonregularized partial correlation network and Tamer Khraisha for advice on coding and visualizations.

Supplementary Materials

More details about procedures and results of the analyses are organized in 13 sections of the Supplementary Materials: data processing, sample description, description of missing data, descriptive statistics of 26 psychological attributes, the choice of the estimation method, robustness analyses, network of 26 psychological attributes, analysis of network ties, centrality analysis, correlations between four centrality measures in full network and in MST, the MST with different distance measures, the effect of reverse coding on the analyses, and participation coefficient based on empirical (data-driven) communities. (*Supplementary Materials*)

References

- [1] H. L. J. Van Der Maas, C. V. Dolan, R. P. P. Grasman, J. M. Wicherts, H. M. Huizenga, and M. E. J. Raijmakers, “A dynamical model of general intelligence: The positive manifold of intelligence by mutualism,” *Psychological Review*, vol. 113, no. 4, pp. 842–861, 2006.
- [2] A. O. J. Cramer, C. D. Van Borkulo, E. J. Giltay et al., “Major depression as a complex dynamic system,” *PLoS ONE*, vol. 11, no. 12, article e0167490, 2016.
- [3] G. Costantini, J. Richetin, E. Preti, E. Casini, S. Epskamp, and M. Perugini, “Stability and variability of personality networks. A tutorial on recent developments in network psychometrics,” *Personality and Individual Differences*, vol. 136, pp. 68–78, 2017.
- [4] J. Dalege, D. Borsboom, F. van Harreveld, and H. L. J. van der Maas, “Network analysis on attitudes: a brief tutorial,” *Social Psychological and Personality Science*, vol. 8, no. 5, pp. 528–537, 2017.
- [5] V. D. Schmittmann, A. O. J. Cramer, L. J. Waldorp, S. Epskamp, R. A. Kievit, and D. Borsboom, “Deconstructing the construct: A network perspective on psychological phenomena,” *New Ideas in Psychology*, vol. 31, no. 1, pp. 43–53, 2013.
- [6] A. O. J. Cramer, S. van der Sluis, A. Noordhof et al., “Dimensions of normal personality as networks in search of equilibrium: you can't like parties if you don't like people: dimensions of normal personality as networks,” *European Journal of Personality*, vol. 26, no. 4, pp. 414–431, 2012.
- [7] D. Borsboom and A. O. J. Cramer, “Network analysis: An integrative approach to the structure of psychopathology,” *Annual Review of Clinical Psychology*, vol. 9, pp. 91–121, 2013.
- [8] J. Kossakowski and A. O. J. Cramer, “Complex dynamical systems in psychology,” in *Network Science in Cognitive Psychology*,

- M. Vitevitch, Ed., chapter: Complex Dynamical Systems in Psychology, Rutledge Taylor & Francis Group, 2017.
- [9] S. Epskamp, L. J. Waldorp, R. Mötus, and D. Borsboom, "The gaussian graphical model in cross-sectional and time-series data," *Multivariate Behavioral Research*, vol. 53, no. 4, pp. 453–480, 2018.
 - [10] C. D. van Borkulo, D. Borsboom, S. Epskamp et al., "A new method for constructing networks from binary data," *Scientific Reports*, vol. 4, no. 5918, 2015.
 - [11] S. Epskamp and E. I. Fried, "A tutorial on regularized partial correlation networks," *Psychological Methods*, vol. 23, no. 4, pp. 617–634, 2018.
 - [12] C. D. van Borkulo, L. Boschloo, J. Kossakowski et al., "Comparing network structures on three aspects: A permutation test," In press.
 - [13] R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B*, vol. 11, no. 1, pp. 193–197, 1999.
 - [14] T. F. Blanken, M. K. Deserno, J. Dalege et al., "The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks," *Scientific Reports*, vol. 8, no. 1, 2018.
 - [15] P. J. Jones, A. Heeren, and R. J. McNally, "Commentary: A network theory of mental disorders," *Frontiers in Psychology*, vol. 8, Article ID 1305, 2017.
 - [16] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, 2005.
 - [17] P. W. Holland and S. Leinhardt, "Local structure in social networks," *Sociological Methodology*, vol. 7, pp. 1–45, 1976.
 - [18] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.
 - [19] P. L. Ackerman and E. D. Heggstad, "Intelligence, personality, and interests: Evidence for overlapping traits," *Psychological Bulletin*, vol. 121, no. 2, pp. 219–245, 1997.
 - [20] A. R. Jensen, *The g factor: The Science of Mental Ability*, Praeger Publishers/Greenwood Publishing Group, Westport, CT, USA, 1998.
 - [21] A. S. Griffin, L. M. Guillette, and S. D. Healy, "Cognition and personality: an analysis of an emerging field," *Trends in Ecology & Evolution*, vol. 30, no. 4, pp. 207–214, 2015.
 - [22] H. J. Eysenck, "Personality and intelligence: psychometric and experimental approaches," in *Personality and Intelligence*, R. J. Sternberg and P. Ruzgis, Eds., pp. 221–247, Cambridge University Press, Cambridge, UK, 1994.
 - [23] A. J. Berg, G. M. Ingersoll, and R. L. Terry, "Canonical analysis of the MMPI and WAIS in a psychiatric sample," *Psychological Reports*, vol. 56, no. 1, pp. 115–122, 1985.
 - [24] R. Howard and M. McKillen, "Extraversion and performance in the perceptual maze test," *Personality and Individual Differences*, vol. 11, no. 4, pp. 391–396, 1990.
 - [25] P. Salovey and J. Mayer, "Some final thoughts about personality and intelligence," in *Personality and Intelligence*, R. J. Sternberg and P. Ruzgis, Eds., pp. 221–247, Cambridge University Press, Cambridge, UK, 1994.
 - [26] J. F. Kihlstrom and N. Cantor, "Social Intelligence," in *Handbook of intelligence*, R. J. Sternberg and S. B. Kaufman, Eds., pp. 564–581, Cambridge University Press, Cambridge, UK, 2011.
 - [27] D. H. Ford, *Humans as Self-Constructing Living Systems: A Developmental Perspective on Behavior and Personality*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, USA, 1987.
 - [28] S. D. Smirnov, "Intelligence and personality in the psychological theory of activity," in *Personality and Intelligence*, R. J. Sternberg and P. Ruzgis, Eds., pp. 221–247, Cambridge University Press, Cambridge, UK, 1994.
 - [29] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *American Psychologist*, vol. 70, no. 6, pp. 543–556, 2015.
 - [30] D. J. Stillwell and M. Kosinski, "MyPersonality project: example of successful utilization of online social networks for large-scale social research," in *Proceedings of the 1st ACM Workshop on Mobile Systems for Computational Social Science (MobiSys)*, 2012, https://www.gsb.stanford.edu/sites/gsb/files/conf-presentations/stillwell_and_kosinski_2012.pdf.
 - [31] S. H. Schwartz, "An overview of the schwartz theory of basic values," *Online Readings in Psychology and Culture*, vol. 2, no. 2, 2012.
 - [32] L. R. Goldberg, J. A. Johnson, H. W. Eber et al., "The international personality item pool and the future of public-domain personality measures," *Journal of Research in Personality*, vol. 40, no. 1, pp. 84–96, 2006.
 - [33] V. Egan and V. Campbell, "Sensational interests, sustaining fantasies and personality predict physical aggression," *Personality and Individual Differences*, vol. 47, no. 5, pp. 464–469, 2009.
 - [34] L. C. Miller, R. Murphy, and A. H. Buss, "Consciousness of body: private and public," *Journal of Personality and Social Psychology*, vol. 41, no. 2, pp. 397–406, 1981.
 - [35] J. Rust, "The validation of the orpheus minor scales in a working population," *Social Behavior and Personality*, vol. 26, no. 4, pp. 399–406, 1998.
 - [36] D. Watson and J. W. Pennebaker, "Health complaints, stress, and distress: exploring the central role of negative affectivity," *Psychological Review*, vol. 96, no. 2, pp. 234–254, 1989.
 - [37] M. Snyder, "Self-monitoring of expressive behavior," *Journal of Personality and Social Psychology*, vol. 30, no. 4, pp. 526–537, 1974.
 - [38] P. M. Lewinsohn, J. R. Seeley, R. E. Roberts, and N. B. Allen, "Center for epidemiologic studies depression scale (CES-D) as a screening instrument for depression among community-residing older adults," *Psychology and Aging*, vol. 12, no. 2, pp. 277–287, 1997.
 - [39] S. Baron-Cohen and S. Wheelwright, "The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences," *Journal of Autism and Developmental Disorders*, vol. 34, no. 2, pp. 163–175, 2004.
 - [40] W. Pavot and E. Diener, "Review of the satisfaction with life scale," in *Assessing Well-Being*, E. Diener, Ed., vol. 39 of *Social Indicators Research Series*, pp. 101–117, Springer, Dordrecht, Netherlands, 2009.
 - [41] J. Raven, "The Raven's progressive matrices: change and stability over culture and time," *Cognitive Psychology*, vol. 41, no. 1, pp. 1–48, 2000.
 - [42] A. Mohammadi and E. C. Wit, "Bayesian Structure Learning in Sparse Gaussian Graphical Models," *Bayesian Analysis*, vol. 10, no. 1, pp. 109–138, 2015.
 - [43] S. Epskamp, A. O. J. Cramer, L. J. Waldorp, V. D. Schmittmann, and D. Borsboom, "Qgraph: Network visualizations of relationships in psychometric data," *Journal of Statistical Software*, vol. 48, no. 4, pp. 1–18, 2012.

- [44] P. J. Jones, "Package "Networktools";" 2018, <https://cran.r-project.org/web/packages/networktools/networktools.pdf>.
- [45] "NetworkX — NetworkX [Internet];" 2018, <https://networkx.github.io/>.
- [46] V. D. Schmittmann, S. Jahfari, D. Borsboom, A. O. Savi, and L. J. Waldorp, "Making large-scale networks from fMRI data," *PLoS ONE*, vol. 10, no. 9, article e0129074, 2015.
- [47] D. R. Williams, M. Rhemtulla, A. Wysocki, and P. Rast, "On non-regularized estimation of psychological networks," *PsyArXiv [Internet]*, 2018, <https://psyarxiv.com/xr2vfl>.
- [48] D. R. Williams and P. Rast, "Back to the basics: rethinking partial correlation network methodology," *Open Sci Framew [Internet]*, 2018, <https://osf.io/fndru/>.
- [49] M. K. Forbes, A. G. C. Wright, K. E. Markon, and R. F. Krueger, "Evidence that psychopathology symptom networks have limited replicability," *Journal of Abnormal Psychology*, vol. 126, no. 7, pp. 969–988, 2017.
- [50] D. J. Robinaugh, A. J. Millner, and R. J. McNally, "Identifying highly influential nodes in the complicated grief network," *Journal of Abnormal Psychology*, vol. 125, no. 6, pp. 747–757, 2016.
- [51] J. C. Gower, "Some distance properties of latent root and vector methods used in multivariate analysis," *Biometrika*, vol. 53, no. 3-4, pp. 325–338, 1966.
- [52] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, vol. 32, no. 3, pp. 245–251, 2010.
- [53] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, "Intensity and coherence of motifs in weighted complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 6, 2005.
- [54] J. Siltaloppi and S. L. Vargo, "Triads: A review and analytical framework," *Marketing Theory*, vol. 17, no. 4, pp. 395–414, 2017.
- [55] Y. Kalish and G. Robins, "Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure," *Social Networks*, vol. 28, no. 1, pp. 56–84, 2006.
- [56] M. Newman, *Networks: An Introduction*, Oxford University Press, New York, NY, USA, 2010.
- [57] M. A. Serrano, M. Boguñá, and R. Pastor-Satorras, "Correlations in weighted networks," *Physical Review E*, vol. 74, no. 5, Article ID 055101, pp. 1–4, 2006.
- [58] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory," *Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.
- [59] J. A. Davis, "Structural balance, mechanical solidarity, and interpersonal relations," *American Journal of Sociology*, vol. 68, no. 4, pp. 444–462, 1963.
- [60] F. Heider, "Attitudes and cognitive organization," *The Journal of Psychology: Interdisciplinary and Applied*, vol. 21, no. 1, pp. 107–112, 1946.
- [61] R. Toivonen, M. Kivelä, J. Saramäki, M. Viinikainen, M. Vanhatalo, and M. Sams, "Networks of emotion concepts," *PLoS ONE*, vol. 7, no. 1, Article ID e28883, 2012.
- [62] A. P. Christensen, Y. N. Kenett, T. Aste, P. J. Silvia, and T. R. Kwapil, "Network structure of the wisconsin schizotypy scales/short forms: examining psychometric network filtering approaches," *Behavior Research Methods*, vol. 50, no. 6, pp. 2531–2550, 2018.
- [63] S. Epskamp, D. Borsboom, and E. I. Fried, "Estimating psychological networks and their accuracy: A tutorial paper," *Behavior Research Methods*, vol. 50, no. 1, pp. 195–212, 2018.
- [64] D. Borsboom, D. J. Robinaugh, M. Rhemtulla, and A. O. J. Cramer, "Robustness and replicability of psychopathology networks," *World Psychiatry*, vol. 17, no. 2, pp. 143–144, 2018.
- [65] L. Stankov, "Low correlations between intelligence and big five personality traits: need to broaden the domain of personality," *Journal of Intelligence*, vol. 6, no. 2, p. 26, 2018.
- [66] F. S. Ahmed and L. Stephen Miller, "Executive function mechanisms of theory of mind," *Journal of Autism and Developmental Disorders*, vol. 41, no. 5, pp. 667–678, 2011.
- [67] M. A. Brandimonte, N. Bruno, and S. Collina, "Cognition," in *Psychological Concepts: An International Historical Perspective*, P. Pawlik and G. d'Ydewalle, Eds., Psychology Press, Howe, UK, 2006.
- [68] S. Epskamp, *Network Psychometrics [Ph.D. Dissertation]*, University of Amsterdam, Amsterdam, Netherlands, 2015.
- [69] F. T. McAndrew and H. S. Jeong, "Who does what on Facebook? Age, sex, and relationship status as predictors of Facebook use," *Computers in Human Behavior*, vol. 28, no. 6, pp. 2359–2365, 2012.