Network Analysis for the Visualization and Analysis of Qualitative Data

Article *in* Psychological Methods · June 2017

DOI: 10.1037/met0000129

6 authors, including:


Jen Pokorny
University of California, Davis


Alex Norman
Western Sydney University


Anthony Zanesco
University of Miami


Susan Bauer-Wu
Mind & Life Institute, Charlottesville, Virginia

Psychological Methods

Network Analysis for the Visualization and Analysis of Qualitative Data

Jennifer J. Pokorny, Alex Norman, Anthony P. Zanesco, Susan Bauer-Wu, Baljinder K. Sahdra, and Clifford D. Saron

Online First Publication, June 1, 2017. http://dx.doi.org/10.1037/met0000129

Psychological Methods

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000129

Network Analysis for the Visualization and Analysis of Qualitative Data

Jennifer J. Pokorny
University of California, Davis

Alex Norman
Western Sydney University, Sydney

Anthony P. Zanesco
University of California, Davis

Susan Bauer-Wu
Mind and Life Institute, Charlottesville, Virginia

Baljinder K. Sahdra
Australian Catholic University

Clifford D. Saron
University of California, Davis

Abstract

We present a novel manner in which to visualize the coding of qualitative data that enables representation and analysis of connections between codes using graph theory and network analysis. Network graphs are created from codes applied to a transcript or audio file using the code names and their chronological location. The resulting network is a representation of the coding data that characterizes the interrelations of codes. This approach enables quantification of qualitative codes using network analysis and facilitates examination of associations of network indices with other quantitative variables using common statistical procedures. Here, as a proof of concept, we applied this method to a set of interview transcripts that had been coded in 2 different ways and the resultant network graphs were examined. The creation of network graphs allows researchers an opportunity to view and share their qualitative data in an innovative way that may provide new insights and enhance transparency of the analytical process by which they reach their conclusions.

Translational Abstract

Researchers analyzing qualitative data, such as interviews or written statements, often apply codes to these texts as a way to label and organize concepts and themes found in the data. These codes become the basis of analysis and the way in which researchers both understand and explain what is occurring in the data. Determining how important a particular code is compared with the others and how the codes relate to one another is typically limited to tallying the number of times a code occurs (frequency counts) or instances when codes overlap with one another (code co-occurrence). Presenting and communicating findings from the research is therefore also limited to reporting these same measures, along with providing excerpts from the data that correspond to the codes of interest. Visual representations are rare because of the verbal nature of the data. In this article, we create networks using the chronological location of the codes as they were applied to the text, resulting in a visualization that illustrates the interrelations of the codes in the data. By applying methods from network analysis, additional measures reflecting the relative importance of the codes to one another can be extracted from the networks and illustrated visually in the network graphs. Using this method of analysis can help researchers better

Jennifer J. Pokorny, Center for Mind and Brain, University of California, Davis; Alex Norman, Graduate Research School, Western Sydney University, Sydney; Anthony P. Zanesco, Department of Psychology, University of California, Davis; Susan Bauer-Wu, Mind and Life Institute, Charlottesville, Virginia; Baljinder K. Sahdra, Institute for Positive Psychology and Education, Australian Catholic University; Clifford D. Saron, Center for Mind and Brain, University of California, Davis, and MIND Institute, Davis Medical School, University of California, Davis.

This work was supported by Fetzer Institute Grant 2191 and John Templeton Foundation Grant 39970 to Clifford D. Saron, and by gifts from the Hershey Family, Tan Teo, Yoga Research and Education, Mental Insight, and Baumann Foundations, the Santa Barbara Institute for Consciousness Studies, Grant Couch and Louise Pearson, Caroline Zecca-Ferris and anonymous, and other individual donors, all to the laboratory of Clifford D. Saron. Sponsorship in the form of publicity for participant recruitment and discount services were provided by the Shambhala Mountain Center. Portions of this work were previously presented in a poster presentation at the International Symposium for Contemplative Studies in November 2014, and at the UC Davis Center for Mind and Brain, 2015 Research Summit: Perspectives on Mindfulness: The Complex Role of Meditation Research, in May 2015. We thank Alan Wallace for providing expert meditation training. We thank Tonya Jacobs, Brandon King, David Bridwell, and Stephen Aichele for data collection support, and Rachael Whitworth for initial coding of these data. We thank Saron lab staff, undergraduate research assistants, and volunteers for project implementation and subsequent data management support, including Kimberly Waterman and Benhood Saheli for transcription preparation, and Noelle Reis and the Center for Mind and Brain administrative staff for innumerable assists during the project. We also thank Ezequiel Di Paolo at Ikerbasque for initial network visualization prototyping. Finally, we thank our participants for sharing their experiences with us.

Correspondence concerning this article should be addressed to Jennifer J. Pokorny, Center for Mind and Brain, University of California, Davis, 202 Cousteau Place, Suite 268 Davis, CA 95618. E-mail: jenpokorny@gmail.com

understand the relationship of all codes applied to a data set, and offers an additional way to communicate one's analyses and findings in the form of a network visualization.

Keywords: qualitative methods, visualization, networks, interviews, network graphs

Supplemental materials: http://dx.doi.org/10.1037/met0000129.supp

Qualitative research is an important and growing component within the social sciences (Atkinson, Coffey, & Delamont, 2001), yet there remains ongoing debate concerning its legitimacy and accuracy in psychological fields (Bhati, Hoyt, & Huffman, 2014; Kidd, 2002). This is likely driven by a post-positivist perspective in which transparency, multiple measurements, visualization of results, and replicability are seen as valuable and important to the research process. Accordingly, contributing to this general uncertainty may be factors that include difficulty analyzing qualitative data using common statistical methods, trouble relating qualitative findings to other quantitative measures, as well as a general unease at a perceived lack of transparency in reported studies (Hiles, 2008). Furthermore, researchers utilizing qualitative data have struggled to find ways to represent their findings visually. In an era of graphical depiction of data analysis (Healy & Moody, 2014), the inability to provide appropriate visualizations of qualitative analysis is an important challenge to be overcome, as visual communication of results is becoming a norm (Tufte, 2006). These issues are worth resolving, as qualitative research can provide valuable insights that lead to hypothesis generation and experimental testing (Burt & Oaksford, 1999; Paluck, 2010), and ultimately inform real-world application. Moreover, integrating both qualitative and quantitative methodologies, in what is called "triangulation" of data sources (Creswell, 2013), helps minimize the limitations and biases present in both.

Although there are different approaches to the analysis of qualitative data (Madill & Gough, 2008), a common one is the use of coding to identify, relate, and theorize about common content and/or underlying themes (Ezzy, 2002). As the codes themselves do not contain information beyond indexing sections of text, they are used only to indicate the presence of a particular researcher-defined theme or concept in the data. Research articles utilizing qualitative analysis of interview data often provide little detailed description or reporting (and thus transparency) of the analytic process or on the complete set of codes applied to a data set from which findings and conclusions are made (Bringer, Johnston, & Brackenridge, 2004; Hiles, 2008). In an effort to convey coding patterns observed in the data as well as how codes are interrelated, researchers typically report frequencies of code applications and code co-occurrence matrices (Namey, Guest, Thairu, & Johnson, 2007; Ryan, 1999; Saldaña, 2013). However, code frequency and co-occurrence tables may not sufficiently represent the complexity of coded data, particularly in terms of code interrelations. Scientists need to be able to examine the evidence and analytical process by which evidence comes to support conclusions, whether qualitative or quantitative data. In quantitative studies, the researchers may report the results of different analyses that were conducted and how this is then interpreted by the researcher. For example, a researcher hypothesizes that one group of individuals will respond differently to a particular questionnaire than another group of individuals. They have both groups of individuals complete the questionnaire and conclude that, yes, the two groups respond differently than one another. To provide evidence that this was in fact the case, they report the mean of the total score received on a given questionnaire for the two groups of participants. They also conduct an analysis to test whether the scores are significantly different between the two groups. Thus, other researchers (i.e., the readers of the published work) are able to examine the evidence and analytical process by which the initial researchers came to make their claim that the groups respond differently.

This is difficult to do with qualitative analysis. If researchers code their data, they may have hundreds of codes that interrelate to one another. Through their analytic process, the researchers determine that of these hundreds of codes, there are a few core codes of interest, which they then present in a publication. Other researchers typically do not know how many codes there were originally, what those codes were, how prevalent and representative the few codes selected were to the overall data set (within a participant or across participants), or how these codes may have interrelated to other codes in the data set. Therefore, it is a challenge for other researchers to critically examine the conclusions that the initial researchers made, given a lack of transparency in the data itself or the analytic process by which the researchers made their claims. Furthermore, revealing all of the codes and how the codes may interrelate to one another can motivate and prompt research questions by others. The initial researcher may have been focused on examining one particular aspect of the data (as it is an insurmountable effort to explain or address all of the data), whereas others may in fact be quite interested in other aspects that would be entirely missed or unknown as to even existing when only presenting the few themes that are of interest to the initial researcher.

Thus, although qualitative analysis may be appealing to researchers in psychological and social sciences precisely for the ability to preserve the rich subjective voice of study participants, the difficulties in representing the complexity of the data as well as a lack of transparency of the analytical process may make qualitative analysis a less attractive option for many researchers. Some have made inroads to this problem, for example, by performing code co-occurrence cluster analysis, which identifies groupings of codes applied to overlapping portions of text (Guest & McLellan, 2003), and developing additional ways in which to visualize hierarchical clusters (Freeman, 1994). However, there are other important relations between codes that are not captured by such analysis, such as patterns or sequences of codes that occur with temporal consistency but do not overlap. Accordingly, additional methods to identify, measure, and report the relationships of codes applied to a data set are warranted.

One way to try and understand the relations of codes to each other within a data set is to visualize them. Visualization of data is central to the processes of discovery, understanding, and communication in scientific fields (e.g., Tufte, 2001). Recently, data visualization has entered a new epoch, as visualization tools have become more widely accessible (Healy & Moody, 2014). Visualization offers researchers a way to describe, document, and share complex data in a clear and precise manner. Analytically, the graphing of data can assist researchers by showing their properties or "shape", or by depicting patterns. In turn, this can help suggest modeling or theoretical strategies to researchers as well as help to "debug" analyses by showing errors or missing data. Graphical depiction of data is also an efficient and effective way to communicate one's results or findings (Tufte, 2001). Visualization can help in the telling of one's analytical story, by turning complex ideas into simple visual summaries.

Summary descriptions and analysis of qualitative data, especially of interviews, has proved difficult to represent visually, given that qualitative coding is text-based (i.e., recordings or transcriptions of the spoken word) rather than numerical. However, just as researchers working quantitatively have benefitted from the recent influx of data visualization tools and methods—helping them to communicate the interest or importance of even obscure or arcane research topics—so, too, can those working with coding of interview data benefit. Some methods have attempted to find useful ways to graph qualitative data and/or analysis, but qualitative visualization conventions have tended toward code co-occurrence tables (Namey et al., 2007), which can be difficult to read, or word clouds (McNaught & Lam, 2010), which are usually not code-based. Given the difficulty of visualization and few available methods to visualize the core work of coding, most published qualitative research does not visualize the data or analysis at all. What is needed is a way to visualize coding in a way that accurately represents the codes as they were applied to the data as well as conveys a sense of the complexity of the data, as qualitative work typically seeks not to reduce complexity (Ezzy, 2002).

In order to be in keeping with the ethos of transparency in qualitative research (Miguel et al., 2014; Moravcsik, 2014; Tong, Flemming, McInnes, Oliver, & Craig, 2012), a method that seeks to visualize data should be reproducible, that is, able to be produced by anyone with the appropriate tools. Implied in such a suggestion are mathematical or statistical analyses and/or the use of programming to achieve visualization. An advantage of pursuing this strategy for transparency's sake is the potential to add analytical value by providing calculation of the multiple interrelations between codes, such that they might be subject to statistical analysis and related with other measures. In order to answer these gaps in the qualitative analytical tool kit, we sought an algorithmic and reproducible visualization strategy that could help provide transparency of the underlying qualitative data and reveal the interrelations of codes applied to a text, as well as provide quantitative measures of the importance of codes relative to each other.

In the context of scientific studies, in which data are typically numerical in nature, there is the additional predicament of finding a way to relate text to other quantitative dependent measures. For instance, some codes may be thought to correspond with attributes found in other measures, such as questions on a self-report instrument using a Likert scale, but how does one make this comparison? Some have approached this problem by way of mixed methods research that seeks to integrate qualitative data with other quantitative measures (Creswell, 2013). However, conceiving of how exactly this might be done remains a difficult prospect for many researchers, and, in practice, typically means gathering both types of data, then analyzing and discussing them separately. Here, we present a novel approach that speaks to the concerns raised earlier by utilizing recently developed software tools employing network analysis to create a network of qualitative coding that one can visualize, quantify, and reproduce.

Network Analysis

Complex systems, such as a power grid, the World Wide Web, activity within different brain regions, or people within a community, can be understood, studied, and visualized in terms of their connections in a network. Networks consist of two elements of information: individual entities or objects, referred to as *nodes*, and the relationship(s) or interaction(s) between the objects, which are referred to as *edges* (Newman, 2010). Research in the social sciences has been instrumental in furthering network analysis approaches and visualizations, as they seek to identify and understand complex interactions and relationships of individuals (Borgatti, Mehra, Brass, & Labianca, 2009; Brandes, Freeman, & Wagner, 2013; Ward, Stovel, & Sacks, 2011), beginning as early as the 1920s (Freeman, 1996). Qualitative data, such as a coded interview, is itself a complex system of topics, concepts, and themes, with the nodes representing a particular code and the edges reflecting relationships between codes. In some systems, the directionality of the relationship between two nodes is not of relevance or is symmetrical and the edges are undirected (e.g., A–B), whereas in other systems, edges are directional to indicate the flow of information or the direction of an interaction. An example is Individual A making a phone call to Individual B (A → B), or the code "Rumination" preceding the code "Sadness" in a coded interview (Rumination → Sadness). Edges can also be "weighted," meaning the connections have a strength associated with them based on the frequency of a given connection in the network. In a publication citation network, individuals who frequently publish together as coauthors will have higher weighted edges between them than they have with another individual with whom they published infrequently (e.g., Newman, 2001).

From these network representations, one can acquire descriptive information about the network as a whole as well as information regarding specific components of the network (for a review, see Newman, 2003; Wasserman & Faust, 1994). The structure of a network can be assessed in order to gain information about its contents. Typically, it is useful to know the *number of nodes* and *edges* a network contains, to get an estimate of the network's size. Other measures indicating a network's size are *the average path length* and *graph density*. The path length is the number of edges traveled in order to connect any two randomly selected nodes in the graph (Albert & Barabasi, 2002; Newman, 2003). The minimum value is 0 and the maximum value possible is the diameter of the network, which is the longest path between any two nodes. The average path length, therefore, is an average of all the path lengths in the network. Lower values indicate that the network is more interconnected, with many combinations occurring repeatedly. Networks that are highly interconnected (i.e., low average path length) allow information to travel quickly through them, as the nodes are separated only by a few steps. The graph density indicates how close the network is to being theoretically complete, meaning that all nodes would be connected to all other nodes. It is a ratio of the number of edges per node to the number of possible edges. A complete network would have a density value of 1, whereas networks with only a few edges are considered sparse and would have a value closer to 0.

One can also describe the components within the network, such as individual nodes. The *degree* of each node is the number of connections each node has and is one measure of how connected a node is within a network. For directed networks, this can be split into *in-degree* and *out-degree*, which are the number of edges directed toward a node and the number of edges going out of a node, respectively. The *weight of the edges* connecting the nodes can also provide useful information, as it may be a reflection of the quality, intensity, or duration of the relationship between two nodes, and information will travel differently along the network based upon differences in edge weights. For example, individuals who are close friends and frequently interact with one another, and thus have high edge weights between them in a social network representing contact between individuals, will be more likely to exchange goods, and the flu (Christakis & Fowler, 2010; Mossong et al., 2008). Thus, *weighted degree*, which takes into account both the number (degree) and strength (weight) of the edges, can also be used as a measure of a node's connectedness and relative importance within a network (Opsahl, Agneessens, & Skvoretz, 2010).

Networks also show structural variability, as connections between nodes do not occur randomly. Instead, some nodes are well connected to particular others in the network, typically resulting in clusters of nodes. These clusters, or "communities," are groups of nodes that are closely connected with one another but are sparsely connected between other groups of nodes. Various algorithms have been developed to detect communities within networks (Fortunato, 2010), using different techniques and factors to determine how exactly to partition one community from another. The *modularity value* (ranging from −1 to 1) measures the density of edges within a community compared with what would be expected by chance, with positive values indicating the edges inside the community are greater than expected (Newman & Girvan, 2004). The nodes within each community may have similar properties, serve a similar function, or in the case of codes applied to an interview, are more likely to occur close to one another in a given transcript.

How networks are structured then affects how information flows within the network. One measure of flow in a network is *PageRank*, based on work from Philip Bonacich (1972a, 1987), initially developed as a way to determine the importance of web pages within a network based on how likely it was that a selected web page would be visited given clicks on hyperlinks of other web pages (Brin & Page, 1998). Nodes have higher PageRank values if they receive input from other highly ranked nodes and if they send out the information to other nodes. Nodes that may not have many out connections can still gain in rank if they receive input from nodes that have high ranks, as they benefit from the connection and flow of traffic from the higher ranked node. Thus, the PageRank value is a measure of centrality—how important a node is to the flow of information in the network. There are several other measures of centrality that are often used in network analyses, including betweenness centrality, a measure of how frequently a given node is located along the shortest path between two other nodes (Brandes, 2001; Freeman, 1977), closeness centrality, a measure of how close a given node is to all other nodes (Sabidussi, 1966), and eigenvector centrality, in which each node's value is based on the values of the other connected nodes (Bonacich, 1972b), among others.

Qualitative Codes as Networks

When creating a network, one must determine what relational information to use when constructing the edges, although this will depend on the specific needs and desires of each researcher and what the network is intended to depict. Any number of approaches to applying edges between nodes might be utilized to address different research questions. For instance, with qualitative data such as a coded interview, edges could be drawn between codes that share conceptual meaning in a researcher-defined manner to illustrate codes that are conceptually related to one another. Alternatively, data-driven approaches can be applied. One could use proximity, creating links between a word in the text and all other words that occur within a two-word window around the word (Paranyushkin, 2011a, 2011b), which would provide information about how certain words were used with one another in texts. In an interview, one may be interested in the responses participants gave after each interviewer probe, in which case the probe and response could both be coded and connections be drawn from the probe code to the appropriate response codes.

One may also consider whether the edges are weighted or directed. This again depends on the specifics of one's research project. In the case of the proximity approach employed by Paranyushkin (2011a, 2011b), edges were weighted, with weight added for each repeated appearance of word pairs, while the direction of the relationship was not of concern, and, as such, edges were undirected. In the case of map analysis, researchers identify concepts of interest in the text and then define relationships between the concepts with the ability to specify the strength, meaning, sign, and direction of the relationship (Carley, 1993; Carley & Palmquist, 1992). Another approach, similar to what we propose in this article, is to use the chronological location of codes as they were applied in a text to create networks, as this would illustrate the flow of codes, or concepts, over time. This approach was used by Bodin (2012) in order to analyze how undergraduate students enrolled in a mathematical modeling course described their working process during a particle spring simulation. Codes were applied to transcripts of the students' responses in a mutually exclusive manner, meaning that each segment of text had only a single code applied to it, that is, code applications did not overlap. Edges in the network were undirected and created based on code adjacency, with additional weight being given to repeated adjacencies of code pairs. Therefore, codes that occurred next to one another (i.e., were adjacent) in the text were also connected by an edge in the network. In all of the situations just described, the researchers themselves determined the relationships between concepts, which may correspond to the spatial location of concepts to one another in a text, such as proximity or distance, though typically also incorporates the researchers' own interpretation and knowledge of the subject material. Interpretations drawn from network analyses will be influenced by the approach adopted by the researcher for determining edges between nodes. Note that the decisions made concerning how the network is constructed also affect what analyses are possible, as some analyses are not appropriate for use on directed or weighted networks (for a review, see Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006).

Generally, the method proposed here involves exporting data about codes attached to texts within a qualitative analysis software program and using R (R Development Core Team, 2015) to generate the input for the freeware network graphing program, Gephi (Bastian, Heymann, & Jacomy, 2009). The resulting network consists of nodes, which are the code labels attached to the text, and the edges between the nodes, which represent the chronological location of the codes in relation to one another, starting at the beginning and following the order in which they occur within the text. Through this networking process, researchers are able to reliably present qualitative data and analysis in a manner that portrays complex interrelations of codes such that they are easy to read and visually useful for illustrating connections between the codes applied to a data set. Network metrics can then be applied to analyze and explore the qualitative data in novel ways. We present a description of the method and illustrate its application with two different coding techniques that were applied to interview transcripts. Differences in the resulting networks are because of differences in the coding approach applied to the text as the creation of the networks themselves is done in an identical manner in both. Included with each example are the resulting network graphs, along with descriptive network metrics and interpretations.

Overview of the Method

The method presented assumes that one has already coded their text-based data (e.g., recordings or transcripts of interviews). In short, codes are a form of indexing for texts (including written material, interviews, diaries, etc.). A code, which takes the form of a label, is attached to a section of text, which may range in size from a single word to the whole text. Codes are typically used to indicate what kind of information can be found in the section of coded text. Multiple codes can thus be used to indicate the presence of different kinds of information in a single section of text (e.g., an opinion, the gender of a person being spoken about, and an intended action). The indexes of books work in a similar manner—a word or idea functions as the code, and the section of text, or unit of analysis in qualitative terms, is the page number. As such, in their most basic form, codes contain no information other than their labels and the location data for texts to which they have been attached.

Although originally made popular by Glaser and Strauss (1967) as part of their grounded theory methodology, coding of texts has been widely adopted as a technique of data analysis applied to qualitative data, and many different coding techniques have been developed (see Bernard & Ryan, 2010; Saldaña, 2013). It is also worth noting that although certain methods of coding may have been developed within a particular research paradigm, research method and research paradigm are often not mutually exclusive, and one method may be utilized across a number of paradigms (Madill & Gough, 2008), though some considerations should be made (Yanchar, 1997; Yanchar & Williams, 2006). Codes may be applied to texts using a variety of technologies, including pen and paper, but commonly qualitative data analysis (QDA) software is used. Many QDA programs are available for researchers to code their data (e.g., NVivo [www.qsrinternational.com]; ATLAS.ti [atlasti.com]; RQDA [Huang, 2014]; Dedoose [www.dedoose.com]). For the method we provide here, it is recommended that researchers utilize QDA software that is capable of exporting data about the exact location of codes attached to a text (i.e., the character range or precise timestamp of the text where the code is applied), as this information will be used to draw edges between codes. The examples provided here were coded using the online web application Dedoose (Version 6.1.11), though RQDA is also capable of supplying the required data. Note that this network creation method is agnostic in regard to type of coding approach taken by the researcher, and can likely be used with many coding approaches. The creation of the networks themselves is done in a similar manner for any coding style.

Once coding is complete, the coding information can be extracted and graphed as networks. There are several different freely available programs to assist in the creation and analysis of networks, such as Pajek (Batagelj & Mrvar, 2003), Gephi (Bastian et al., 2009), and packages in R (R Development Core Team, 2015), such as *igraph* (Csárdi & Nepusz, 2006), *statnet* (Handcock, Hunter, Butts, Goodreau, & Morris, 2003), and *network* (Butts, 2008, 2015). Here, we have chosen to use Gephi, though the principles and process we outline could be implemented using any of these software tools.

In the examples presented in the following sections, the connections between nodes represent the chronological location of codes as they were applied in the transcript, starting at the beginning and following in order of their appearance. We have chosen to use chronology, as it is not an arbitrary relation between codes and instead represents how the interviews unfolded. An interview is an event that takes place over time, and the network therefore maps that event. Although similar to Bodin (2012), who also used chronological location of codes to create networks, our coding approach allowed for multiple codes to be attached to the same portion of text either completely or partially overlapping (i.e., not mutually exclusive). Furthermore, we were interested in the order in which codes were applied in the text, which was represented as directed edges in the networks. Thus, when codes overlap in the text, fully or partially, each pair of codes is given a bidirectional relationship in the network, for example, $A \leftrightarrow B$. When one code follows another code in the transcript, but does not overlap, there is a directional relationship from the preceding to the following code, for example, $A \rightarrow B$. When pairs of codes were repeated, additional weight was added to the edge connecting the two nodes representing those codes. Therefore, the resulting directed and weighted networks displayed the chronological relationship of codes as they were applied in the transcripts (see Figure 1).

We created an R function that takes as input a Dedoose export file containing code applications and creates a ".gexf" file (Graph Exchange XML Format), which can be opened by Gephi. The output file contains both the list of nodes and corresponding edges, created in the manner described in the preceding paragraph. The source code for this function, instructions as to how to run the R function, and a sample data set are included in the online supplemental materials. Gephi (Bastian et al., 2009) was designed with the explicit intention of visualizing and manipulating large data sets in network graphs. Thus, users can dynamically modify the presentation of the data depending on a priori research questions, patterns revealed in the network graph, or results of analyses conducted on the network itself. This includes changing the layout

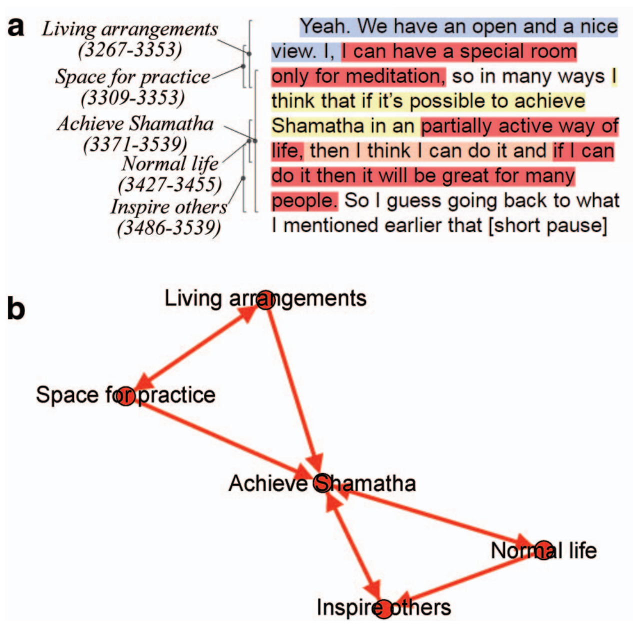

Figure 1. Example of a transcript segment with theme codes applied (1a). The transcript text is on the right. Portions that are highlighted indicate where codes have been applied, with red indicating that more than one code is applied to that portion of the text. On the left are the names of the codes along with the character range of where the coding (highlighting) starts and ends in the transcript. This information is then used to create the corresponding network (1b). Starting at the beginning with Living Arrangements, there is a bidirectional edge between this node and Space for Practice, as these codes overlap with one another in the text. Both codes are then followed by the code Achieve Shamatha, so a directional edge is created from Living Arrangements to Achieve Shamatha, and likewise for Space for Practice. Achieve Shamatha overlaps with Normal Life and Inspire Others in the text and is represented by bidirectional edges between them. Inspire Others follows Normal Life; thus, a directional edge is created from Normal Life to Inspire Others.

of the graph or how nodes and edges are represented (e.g., color, shape, size). Furthermore, users of Gephi may obtain basic network descriptives, such as network diameter and average degree, as well as more complex statistics such as community detection and clustering analyses. Similar to the volunteers-driven development of R packages, the Gephi user community has created many additional plugins to extend the functionality of the program. Note that network modeling methods are currently not available in Gephi, but are available in R packages such as *statnet* (Handcock et al., 2003).

The process for creating each of the networks depicted in the Examples section was the same. Once the network data were loaded into Gephi, we calculated metrics of interest, including average weighted degree, PageRank, and modularity, which is a community detection algorithm based on Blondel, Guillaume, Lambiotte, and Lefebvre (2008). Importantly, this information can be represented visually in the network graph. In our examples, the size of each node corresponds to the weighted degree, and the color of the node represents membership in the assigned community, with nodes of the same color belonging to the same community. Note that the color of a particular node may not be consistent between different networks, as the color represents groupings of communities within the given network, and a node may belong to one community in Network A and a different community in Network B, thus being represented by a different color in another network. The minimum and maximum node size of each network corresponds to the minimum and maximum weighted degrees found in each individual network. We therefore scaled these values using the minimum and maximum values obtained across all our networks to allow for comparisons between different networks. The layout of the networks was determined by applying a ForceAtlas2 algorithm (Jacomy, Venturini, Heymann, & Bastian, 2014), with LinLog mode and Prevent Overlap options selected in Gephi. This algorithm can run continuously and operates such that nodes repel one another while edges pull their nodes toward one another. Given that this is a continuously running algorithm, the spatial location of the nodes provides relative, but not absolute, relational information. As such, every time the ForceAtlas2 layout is applied, even on the same network, the resulting graph will be visually different but relationally the same. The metrics obtained from these networks may be used in conjunction with quantitative measures (e.g., Bruun & Brewe, 2013), such as correlating the weighted degree of codes of interest with scores on a questionnaire. A preliminary example will be provided in the general discussion.

Examples

In this section, we provide examples of networks created from two different types of qualitative coding run through our R script for rendering into network graphic form. The purpose for providing multiple examples is to illustrate how this method can be used with different types of coding approaches. We also include an example of how metrics from the networks may be obtained and used in conjunction with other quantitative measures.

Data Set Used for Examples

Data for these examples are from a study examining biological, physiological, and cognitive effects of participation in a 3-month Buddhist meditation retreat taught by B. Alan Wallace (Wallace, 2006; see Jacobs et al., 2011; MacLean et al., 2010; Rosenberg et al., 2015; Saggar et al., 2012; Sahdra et al., 2011). The Institutional Review Board of University of California, Davis, approved all study procedures and informed consent was obtained. One aspect of the project involved conducting semi-structured interviews with participants in which one of the questions asked was,

> How would you describe, or summarize, your current perspective on and approach to life, including your goals, priorities, and everyday activities? [Follow up prompt] What's most important to you? What's your top priority in life?

The data subset provided here, for illustrative purposes only, consists of responses from six randomly selected participants. We show the network for one individual—"ID A"—along with an aggregated network of the six individuals in the subset.

Subject Code Networks

Method. In this example, we provide a network depiction of Subject codes, which summarize the contents of an interview in terms of "what is being talked *about*" (the subject of thoughts being expressed). What we are here calling "Subject coding" is analogous to "open coding" in the grounded theory framework (Corbin & Strauss, 1990), in that it is an attempt to summarize what exists in the data without reliance upon or reference to theories or research questions; they are data-prompted rather than a priori codes. Subject codes attempt to summarize both general topics of discussion, as well as subtopics within larger general categories, in a way that moves toward generalizability across the data set. They may be applied to sentence fragments, whole sentences, or multiple sentences.

Results. Figure 2a illustrates the Subject code network graph for the entire response given by one participant (ID A), and Figure 2b is the aggregated network graph of Subject coding for all six participants, including ID A. Note that some Subject code labels have been redacted to ensure participant anonymity.

Figure 2a depicts ID A's response to the interview question. The graph illustrates the main concerns of the participant, all of which revolve around the idea of achieving the meditative state of *shamatha* (an advanced Buddhist meditative state; Wallace, 1999) in the course of the participant's normal life (as opposed to in a monastery or retreat). The node Priority (weighted degree [wd] = 54, PageRank [PR] = 0.070) dominates the graph, as one might expect given the question asked, with four communities of nodes surrounding it. Surrounding the Priority node are six nodes with the subsequent highest weighted degrees in the network: Shamatha in Normal Life (wd = 24, PR = 0.034), Changed Priorities (wd 25, PR = 0.029), Achieve Shamatha (wd = 30, PR = 0.042), Normal Life (wd = 28, PR = 0.033), Personal Circumstances (wd = 24, PR = 0.043), and Rearrange Life (wd = 21, PR = 0.026). The first four of these contribute to the network's largest

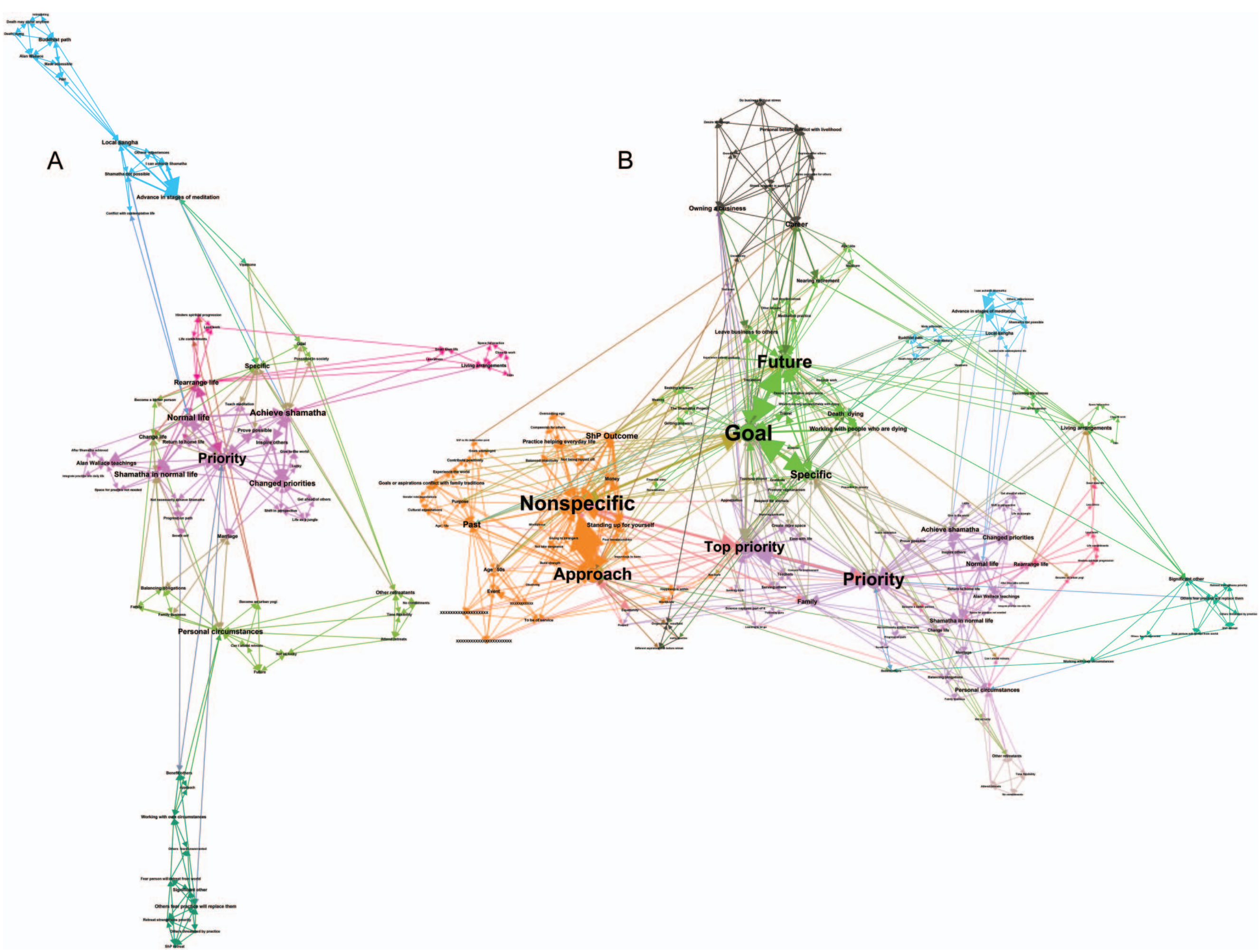

Figure 2. (a) ID/A subject code network. For all following network graphs, node size corresponds to the node's weighted degree, and the colors correspond to communities of nodes determined by the modularity algorithm (Blondel et al., 2008). Node size is relative only to the other nodes within the same network graph. Likewise, colors distinguish communities identified within the one network. Larger arrows correspond to higher edge weight between the pair of nodes. (b) Group subject code network.

community, constituting 30.14% of the nodes, with Priority as the central node (pink in color).

In the group network graph (Figure 2b), one can see that the nodes Goal (wd = 122, PR = 0.047), Future (wd = 98, PR = 0.043), and Priority (wd = 97, PR = 0.046), as well as Nonspecific (wd = 114, PR = 0.036), Approach (wd = 92, PR = 0.028), and Top Priority (wd = 78, PR = 0.044), are central in this network given their large node sizes. Also apparent in the network is that these highly connected nodes form the basis for three of the eight identified communities, constituting 71% of all of the nodes in the network (colors green, pink, and orange). The largest community (25.9%) is the Priority/Top Priority community (in pink), with nodes such as Family, Personal Circumstances, and Normal Life connecting. The second largest community (22.8%, in green) is the cluster around Goal, Future, and Specific. Here one can see that goals are often discussed in the context of the future and participants discussed topics such as Death/Dying, Nearing Retirement, Travel, and Living Arrangements. The Approach/Nonspecific community is the third largest, consisting of 22.2% of all of the nodes (in orange), and includes nodes such as Past, Money, and Cultural Expectations.

These networks are sparsely interconnected with nodes connected to a few others, as evidenced by the low density and average degree values, respectively (see Table 1). Both networks also have relatively high average path lengths and network diameters, which further indicate less interconnections between the nodes in the networks and that it can take many steps to get from one node to another randomly selected node. Many of the edges have a weight of 1, particularly in the individual network, as people may talk about different topics without repeated overlap. However, there are some codes that have highly weighted edges connecting them, such as the edge connecting Nonspecific and Approach in the group network, which has a weight of 8.

Discussion. The Subject code network may serve as a summary of participant responses; if a researcher wishes to see what subjects a participant was talking about without having to consult the whole transcript, a glance at the participant's network may suffice. The graph in Figure 2a also shows the concerns of ID A that are contributors to the central Priority cluster concerning achieving shamatha in normal, or daily, life. For instance, the participant's personal circumstances naturally play a part in this priority, reflected in the Personal Circumstances node (colored olive) and its attendant community nodes such as Balancing Obligations and Family. Separate from this, another community (colored blue-green) shows a number of nodes indexing discussion about others (e.g., Benefit Others, Others Fear Practice Will Replace Them, Significant Other). Distant to these topics of discussion are nodes indexing the participant's concerns regarding their Buddhist practice and community (colored turquoise), with the nodes Local *Sangha* (a community of meditation practitioners) and Advance in Stages of Meditation being central concerns.

The group Subject code network displays the aggregated responses from all participants, summarizing what subjects the group was talking about overall. Within this network, one can see the contribution from the individual networks, such as Normal Life, Achieve Shamatha, and Shamatha in Normal Life surrounding the Priority node in the group graph (colored pink), which was present in the individual network as well (Figure 2a). One also sees the contribution from other individuals, such as Create More Space, Following Guru, Appreciation, Ease With Life, and Concern for Environment as being the priorities that they discussed. Another subject of importance was a discussion concerning participants' careers, in particular, the stress involved and that their job may not be consistent with their personal beliefs, reflected in the brown colored nodes Career, Stress Inherent in Success, Greed Driven, Desire to Change, Do Business Without the Stress, and Personal Beliefs Conflict With Livelihood.

Table 1
Network Metrics of Individual (ID A) and Group Subject Coding Networks

Network metrics	ID A	Group
Nodes (*N*)	73	162
Edges (*N*)	326	1010
In-degree (*M* ± *SE*)	4.47 ± .43	6.23 ± .57
Out-degree (*M* ± *SE*)	4.47 ± .38	6.23 ± .52
Weighted in-degree (*M* ± *SE*)	4.77 ± .51	7.10 ± .80
Weighted out-degree (*M* ± *SE*)	4.77 ± .45	7.10 ± .68
Diameter	8	7
Density	.062	.039
Modularity	.57	.52
Communities (*N*)	5	8
Path length (*M*)	3.36	3.01
PageRank (*M* ± *SE*)	.014 ± .001	.006 ± .0006

Network of Theme Codes

Method. The following example is a network of Theme codes both for one participant (Figure 3a) and for the group (Figure 3b). Theme coding allows researchers to group together concepts, events, persons, and concerns that they interpret to fall within a predefined code label. A theme can be thought of a higher level, more general category (Braun & Clarke, 2006) as opposed to short terms and phrases. In particular, themes are developed in order to capture and summarize something in the data deemed important for research questions. The application of the theme codes to the data is a way for researchers to mark such moments of recurrence in the text, and subsequently analyze how common they were, what commonalities in the contents of the recurrence are present, and their relatedness with other themes (Bernard & Ryan, 2010; Braun & Clarke, 2006; Saldaña, 2013).

Results. The number of nodes in the Theme code networks (see Table 2) is far fewer than the Subject code networks, reflecting that Theme codes represent broader categories of topics, which generalize across individuals. Although there are fewer edges in the Theme code networks as well, compared with the Subject code networks, individual nodes have more connections per node (i.e., higher average degree and average weighted degree). The Theme networks also appear more interconnected, which is corroborated by a higher density and lower network diameter and average path length compared with the Subject networks. The individual network has been grouped into two communities, and the group network into three. However, the modularity values are low, 0.18 and 0.21, respectively; thus, caution should be taken in interpreting the identified communities. The low modularity is likely because of the fact that these are small networks (<30 nodes; for a discussion, see Fortunato, 2010).

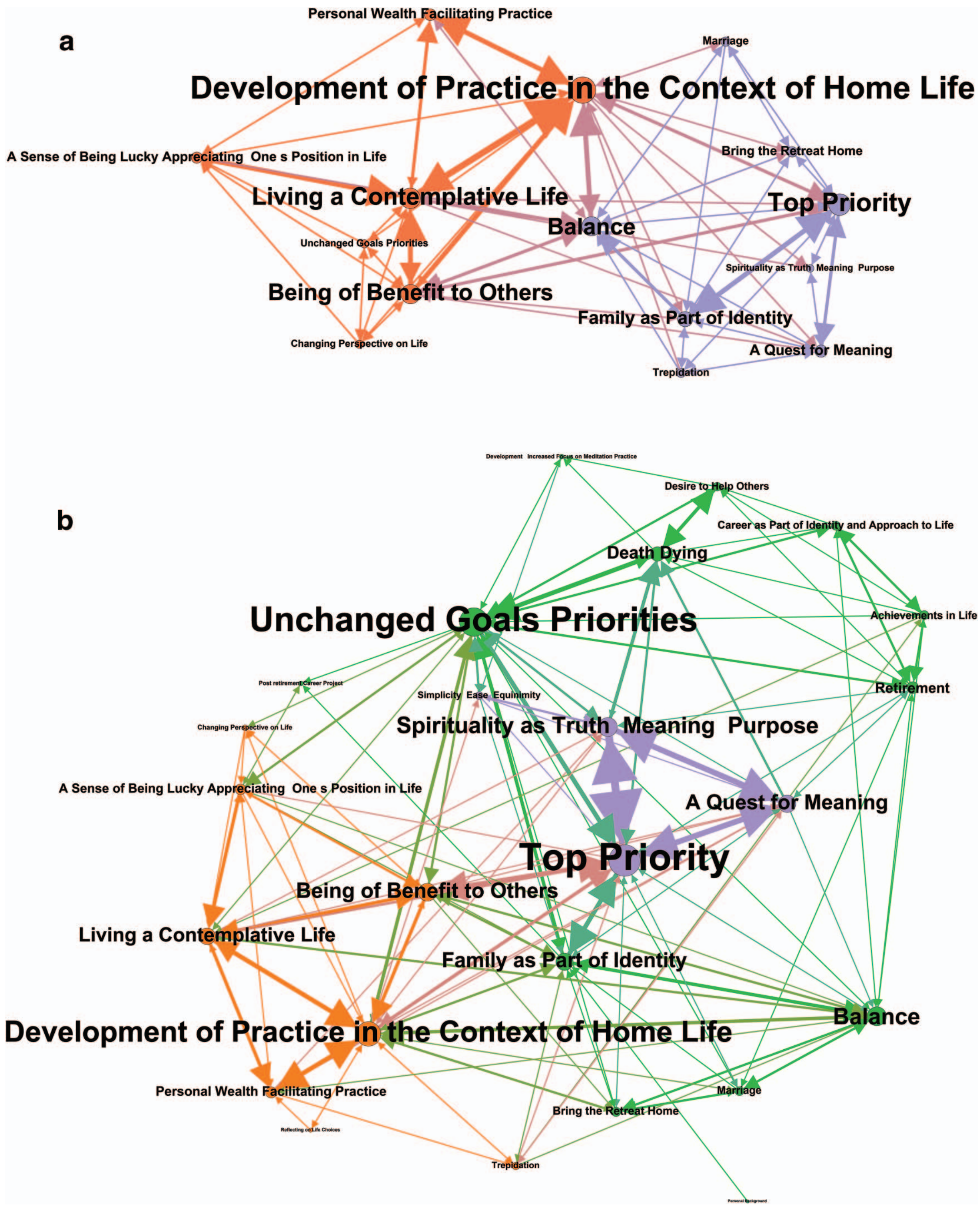

Figure 3. (a) ID A theme code network. (b) Group theme code network.

The network for ID A's transcript reveals several themes as central to the interview, such as Development of Practice in the Context of Home Life (wd = 38, PR = 0.142), Top Priority (wd = 30, PR = 0.122), Living a Contemplative Life (wd = 28, PR = 0.093), as well as Being of Benefit to Others (wd = 26, PR = 0.083) and Balance (wd = 26, PR = 0.107). The central theme node in the group network is Top Priority (wd = 69, PR = 0.110). Other highly central group theme codes identified in the network are Unchanged Goals/Priorities (wd = 62, PR = 0.102), Development of Practice in the Context of Home (wd = 52, PR = 0.076), Spirituality as Truth, Meaning, Purpose (wd = 39, PR = 0.062), and Balance (wd = 39, PR = 0.063).

Discussion. Developing and increasing meditation practice was of central importance to ID A, indicated by the node's relative size (e.g., weighted degree) compared with others in ID A's network. The network also accurately represents the participant's discussion about

Table 2
Network Metrics of Individual (ID A) and Group Theme Coding Networks

Network metrics	ID A	Group
Nodes (*N*)	15	25
Edges (*N*)	93	186
In-degree ($M \pm SE$)	6.2 ± .72	7.44 ± .94
Out-degree ($M \pm SE$)	6.2 ± .69	7.44 ± .87
Weighted in-degree ($M \pm SE$)	8.47 ± 1.36	12.16 ± 1.96
Weighted out-degree ($M \pm SE$)	8.47 ± 1.40	12.16 ± 1.70
Diameter	3	3
Density	.44	.31
Modularity	.18	.21
Communities (*N*)	2	3
Path length (*M*)	1.62	1.75
PageRank ($M \pm SE$)	.067 ± .009	.040 ± .006

balancing obligations, particularly family obligations, with their desire for increased meditation practice, which is depicted by Balance being connected by Development of Practice in the Context of Home Life and Living a Contemplative Life, on one side, and Being of Benefit to Others and Family as Part of Identity, on the other.

In the group network graph (Figure 3b), one can see the Development of Practice in the Context of Home Life node is contained within the second largest community (orange, 32%), along with some important mediating theme nodes, Personal Wealth Facilitating Practice and A Sense of Being Lucky/Appreciating One's Position in Life, indicating that participants saw these as interrelated concepts. The Development of Practice node has edges to nodes in the largest community (green, 52%), such as Unchanged Goals/Priorities, Balance, and Family as Part of Identity, and to nodes in the third and smallest community (purple, 16%), such as Top Priority and A Quest for Meaning. This is a useful, visually simple, and easy-to-read illustration of the complexity of what the theme of Development of Practice in the Context of Home Life is connected with for this subset of participants.

General Discussion

The networks depicted in this article demonstrate how coding that seeks to summarize the content, topics, and themes can be

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	TOT
1. Achievements in Life		1			2																		1		1	9
2. Balance	1			1	2		2	2		1		6	1					5					1	8	9	77
3. A Quest for Meaning							1								5	1								7		22
4. Being of Benefit to Others		1					2					1	1				1							5	1	22
5. Career as Part of Identity and Approach to Life	2	2																					2		2	16
6. Desire to Help Others																				3					1	10
7. Family as Part of Identity		2	1	2								1				1		1						6	2	24
8. Marriage		2										1						1						1	1	10
9. Personal Background																										2
10. Personal Wealth Facilitating Practice		1										3	1													15
11. Post-retirement Career Project																	1							2	2	9
12. Development of Practice in the Context of Home Life		6		1			1	1		3			3	1				6				2		6	1	59
13. Living a Contemplative Life		1		1						1		3					1							1		16
14. Reflecting on Life Choices												1														3
15. Spirituality as Truth, Meaning, Purpose			5																	1		1		8	1	34
16. Trepidation			1				1																	1		3
17. A Sense of Being Lucky/Appreciating One's				1							1		1											2	1	12
18. Bring the Retreat Home		5					1	1				6												5		30
19. Changing Perspective on Life																									1	3
20. Death/Dying						3									1									2	2	20
21. Development & Increased Focus on Meditation Practice																									1	3
22. Simplicity, Ease, Equinimity												2			1										1	12
23. Retirement	1	1			2																			1	2	15
24. Top Priority		8	7	5			6	1			2	6	1		8	1	2	5		2			1		7	115
25. Unchanged Goals/Priorities	1	9		1	2	1	2	1			2	1			1		1		1	2	1	1	2	7		87
Totals	9	77	22	22	16	10	24	10	2	15	9	59	16	3	34	3	12	30	3	20	3	12	15	115	87	

Figure 4. Group theme code co-occurrence matrix. The number in the colored cells is the count of code co-occurrences and the color corresponds to "heat," ranging from blue (low counts) to red (high counts). Numbers across the top row correspond to the numbered theme listed along the first column on the left.

networked to illustrate both overarching patterns and complex interrelations across a qualitative data set. In particular, the metrics extracted from such a network may present a more accurate representation of the relative importance of certain themes and concepts over others, in contrast to frequency counts (e.g., Woodman & Hardy, 2001), code co-occurrence matrices (e.g., Namey et al., 2007), or the particular selection of codes for exposition by a qualitative researcher. To illustrate the problems with code co-occurrence matrices, Figure 4 shows a code co-occurrence matrix for the Theme coding attached to the present group data subset. Although the code co-occurrence matrix presents an accurate depiction of instances in which two or more codes have been attached to the same portion of a text, it is little help in explaining the relations between codes that do not involve overlap. For example, Code A may never overlap with another code, but it might frequently appear before Code B. The code co-occurrence matrix misses this relation, whereas the network graph accurately depicts both it and any other instances in which codes might overlap. Furthermore, network metrics such as reciprocity can also reveal which connections between pairs of codes (dyads) are bidirectional (i.e., reciprocated, A ↔ B) and which are not reciprocated (e.g., Code A always proceeds Code B), while also taking into account differences in edge weights from one code to the other (Squartini, Picciolo, Ruzzenenti, & Garlaschelli, 2013). For instance, if Code A connects to Code B 7 times, while Code B connects to Code A 3 times—the relationship is reciprocated, but it is not symmetrical and is more heavily weighted in the direction of B following A.

Finally, insofar as the purpose of data visualization is to help explain, we argue that the network graphing of coding, such as we have demonstrated here, may be a more effective, easier to read, and more visually engaging method when compared with other solutions, such as co-occurrence matrices like Figure 4. For researchers who may have very large data sets, such as thousands of codes compared with the <200 in these examples, it is possible to filter the data that is visualized in Gephi, such as filtering out edges that are less than a particular edge weight, thereby only looking at codes and connections that are repeated multiple times in the data set. Gephi also has the ability to represent dynamic networks, that is, those that change over time. Although we presented our networks as static, two-dimensional images, a snapshot representative of the entire interview question, the networks are actually dynamic, with codes being applied at different times in the interview. Dedoose can provide a time stamp of where in the interview each code is located. That information can then be passed to Gephi and used to animate the network, essentially playing back the interview with codes appearing and disappearing as they were applied over the interview. Although only a couple of metrics are able to calculated dynamically in Gephi currently, such as degree and clustering coefficient, there are several methods available for dynamic network data to examine and model how the structure and behavior of such networks change over time (see Blonder, Wey, Dornhaus, James, & Sih, 2012; National Research Council, 2003), some of which are available through R packages such as Siena (Snijders, 2014), timeordered (Blonder, 2015), and networkDynamic (Butts, Leslie-Cook, Krivitsky, & Bender-deMoll, 2016). Thus, a strength of the approach we present and the tools we have used is that it is very flexible, allowing researchers to adapt it for their own data and particular research questions.

It may be possible to use the metrics extracted from the networks in conjunction with other quantified data (e.g., Bruun & Bearden, 2014; Bruun & Brewe, 2013), such as scores on questionnaires. For instance, Bruun and Brewe (2013) were interested in the role that student interactions may have on learning course material in the context of an undergraduate physics course. They created multiple social networks of the students in the course based on the frequency and types of interactions students had with one another. Various network centrality measures, such as PageRank and indegree, were then calculated for each individual (i.e., each node) in each of the networks. Analyses included performing correlations between various centrality measures and outcome measures of student knowledge, such as student grades at the end of the course. They found highly significant correlations between network centrality measures and grades in each of the social networks. Note that not every network centrality measure in each social network was correlated with grades—there were some differences based on the type of social network and which centrality measure. However, this is an example in which metrics from the networks were found to relate to other measures of interest.

As an example from our illustrative data set, one of the subscales of Ryff's Psychological Well-Being Scale (PWB; Ryff & Keyes, 1995), the Purpose in Life subscale, may be potentially relevant to the question participants in our sample were asked regarding their goals and priorities in life. According to Ryff and Keyes (1995), a person who scores high on the Purpose in Life subscale has "goals in life and a sense of directedness, feels there is meaning to present and past life, holds beliefs that give life purpose, has aims and objectives for living" (p. 727). Related themes that were developed for this example data set include Future Plans Discussion and Unchanging Goals and Priorities. One could take individual's weighted degree (or other centrality measure) for these Theme codes and relate them to their scores on the PWB Purpose in Life subscale. Another option, however, is to code the transcripts specifically for the PWB and the subscales of interest, using Ryff and Keyes's definitions of each subscale. These codes would then be included in the individual networks and the relevant metric (e.g., weighted degree, PageRank) would be extracted and used in relation to individual's scores on the PWB subscale.

Given that we are only including the small sample of six participants here, the example is meant for illustrative purposes. Our intent was to develop a method that could be used, in a larger sample (e.g., the full study to be reported elsewhere), to see if there was an association between self-report measures and what participants revealed during an interview. We coded the transcripts (used in the previous examples) for High Purpose in Life using Ryff and Keyes's (1995) definition of the Purpose in Life subscale. The code was included in every participant's Subject code network, along with the previously identified subject codes (presented in the Subject Code Networks section). Figure 5 shows the resulting network graph for our example participant, ID A. One can see the codes that were most often coded along with the Ryff High Purpose in Life code, indicating these are the concepts that reflect what gives the participant purpose in life, that is, what their goals are and what gives them direction, purpose, aims, and objectives for their life. For this participant, those would be codes reflecting the participant's plan to change and rearrange their life and priorities so that they can try to achieve shamatha in their normal daily

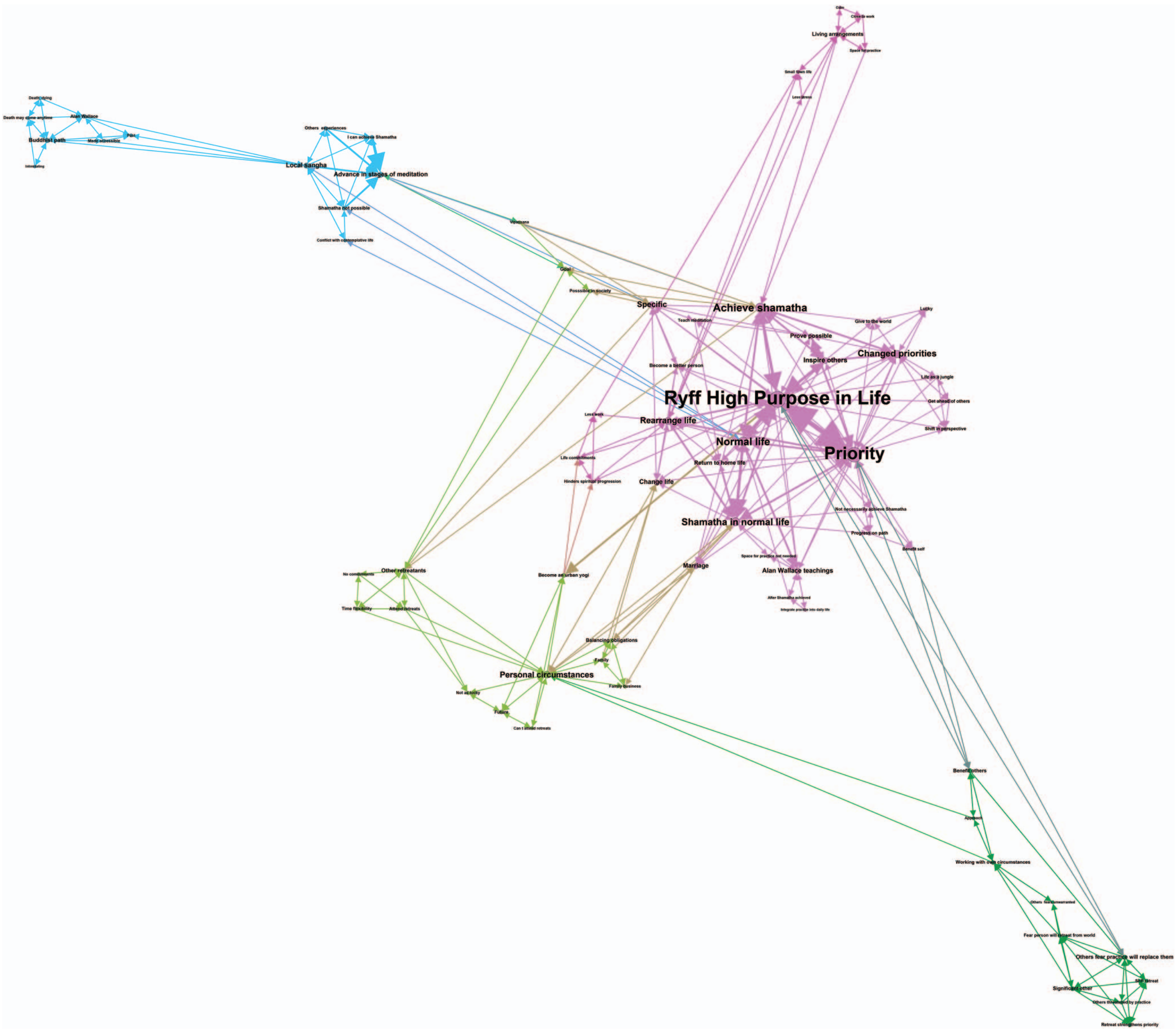

Figure 5. Subject code network for ID A with Ryff High Purpose in Life.

life. From these individual networks, the PageRank and weighted degree values were extracted for the High Purpose in Life code and can be found in Table 3. The tabulation of the metrics is only to show the paring of dependent measures, as the correlations are not statistically significant (wd: $r[4] = -0.18$, $p = ns$; PR: $r[4] = -0.40$, $p = ns$), likely because of the small sample size. This type of analysis demonstrates the ability to relate network metrics to other data (see also Bodin, 2012; Bruun & Brewe, 2013), and will be explored with a larger data set.

Researchers may be interested in determining whether the metrics obtained from the network are meaningful or representative of the larger population, particularly when interpreting group-level data and drawing conclusions based on smaller sample sizes. A common approach in dealing with small sample sizes is to obtain confidence intervals of the statistics using a bootstrapping technique (Efron, 1979; Efron & Tibshirani, 1994). In applying this technique, the data are resampled a large number of times (i.e., 1000) and the statistic of interest is extracted from each resampled set of data, providing a distribution of the statistic from which one can obtain confidence intervals. Typically, this is done by resampling the cases with replacement (i.e., individual samples, represented as rows in a data file, for instance) from the original data set. In extending this to networks, a bootstrapping process resamples the original data, assuming that each row (i.e., node) is an individual sample/case, and creates new networks based upon the resampled data. The metric of interest is then extracted from these resampled networks (see *bootnet* [Epskamp, 2015] for bootstrapping applied to network data). Thus, the data are resampled at the level of the node, which corresponds to an individual. However, in our networks, a “case” (i.e., participant, individual sample) is represented by an entire network consisting of multiple nodes and edges instead of by a single node. We believe a bootstrapping

Table 3
Individual Purpose in Life Measures: Individual Participant's Scores on the Purpose in Life Subscale of the Scales of Psychological Well-Being (Ryff & Keyes, 1995) and the PageRank and Weighted Degree Values of the High Purpose in Life Code Extracted from Each Individual's Subject Code Network

Participant	Purpose in Life score	PageRank	Weighted degree
Participant 1	4.89	.194	38
Participant 2	5.11	.111	65
Participant 3	6.33	.080	39
Participant 4	5.33	.161	31
Participant 5	5.67	.065	71
Participant 6	4.44	.092	59

approach would instead require resampling of the *individual networks*, which aggregate to construct any group-level network. Our networks are also dynamical systems representing flow and change over time. We are working to develop alternative approaches to calculating the confidence intervals of network metrics from networks such as ours, but this is beyond the scope of the current article. As such, as the examples presented here are illustrative and we are not making population-level inferences about the data, as it is based on a small sample size, we did not include confidence intervals in the estimates from our group networks.

Conclusion

We have presented a novel application of using network analysis to graphically depict qualitative data that also provides a potential avenue for the integration of qualitative and quantitative analyses. We hope this approach promotes the use of qualitative methods in the psychological and social sciences, and contributes to the transparency and rigor of qualitative methodologies broadly. Qualitative approaches can strengthen research investigations by capturing complex and rich information about individuals that may be neglected by other methodologies. This allows researchers increased variation in their data collection and analysis methods, which results in greater validity through triangulation of findings.

Depicting codes applied to a text as a network enables the representation and communication of the complex interrelations between codes, beyond code co-occurrence matrices. Although both represent relations between pairs of codes, the network graph provides the relation between a given pair of codes in context, and in relation to all the other codes in the network (i.e., in the interview). The networks are also created using predefined criteria (i.e., chronology) to indicate the relations between codes, rather than on the researcher's interpretation of code relations within a given type of coding (Attride-Stirling, 2001; Carley & Palmquist, 1992). This means that measures derived from the relations between codes are data-driven and not researcher-driven, allowing others the ability to reproduce the graphs from the same code location data. This offers researchers a way to increase the transparency of the evidence used to support their research conclusions. The creation of network graphs also allows researchers an opportunity to view their data in a novel way that provides insights and drives further analysis. Researchers who utilize network analysis to visualize and analyze their qualitative data can make the networks available, revealing all the coding work done on a project allowing other researchers to reinterpret the data, provide potentially competing interpretations, and explore other research questions of interest.

References

Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74,* 47–97. http://dx.doi.org/10.1103/RevModPhys.74.47

Atkinson, P., Coffey, A., & Delamont, S. (2001). A debate about our canon. *Qualitative Research, 1,* 5–21. http://dx.doi.org/10.1177/146879410100100101

Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research, 1,* 385–405. http://dx.doi.org/10.1177/146879410100100307

Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An open source software for exploring and manipulating networks*. Paper presented at the International AAAI Conference on Weblogs and Social Media, San Jose, CA.

Batagelj, V., & Mrvar, A. (2003). Pajek—Analysis and visualization of large networks. In M. Junger & P. Mutzel (Eds.), *Graph drawing software* (pp. 77–103). Berlin, Germany: Springer.

Bernard, H. R., & Ryan, G. W. (2010). *Analyzing qualitative data: Systematic approaches*. Thousand Oaks, CA: Sage.

Bhati, K. S., Hoyt, W. T., & Huffman, K. L. (2014). Integration or assimilation? Locating qualitative research in psychology. *Qualitative Research in Psychology, 11,* 98–114. http://dx.doi.org/10.1080/14780887.2013.772684

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics, 2008,* P10008. http://dx.doi.org/10.1088/1742-5468/2008/10/P10008

Blonder, B. (2015). *timeordered: Time-ordered and time-aggregated network analyses*. R package version 0.9.8. Retrieved from https://CRAN.R-project.org/package=timeordered

Blonder, B., Wey, T. W., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution, 3,* 958–972. http://dx.doi.org/10.1111/j.2041-210X.2012.00236.x

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

Bodin, M. (2012). Mapping university students' epistemic framing of computational physics using network analysis. *Physical Review Special Topics-Physics Education Research, 8,* 010115. http://dx.doi.org/10.1103/PhysRevSTPER.8.010115

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

Bonacich, P. (1972b). Technique for analyzing overlapping memberships. *Sociological Methodology, 4,* 176–185. http://dx.doi.org/10.2307/270732

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

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323,* 892–895. http://dx.doi.org/10.1126/science.1165821

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

Brandes, U., Freeman, L. C., & Wagner, D. (2013). Social networks. In R. Tamassia (Ed.), *The handbook of graph drawing and visualization* (pp. 803–837). Boca Raton, FL: CRC Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3,* 77–101. http://dx.doi.org/10.1191/1478088706qp063oa

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30,* 107–117. http://dx.doi.org/10.1016/S0169-7552(98)00110-X

Bringer, J. D., Johnston, L. H., & Brackenridge, C. H. (2004). Maximizing transparency in a doctoral thesis: The complexities of writing about the use of QSR*NVIVO within a grounded theory study. *Qualitative Research, 4,* 247–265.

Bruun, J., & Bearden, I. G. (2014). Time development in the early history of social networks: Link stabilization, group dynamics, and segregation. *PLoS ONE, 9*(11), e112775. http://dx.doi.org/10.1371/journal.pone.0112775

Bruun, J., & Brewe, E. (2013). Talking and learning physics: Predicting future grades from network measures and Force Concept Inventory pretest scores. *Physical Review Special Topics*. Physics Education Research, *9,* 020109. http://dx.doi.org/10.1103/PhysRevSTPER.9.020109

Burt, K., & Oaksford, M. (1999). Qualitative methods: Beyond beliefs and desires. *The Psychologist, 12,* 332–335.

Butts, C. T. (2008). network: A package for managing relational data in R. *Journal of Statistical Software, 24,* 1–36. http://dx.doi.org/10.18637/jss.v024.i02

Butts, C. T. (2015). *network: Classes for relational data*. R package version 1. 13.0. Retrieved from http://CRAN.R-project.org/package=network

Butts, C. T., Leslie-Cook, A., Krivitsky, P. N., & Bender-deMoll, S. (2016). *networkDynamic: Dynamic Extensions for Network Objects*. R package version 0.9.0. Retrieved from https://CRAN.R-project.org/package=networkDynamic

Carley, K. M. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological Methodology, 23,* 75–126. http://dx.doi.org/10.2307/271007

Carley, K. M., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social Forces, 70,* 601–636. http://dx.doi.org/10.2307/2579746

Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks. *PLoS ONE, 5*(9), e12948. http://dx.doi.org/10.1371/journal.pone.0012948

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Qualitative Sociology, 13,* 3–21. http://dx.doi.org/10.1007/BF00988593

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *Inter Journal Complex Systems, 1695,* 1–9.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7,* 1–26. http://dx.doi.org/10.1214/aos/1176344552

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. New York, NY: Chapman and Hall/CRC.

Epskamp, S. (2015). *bootnet: Bootstrap methods for various network estimation routines*. Retrieved from http://cran.r-project.org/package=bootnet

Ezzy, D. (2002). *Qualitative analysis: Practice and innovation*. London, UK: Routledge.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3–5), 75–174. http://dx.doi.org/10.1016/j.physrep.2009.11.002

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40,* 35–41. http://dx.doi.org/10.2307/3033543

Freeman, L. C. (1994). Displaying hierarchical clusters. *Connections, 17,* 46–52.

Freeman, L. C. (1996). Some antecedents of social network analysis. *Connections, 19,* 39–42.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York, NY: Aldine.

Guest, G., & McLellan, E. (2003). Distinguishing the trees from the forest: Applying cluster analysis to thematic qualitative data. *Field Methods, 15,* 186–201. http://dx.doi.org/10.1177/1525822X03015002005

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). statnet: Software tools for the statistical modeling of network data. *Journal of Statistical Software, 24,* 1548–7660.

Healy, K., & Moody, J. (2014). Data visualization in sociology. *Annual Review of Sociology, 40,* 105–128. http://dx.doi.org/10.1146/annurev-soc-071312-145551

Hiles, D. R. (2008). Transparency. In L. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 890–892). Thousand Oaks, CA: Sage.

Huang, R. (2014). *RQDA: R-based qualitative data analysis*. Retrieved from http://rqda.r-forge.r-project.org/

Jacobs, T. L., Epel, E. S., Lin, J., Blackburn, E. H., Wolkowitz, O. M., Bridwell, D. A., . . . Saron, C. D. (2011). Intensive meditation training, immune cell telomerase activity, and psychological mediators. *Psychoneuroendocrinology, 36,* 664–681. http://dx.doi.org/10.1016/j.psyneuen.2010.09.010

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE, 9*(6), e98679. http://dx.doi.org/10.1371/journal.pone.0098679

Kidd, S. A. (2002). The role of qualitative research in psychological journals. *Psychological Methods, 7,* 126–138. http://dx.doi.org/10.1037/1082-989X.7.1.126

MacLean, K. A., Ferrer, E., Aichele, S. R., Bridwell, D. A., Zanesco, A. P., Jacobs, T. L., . . . Saron, C. D. (2010). Intensive meditation training improves perceptual discrimination and sustained attention. *Psychological Science, 21,* 829–839. http://dx.doi.org/10.1177/0956797610371339

Madill, A., & Gough, B. (2008). Qualitative research and its place in psychological science. *Psychological Methods, 13,* 254–271. http://dx.doi.org/10.1037/a0013220

McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *Qualitative Report, 15,* 630–643.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., . . . Van der Laan, M. (2014). Promoting transparency in social science research. *Science, 343,* 30–31. http://dx.doi.org/10.1126/science.1245317

Moravcsik, A. (2014). Transparency: The revolution in qualitative research. *PS: Political Science & Politics, 47,* 48–53. http://dx.doi.org/10.1017/S1049096513001789

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., . . . Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine, 5,* 0381–0391. http://dx.doi.org/10.1371/journal.pmed.0050074

Namey, E., Guest, G., Thairu, L., & Johnson, L. (2007). Data reduction techniques for large qualitative data sets. In G. Guest & K. MacQueen (Eds.), *Handbook for team-based qualitative research* (pp. 137–163). Lanham, MD: Altamira Press.

National Research Council. (2003). *Dynamic social network modeling and analysis* (R. Breiger, K. Carley, & P. Pattison, Eds.). Washington, DC: National Academies Press.

Newman, M. E. J. (2001). Scientific collaboration networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 64,* 1–8. http://dx.doi.org/10.1103/PhysRevE.64.016131

Newman, M. E. J. (2003). Structure and function of complex networks. *SIAM Review, 45,* 167–256. http://dx.doi.org/10.1137/S003614450342480

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199206650.001.0001

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E: Covering Statistical, Nonlinear, and Soft Matter Physics, 69,* 026113. http://dx.doi.org/10.1103/PhysRevE.69.026113

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

Paluck, E. L. (2010). The promising integration of qualitative methods and field experiments. *Annals of the American Academy of Political and Social Science, 628,* 59–71. http://dx.doi.org/10.1177/0002716209351510

Paranyushkin, D. (2011a). *Identifying the pathways for meaning circulation using text network analysis*. Berlin, Germany: Nodus Labs.

Paranyushkin, D. (2011b). Visualization of text's polysingularity using network analysis. *Prototype Letters, 2,* 256–278. Retrieved from http://noduslabs.com/publications/text-polysingularity-network-analysis.pdf

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Rosenberg, E. L., Zanesco, A. P., King, B. G., Aichele, S. R., Jacobs, T. L., Bridwell, D. A., . . . Saron, C. D. (2015). Intensive meditation training influences emotional responses to suffering. *Emotion, 15,* 775–790. http://dx.doi.org/10.1037/emo0000080

Ryan, G. W. (1999). Measuring the typicality of text: Using multiple coders for more than just reliability and validity checks. *Human Organization, 58,* 313–322. http://dx.doi.org/10.17730/humo.58.3.g224147522545rln

Ryff, C. D., & Keyes, C. L. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69,* 719–727. http://dx.doi.org/10.1037/0022-3514.69.4.719

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

Saggar, M., King, B. G., Zanesco, A. P., Maclean, K. A., Aichele, S. R., Jacobs, T. L., . . . Saron, C. D. (2012). Intensive training induces longitudinal changes in meditation state-related EEG oscillatory activity. *Frontiers in Human Neuroscience, 6,* 256. http://dx.doi.org/10.3389/fnhum.2012.00256

Sahdra, B. K., MacLean, K. A., Ferrer, E., Shaver, P. R., Rosenberg, E. L., Jacobs, T. L., . . . Saron, C. D. (2011). Enhanced response inhibition during intensive meditation training predicts improvements in self-reported adaptive socioemotional functioning. *Emotion, 11,* 299–312. http://dx.doi.org/10.1037/a0022764

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Thousand Oaks, CA: Sage.

Snijders, T. A. B. (2014). Siena: Statistical modeling of longitudinal network data. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining* (pp. 1718–1725). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-6170-8_312

Squartini, T., Picciolo, F., Ruzzenenti, F., & Garlaschelli, D. (2013). Reciprocity of weighted networks. *Scientific Reports, 3,* 2729. http://dx.doi.org/10.1038/srep02729

Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology, 12,* 181. http://dx.doi.org/10.1186/1471-2288-12-181

Tufte, E. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Tufte, E. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.

Wallace, B. A. (1999). The Buddhist tradition of Samatha: Methods for refining and examining consciousness. *Journal of Consciousness Studies, 6,* 175–187.

Wallace, B. A. (2006). *The attention revolution*. Somerville, MA: Wisdom.

Ward, M. D., Stovel, K., & Sacks, A. (2011). Network analysis and political science. *Annual Review of Political Science, 14,* 245–264. http://dx.doi.org/10.1146/annurev.polisci.12.040907.115949

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815478

Woodman, T., & Hardy, L. (2001). A case study of organizational stress in elite sport. *Journal of Applied Sport Psychology, 13,* 207–238. http://dx.doi.org/10.1080/104132001753149892

Yanchar, S. C. (1997). William James and the challenge of methodological pluralism. *Journal of Mind and Behavior, 18,* 425–442.

Yanchar, S. C., & Williams, D. D. (2006). Reconsidering the compatibility thesis and eclecticism: Five proposed guidelines for method use. *Educational Researcher, 35,* 3–12. http://dx.doi.org/10.3102/0013189X035009003

Received January 21, 2016
Revision received November 14, 2016
Accepted November 16, 2016 ■