doi:10.3233/978-1-61499-890-7-86

Application of the Social Network Analysis to Better Comprehend Relationships in Qualitative Health Data

Lawrence T. LAM[a], Mary K. LAM[b,c,1], Charlie S. BENRIMOJ[a] and Victoria G. CARDENS[a]
[a]*Graduate School of Health, University of Technology Sydney, Australia*
[b]*Faculty of Health, University of Technology Sydney, Australia*
[c]*School of Information Technologies, The University of Sydney, Australia*

Abstract. Objective: To explore social network analysis (SNA) as an additional approach to elucidate quantifiable insight from qualitative health-related textual data. **Methods:** Key concepts gained from thematic analyses of a set of qualitative health data obtained from an implementation study was analysed using the Excel Add-on module NodeXL **Results**: Our results show that SNA provided useful visualisation and quantifiable information of the relationship between key concepts obtained from the thematic analysis. **Discussion**: SNA is a useful technique for exploring and analysing qualitative data, particularly when the research interest is in complex relationships that may exist among a large number of qualitative variables. In addition to providing a way to visualise the relationship between concepts, SNA provides metric measures that can be further analysed quantitatively. **Conclusion**: The SNA approach allows researchers to explore deeper relationships that may exist among various variables and enable researchers to derive potentially a fuller and more complete appreciation and comprehension of health-related data.

Keywords. Social Network Analysis, implementation study, qualitative data

Introduction

Routinely collected health data or data collection tailored for health-related research projects contain a reasonably large amount of qualitative data. The approach employed to analyse qualitative data differ from those used to analyse quantitative data. Based on established conceptual framework, qualitative data will be analysed using well-established qualitative approaches. These approaches include but are not limited to content analysis, thematic analysis, discourse analysis, and grounded theory [1] to identify recurring themes. Results of these analyses may lead to the generation of hypotheses for further research which may be conducted using quantitative methods.

Recent years have seen rapid development in implementation research, particularly in the field of health-related research. In health-related implementation research, a large amount of qualitative data is often collected alongside quantitative data. This type of research often aims to test hypotheses that require some form of quantification of relationships among different variables including qualitative data. Conventional qualitative data analytical approaches, in this case, may be limited in achieving the

[1] Corresponding Author.

research objective. In this circumstance, a different analytical approach should be explored as a more appropriate alternative to elucidate the dataset and to extract comprehensive information from the data

Social Network Analysis (SNA) is a data analysis and visualisation technique initially developed by social scientists to describe relationships and links between people, and groups [2,3,4]. This analytical technique involves depicting relationships among individuals or groups visually as well as providing mathematical/ statistical analysis on these relationships. Based on the concept of nodes (individuals) and links (network connections), SNA utilises a set of indicators to describe the characteristics of a network being studied. The two main indicators are network density and centrality [5]. Using values that range from 0 to 1, network density measures how tight knitted a network is, with 0 indicates the network is very sparse and 1 indicates a very dense network. Relationships within a network are quantified using centrality measures. These centrality measures include measurement of degree, betweenness, and closeness. Degree represents the number of direct connections an individual node has within the network. It is a measure of the network activity of specific nodes. Betweenness is a measure of the importance of each node as a bridge or connection between different parts of the network. Nodes with high betweenness centrality play an important role in maintaining the network, as the removal of the nodes with high betweenness centrality would cause the collapse of a network. Closeness, on the other hand, measures how close each node on average is to all other nodes in the network.

As an analytical and visualisation technique, SNA has been adopted beyond social sciences into other disciplines and has been applied in different areas of research. These include human resource [6], organisational management [7], social media information flow and digital tourism studies [8,9]. In recent years this analytic technique has been adopted by computing scientists to develop machine learning algorithms for the investigation of a variety of topics, such as disease association prediction [10]. In health-related research, examples have also been found on the utilisation of SNA in examining the dynamics of health care professionals' interactions and communication within the healthcare system [11].

Despite its growing popularity, little has been identified in the literature of the employment of SNA as an approach for analysing qualitative data, particularly as a means for textual analysis with a purpose of visualising and quantifying relationships among textual data. Using qualitative data collected from an implementation research study conducted in Spain, this study aims to provide an example of how SNA could be used as an analytical approach to gain more insight into relationships among qualitative variables.

1. Methods

1.1 Data Source and Description of the Qualitative Data Set

The data utilised in this analytic study were collected from a previous implementation study designed to introduce an intervention program for enhancing primary health practices in community pharmacies in Spain. The intervention program, conducted by an onsite trainer, involved a series of training workshops for pharmacists to integrate primary health services in a community pharmacy setting. The project was conducted in 2015-2016 with data collected from pharmacists via a self-reported questionnaire. As

part of the data collection, both quantitative and qualitative data about the implementation were obtained from each participating pharmacist. Of particular interest to this study were data on the responses to two main open-ended questions. One addressed the main barriers for the implementation of the program in the individual pharmacy, and the other addressed facilitators of successful program implementation. Respondents were asked to provide detailed descriptions on their views about barriers and facilitators. The written text was extracted as complete sentences to be used as raw data. Collected textual data was then coded according to a list of key concepts established from a previous implementation of a similar program in pharmacies. In total, 45 key concepts were identified. A numeric value was then applied to each key concept such that each number represented a key concept. For example, 1 was the code for "communication with patients", 2 the code for "time", and 3 for "recruitment", etc. The key concepts were then extracted and coded from each statement made by the respondent. Furthermore, most key concepts played a role as a "cause", and at the same time they could also be a "result" in the questions on barriers and facilitators. These coded key concepts formed the data set for the SNA. Ethical approval for the project was granted by the Human Research Ethics Committee of the Universidad de Granada, Spain.

1.2 Analytical Software

SNA were applied to the data using the NodeXL, an open source addon template of Microsoft Excel. NodeXL is a project of the Social Media Research Foundation (https://nodexl.codeplex.com/). This template is designed as a graphic exploration of network data. Apart from providing a graphical depiction of a network relationship, it can also provide metric calculation on important SNA parameters, such as degree, betweenness centrality, and closeness centrality.

2. Results and Discussions

Results obtained from the analysis on the barriers suggested an uneven distribution of the clustering pattern with a maximum degree of 28 and a minimum of 0 suggesting an overall moderate level of connections of all nodes with a Graph density of 0.141 (Figure 1). The average betweenness centrality was 38.489 with a maximum value of 374.334 for the key concept of "personal characteristics" indicating that this variable played the most central role within the whole network. The mean closeness centrality was 0.012 with a maximum value of 0.018 for the key concept of "time". Since the value represented a relative measure of closeness of an individual node to all other nodes within the network, a higher value suggested that the node was closer to other nodes than that of a node with lower value. This reflected that "time" had the closest relationship with all other variables in comparison to other variables within the network. Figure 2 depicts the top five barriers identified in the implementation process and their relationships with other variables, with heavy dots represent more important variables.

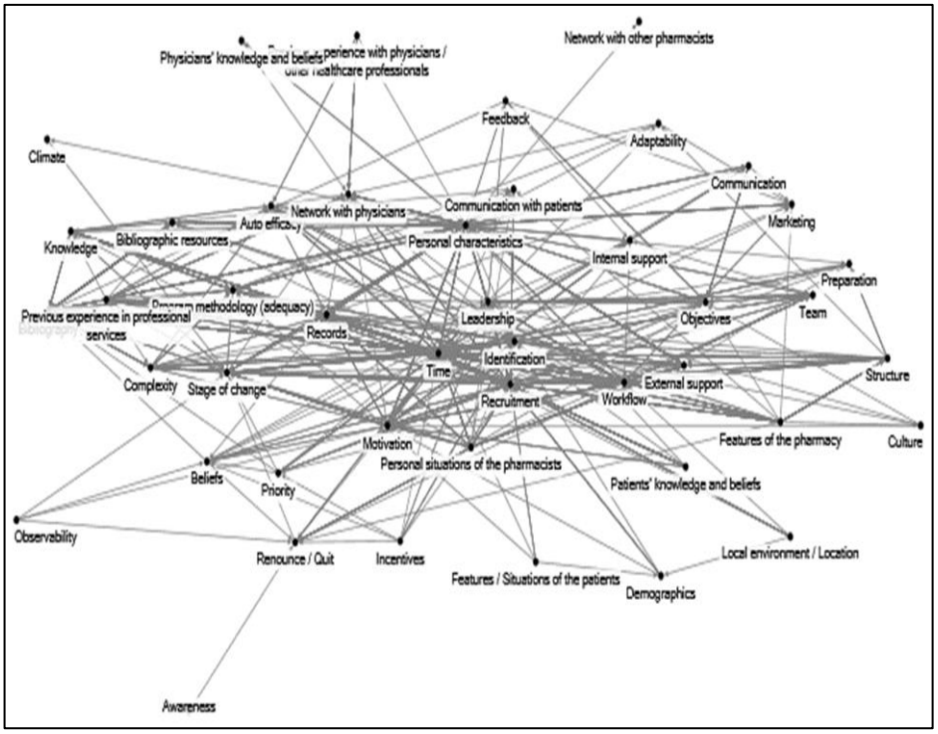

Figure 1. Social Network Analysis of the relationships among variables for barriers.

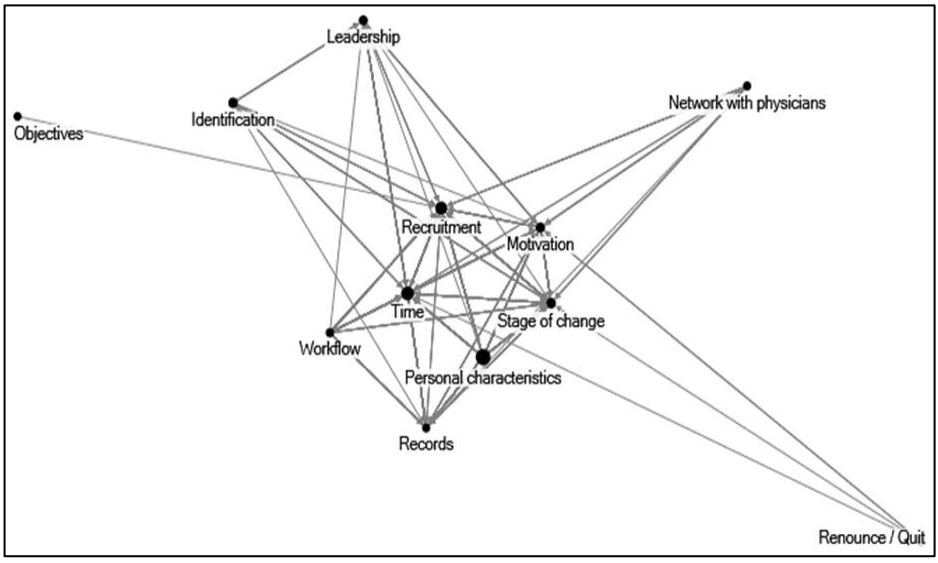

Figure 2. The five most common barriers of the implementation process and the relationships with other variables, with heavy dots representing more important variables.

A similar clustering pattern was identified for facilitators with a maximum degree of 27 and a minimum of 0. Similar to the barriers, "time" was identified as the most important variable within the network with the largest betweenness and closeness

centralities with values of 252.229 and 0.018 respectively. Figure 3 depicts the top five facilitators of the implementation process and their relationship to other variables perceived by the respondents.

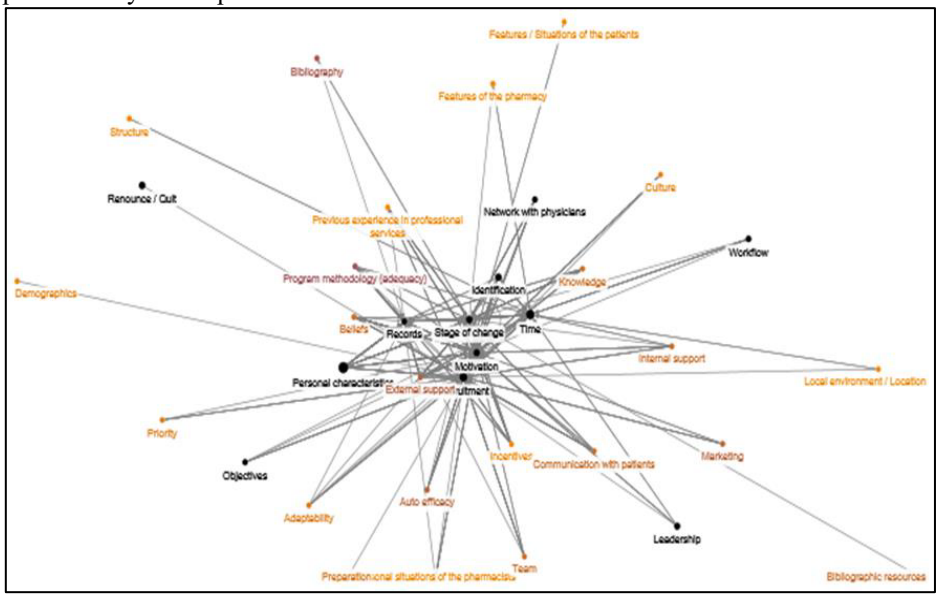

Figure 3. The five most common facilitators of the implementation process and the relationships with other variables, with heavy dots representing more important variables.

Combining the results obtained from the SNA on the barriers and facilitators, the variable "time" seemed to play a central role among all other variables considered. For the interpretation of this variable within the context of the implementation process, time was a crucial element or resource. The lack of time was identified as a main barrier for successful implementation of the intervention program. On the other hand, having sufficient time facilitated the process that could lead on the excepted outcome.

3. Conclusion and Potential Implications

SNA is a useful technique for exploring and analysing qualitative data, particularly when the research focuses on the multiple and complex relationships among a large number of qualitative variables. The advantages of the SNA are the possibility of visualising the data as well as obtaining metric calculation on some relationship measures that can be further analysed quantitatively. SNA also enables management and analysis of large amounts of textual data often collected in a healthcare setting. An additional advantage of this approach is the visualisation of complex relationships among a large number of variables. These relationships may not be easily elucidated with some of the commonly utilised quantitative analytical approaches such as regression modelling. This approach allows researchers to explore the deeper relationships among various variables, and potentially a fuller and more complete appreciation and comprehension of health-related data.

As an analytical approach of qualitative data there are also some limitations that are common to other qualitative analytical approaches. One particular limitation worth

mentioning is that an important aspect of the approach is based on an interpretative process by the researcher. For example, during the extraction and coding process of the textual data, and the conversion of the raw textual data into meaningful themes or concepts, such as culture, climate, etc., it is inevitable that the researcher could have a chance to infuse his/her interpretation into the data. During these processes, systematic biases could be introduced. To safe guard the validity of the resulting data for the SNA, it would be prudent to put in place a process, such as the use of multiple data extractors with proven inter-rater validity, to ensure the quality of data as well as the final results obtained from the SNA.

Another point worth mentioning is that, although the SNA approach plays an important role in the current study by adding much value to a quantitative self-reported survey, it was not in the original study design. The SNA is only supplementing the quantitative data by adding more information to the interpretation of results obtained from quantitative data analyses. To allow the SNA approach to fully demonstrate its usefulness in health-related data analysis, it is important to consider incorporating all elements of a good SNA and qualitative research approach (if appropriate) in the original study design.

References

[1] M.M. Hennink, *Qualitative research methods*, Sage Publication, California, 2011.
[2] S. Wasserman and K. Faust, *Social network analysis: methods and applications*, Cambridge University, New York, 1994.
[3] C. Prell, *Social network analysis*, Sage Publication, California, 2012.
[4] J. Scott, *Social network analysis*, Sage Publication, California, 2013.
[5] A. Martinez, Y. Dimitriadis, B. Rubia, E. Gomez and P. de la Fuente, Combining qualitative evaluation and social network analysis for the study of classroom social interactions, *Computers and Education*, **41** (2003) 353-368.
[6] R. Kase, Z. King and D. Minbaeva, Using social network research in HRM; Scratching the surface of a fundamental basis of HRM, *Human Resource Management* , **52(4)** (2013), DOI: 10.1002/hrm.21546.
[7] T. Grosser, V. Lopez-Kidwell and G. Labianca, A social network analysis of prositive and negative gossip in organisational life, *Group and Organisation Management*, **32 (2)** (2010), 177-212.
[8] I. Himelboim, M. A. Smith, L. Rainie, B. Shneiderman and C. Espina, Classifying twitter topic-networks using social network analysis, *Social Medial and Society*, (2017), DOI 10.1177/2056305117691545.
[9] Q. Luo and D. Zhong, Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites, *Tourism Management*, **46** (2015), 274-282.
[10] Q. Zou, J. Li, Q, Hong, Z. Lin, Y, Wu, H. Shi and Y. Ju, Prediction of micro-RNA-disease associations based on social network analysis methods, *BioMed Research International*, (2015), DOI 10.1155/2015/810514.
[11] L. Clemson, L. Mackenzie, C. Roberts, R. Poulos, A. Tan, M. Lovarini, and et.al., Integrated solutions for sustainable fall prevention in primary care, the iSOLVE project: a type 2 hybrid effectiveness-implementation design, *Implementation Science*, (2017), DOI 10.1186/s13012-016-0529-9.