Social Network Analysis to understand behaviour dynamics in online health communities

A systematic review

Rui E. V. Carvalho
Dept. of Informatics Engineering
Faculty of Engineering, University of Porto
Porto, Portugal
ruievcarvalho@gmail.com

Carla Teixeira Lopes
INESC TEC and
DEI, Faculty of Engineering, University of Porto
Porto, Portugal
ctl@fe.up.pt

Abstract — **Nowadays, online communities are becoming an important resource for health consumers who want to retrieve and share information about health subjects. These communities have the potential to influence patients' health behaviors and increase their engagement with therapies. However, the interaction dynamics in this type of media remains poorly understood what might hinder the development of strategies that facilitate and encourage participation. Social Network Analysis is a technique that tries to expose the hidden channels of communication and information flow, leading to a better understanding of how members relate to each other on online social network. In this study we do a systematic review of the literature regarding the apply this technique in the study of online health communities. We show that this type of studies is scarce and that, in this domain, Social Network Analysis is mainly applied to identify influential key members, as well as the most active members in terms of posting or answering questions.**

Keywords - Social Network Analysis, health, online communities, forums, literature review.

I. Introduction

A. Online Health Communities

Nowadays, social networks, which involves a two-way and direct communication that includes sharing of information between several parties **Error! Reference source not found.**, have the potential to influence patients' health behaviors and increase their engagement with therapies [2]. Online Health Communities (OHC) are an example of a social network that puts in contact individuals with a shared goal or similar interest through a web application. Those groups include health consumers – patients, their family or friends – and might also include health professionals. Members can know each other from the "real world", but the biggest strength of OHC is the possibility to bring together people from different contexts who wouldn't connect otherwise. Within OHCs, members can easily connect with each other using technologies such as blogs, chats, forums, and wikis [3,4].

It has been shown that patients who join OHC experience a significantly improved quality of life, and, most surprisingly, reduced pain levels [5]. Moreover, sharing information about diseases on OHC results in patient empowerment, contributing to the self-efficacy in the use of therapeutics [6].

One could think that the benefits come only from passively reading the forum's content. However, a study **Error! Reference source not found.** showed that posters received more benefits from online communities than lurkers did, including emotional support.

Medical professionals are a little more resistant when it comes to its advantages, being concerned with the misinformation that might be shared among patients. Some OHC use moderators, an external governance, such as health professionals, that moderate discussions and might improve the quality of the discussions. However, literature shows mixed perspectives toward the role of OHC moderators **Error! Reference source not found.**.

B. The use of Social Network Analysis

Network analysts believe that the way that a person lives depends on how that individual is tied into the broader web of social connections. Many go further and believe that the success of societies and organizations might significantly depend on the patterning of their internal structure [8].

Social Network Analysis (SNA) allows the exposure of the hidden channels of communication, collaboration and disconnects between people in groups within a certain organization [9]. It helps to explore the types of relationships that will have some impact in terms of communication and learning, instead of focusing on individual members and relationships. SNA is commonly used in commercial organizations, in order to improve the effectiveness of decision making processes. However, many disciplines have been using it. Although still not having the biggest tradition in healthcare, it seems likely that the study of networks in this domain might be important, for example, in behavior-change interventions, where it can help to identify, target and support relationships that result in better uptake and use of knowledge [8,9].

It was found a systematic review [9] on the literature about the use of SNA in the healthcare domain. The approach is similar to this one, but it mostly addresses the context of physical physical – non-virtual – communities. Actually, only one of the mentioned studies is based on online platforms.

C. Motivation and aim

Due to the continuous spread of SNA into a large number of areas, researchers have developed several ways to identify network effects [10]. This and the importance of the online health communities as structures of support and information exchange motivate this work. Through a systematic literature review on the use of SNA to analyze OHC behavior dynamics, we aim to contribute to a better understanding of the most popular approaches and methods. Besides that, understanding participation as a form of social performance can also enable us to design better systems, encouraging participation. With that in mind, this study is also a help to those who want to perform that analysis on health communities.

II. METHODS

A. Literature Review

A systematic search of 3 scientific online databases - ACM Digital Library, Pubmed and Engineering Village - was conducted at the end of December 2015. In every database we began with a query, to be searched in all fields, containing the terms "social network analysis". Then, since our focus are OHC, we added the term "health" to the query. Since several of the papers found with this last query were related to non-online social networks, we decided to add the term "online".

The number of papers found with these queries, in each system, is presented in Table I. It is interesting to note that, after introducing the term "health", the number of retrieved papers is significantly lower. In the ACM Digital Library and Engineering Village, only 4,4% and 3,2%, respectively, of the studies regarding SNA address health networks. PubMed showed a smaller reduction, which is explained by the fact that the database mainly retrieves medical and health care studies.

TABLE I. NUMBER OF ARTICLES FOUND PER QUERY/DATABASE

	ACM	PubMed	EV
"social network analysis"	1125	726	6695
"social network analysis" health	49	401	213
"social network analysis" health online	12	24	128

After excluding 6 duplicates, we analyzed the titles and abstracts of the remaining 158 articles. To be eligible for this study, articles had to: (1) have the full-text available in the Faculty of Engineering network and (2) describe and report the results of SNA of an online health-related social network. This analysis reduced the list of eligible studies to 24 elements.

After full-text screening, 11 papers were excluded, reducing the list of eligible studies to 13. Most of the papers were excluded because they used techniques other than SNA. Several works used content analysis or text mining techniques as their main analysis methodology. Articles analyzing non-virtual social networks with online surveys, instead of analyzing online social networks, were also excluded.

III. SOCIAL NETWORK ANALYSIS

A. Definition

The Social Network Analysis (SNA) builds on the principles of graph theory to study the relations between actors, and how they influence the overall network. SNA represents communication in terms of *nodes* (which represent actors/members), and *edges* or *arcs* (which represent communication ties). Whereas traditional statistical analysis focuses on actors and their personal attributes, SNA focuses on the structures that emerge out of the relations between actors [11,12,13,14].

B. Types of networks

Social networks can be represented as 1- or 2-mode networks. In 1-mode networks, the nodes are homogeneous, belonging all to the same class. This is the traditional network layout, in which nodes represent people and ties represent some sort of social construct that connects them: advice, friendship, work. 2-mode networks contain two different classes of nodes, and ties exist only from one mode to another. The classes of nodes can represent: members of the online social network platform and the threads they communicate on. The edges would indicate that a specific member has communicated on a specific thread. Because many SNA methods are designed for 1-mode networks, a transformation of the 2-mode network is sometimes necessary [11,12].

C. Concepts

Many SNA concepts will be referred throughout this article, so a brief introduction to their meaning is presented here. Definitions are based on the ones presented in other papers [13,14].

Assortativity – Preference for a network's nodes to attach to similar nodes. "A certain feature is *assortative* in a network if the probability that an arc exists between two nodes having this feature is greater than the probability that an arc exists between two generic nodes" [15].

Betweenness – Reflects the number of members to whom a member is indirectly connecting through their direct links, taking into account the connectivity of the node's neighbors and giving a higher value for nodes which bridge clusters. A node is central if it is used as a path between other nodes.

Bridge - An edge that, if deleted, would cause its endpoints to lie in different components of a graph.

Centralization – A centralized network will "have many of its links dispersed around one or a few nodes", while "in a decentralized one there is little variation between the number of links each node has" [14].

Closeness – Reflects the ability of a node to access information through the other network members. An actor is considered central if he can reach all the other nodes in the fewest possible steps.

Clustering coefficient – Measures the likelihood that two links of a node are themselves linked.

Cohesion – "The degree to which actors are directly connected to each other by cohesive bonds" [14].

Core-periphery – A structure of a network in which some nodes are part of a densely connected core and some are part of a sparsely connected periphery.

Degree centrality – It is calculated by the number of ties an actor (node) has to other actors in the network. For the 2-mode networks mentioned throughout this article, this will be the number of threads each actor communicates on, or the number of participants each thread has.

Density – Individual-level density is the degree in which a respondent's ties know one another proportion of ties among an individual's nominees. Global-level density is the proportion of ties in a network relative to the total number of possible ties.

Diameter – It is the shortest distance between the two most distant nodes in the network.

Eigen-centrality – Takes into account the importance of a node in a network, assigning relative scores/weights to all nodes. Assumes "that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes" [14].

Key players – Actors with high levels of connection to the entire community.

Path length – The distances between pairs of nodes in the network. Average path-length is the average of these distances between all pairs of nodes.

Structural cohesion – The minimum number of members who, if removed from a group, would disconnect the group.

IV. EXAMINATION OF PAPERS

The discussion of the different articles reviewed is divided in three sections. The first one, *Communities of health practitioners,* describes articles that applied SNA techniques to study platforms for information sharing health practitioners [11, 12, 16]. The second section, entitled *Comparison between different communities,* reports works that compare communities either in similar or different online social network platforms [17, 18, 19]. Finally, the *Analysis of Social Networks* section examines articles that analyzed the structure of the social networks per si, not through a perspective of information exchange between practitioners [20, 21, 22, 23, 24, 25, 26, 27].

A. Communities of health practitioners

In 2012, Stewart and Abidi [11] analyzed the communication patterns of an online discussion forum for pediatric pain practitioners. They wanted to understand the participation behaviors across different institutions and occupations; identify relationships between reading and posting on discussion threads; identify the most active and influential members of the community; find a central group of community members. The forum was represented in two different ways: a 2-mode network, where the two classes of nodes are the forum members and the threads they communicate on; an undirected 1-mode network in which a tie between 2 members indicates they have communicated on a thread, being the value of the tie the number of threads they have both communicated on.

In terms of methodology, authors conducted *Kruskal-Wallis* tests to compare reading and posting activities between institutions and occupations. Centrality measures were used to identify the most active and influential members. *Core-periphery* analysis was used to identify the members and threads that are at the center of the 1 and 2-mode networks, respectively. Group centrality analysis was used to explore the interactions between group members in order to determine how different types of professionals, or professionals from different institutions, interact as a community [11].

A different study, conducted by Curran & Abidi [16], used SNA to measure the effectiveness of a discussion forum in terms of the knowledge seeking and sharing patterns of urban and rural emergency department practitioners.

In order to do so, two 1-mode graphs were established: one of them refers to the activity of information seeking, in which *the out-edges* are information requests and the *in-edges* are the nodes that provided the information; the other graph regarding the activity of information sharing, in which the *out-edges* are information share, and the *in-edges* are information received. The density of a binary network reflects the percentage of all possible ties that are actually present in the network, providing an indication of the rate at which information diffuses through the nodes. A density ANOVA using the Structural Blockmodel was run to look for differences in interaction patterns [16].

While the first article applies the 3 main centrality measures in order to identify actors who are in favorable or prestigious positions, the second article uses Freeman's Degree Centrality Measure to identify the out-degree (influence) and in-degree (prestige) points for both networks.

Another study, also from the authors Stewart and Abidi [12], explored the knowledge sharing between health practitioners with SNA in a way similar to the one used in the previous studies. Their goal was to evaluate the communication patterns in the Pediatric Pain Mailing List, identifying content experts and isolating potential subgroups of interest.

As in the previous referred articles, 2-mode and 1-mode networks were established. Firstly, authors identified the most active members of the mailing list using centrality measures - degree, betweenness and closeness. Then, they identify nodes that occupy similar roles within the network through an analysis of structural equivalence. A blockmodel was used to partition the network into exclusive, non-overlapping groups, such that nodes are approximately structural equivalent [12].

B. Comparison between different communities

Chuang and Yang [17] evaluated the transfer of social support in three different online communication formats (forum, journal, notes). For each one of the three computer-mediated communication platforms, two graphs were established in order to differentiate informational and nurturant networks (6 in total). Each node in the six networks was clustered into one of four partitions (isolates, transmit, receiver, carriers) to compare the selections using the blockmodel [17].

SNA was also used by Chomutare, Arsand and Hartvigsen [18] to explore the temporal nature of two large diabetes social networks and compare them with two other non-healthcare social networks: *Slashdot* and *Facebook*. The crawled datasets for each forum were partitioned into periodical sub-datasets.

Firstly, a greedy optimization algorithm was applied to analyze the existent community structures from time-sliced partitions of networks. Then, a similarity analysis, using the *Jaccard* coefficient, was done in order to compare the similarity of the node composition in the network along different periods. During those periods, the clustering coefficient, network diameter, characteristic path length, and average neighbors were analyzed and compared with the two non-health social networks. Finally, a community cohesion analysis tried to understand the bonding factors between actors. Several types of attributes such as years-since-diagnosis, type-of-diabetes, age and gender, were evaluated. Additional network measures were analyzed: degree assortativity and/or homophily network diameter, network density and average degree [18].

Zhang and Yang [19] compared user behaviors between two communication channels on smoking cessation and abstinence - *QuitNet Forum* and *QuitNet Facebook*. Authors combined SNA and user response immediacy results to to identify the differences between channels in terms of their social network structures and actor centralities, and also the difference in user's response immediacy.

To design the graphs, two undirected social networks were built based on the data of both communities. In the network, users were represented as actors (nodes), and the ties connecting them represent the participation in the same post at least once. The number of posts in which both users participated is the tie weight.

The degree centrality, core/periphery structure, density and network centralization were analyzed for both social networks. The authors introduced another metric - the average response time - to represent the mean value of the time taken by a user to make his first comment on the posts he commented [19].

C. Analysis of Social Network

Durant, McCray and Safran [20] aimed to verify the hypothesis that, on an Online Melanoma Discussion Group, users posting questions on Interleukin-2 (IL-2), a treatment prescribed to patients with a more advanced degree of melanoma, receive a stronger response from the Network. After data was extracted from the web, the users were assigned to one of five different user *types,* defined at their member profile: *caregiver*, *patient*, *survivor*, *doctor/nurse* and *member*.

The nodes of the graph represent the members of the forum, and arcs represent the directed communication - from the answerer to the questioner - between them. Then, a sub-network was defined identifying – by the presence of some terms - the threads, posts and users that have discussed a specific theme (IL-2) within the general network.

The authors differentiate the networks by comparing: the density, the arc weights, the node weights, the initial user action (pose or answer a question) and the effect of user type on activity level and on membership duration in both networks.

A quantification of the response provided by both networks to its members was also made, using the additional metrics: the average length of a thread (measured in the number of responses), the average number of days the thread is active (from question posted to last response) and the percentage of unanswered questions. Also, a hub/authority analysis was performed, analyzing the degree centrality in order to identify the influential producers and the consumers whose needs have a higher probability of being satisfied [20].

A study from Dias, Chomutare and Botsis [21] also used SNA techniques to detect user communities in a diabetes health forum, identify the key actors in the network topology and their particular role in the top user communities.

To create the 2-mode graph, users and topics were considered the two classes of nodes of the first and the second mode of the network, respectively. Then, the 2-mode was converted into a 1-mode undirected network, as explained in a previous article [11]. To identify the key actors in the original network, closeness and degree centrality measures were used. To assess the network communities, four standard community detection algorithms were used: the Greedy Optimization, the Affinity Propagation, the Connected Components Cluster, and the Mcode algorithm. The number and size of clusters, and a measure of the connectedness between the communities (modularity), were calculated per algorithm. The authors also evaluated if the key actors of the original network also appeared in the top community for each algorithm used [21].

Cobb, Graham and Abrams [22] adhered to the traditional formal network methods and analytics to characterize the *QuitNet Forum* – which was already mentioned, in the context of Zhang and Yang's study [19] – social network and its participants, describe its structure and identify subgroups, from connections and communication patterns.

The collected data resulted in a large dataset so 5 subsets of participants were delineated. Firstly, 2 subsets of the graph that were connected with a relatively small diameter were identified, resulting in a *strongly connected core*, constituted by "individuals connected by buddy nominations plus observed communications"; and a *densely connected core*, constituted by "individuals connected by symmetric buddy nominations plus a minimum of 5 communications with at least 1 buddy during the observation period". After that, 3 additional subgroups were delineated: a group of new registrants from the initial 4-week period (newcomers), their alters (actors with a tie to another actor of interest, known as an ego), and key players.

Centrality measures (degree, betweenness), network density and core-periphery analysis were the metrics used to compare and/or characterize the network and derived sub-networks [22].

A Twitter based online community was analyzed by Gruzd and Haythornthwaite [23]. SNA was used to examine structures in a 1-month sample of Twitter messages with the hashtag "#hcsmca". The connection between members in the network is implied if a user was mentioned by, replied to, or had a post retweeted by other user. The study was driven in order to evaluate the factors influencing the longevity of the community, its general composition and the importance of a user's professional role on his centrality within the community.

After the representation of the 1-mode network, three social network measures were used to locate influential individual: the total number of posted messages; the number of times that a user's @username is used by another user, representing the in-degree centrality; the number of times a user refers another user's @username, representing the out-degree centrality.

An analysis of variance density test, using the Structural Blockmodel and the Variable Homophily model, was made to verify if different classes of users' professional roles have different interaction patterns and if each class of actors has a different tendency to connect based on social similarity [23].

SNA was also done by Rice, Tulber, Cederbaum, Barman and Milburn [24] to examine the behavior of homeless youth regarding their participation in an online social networking HIV prevention program and which peer leaders were the most essential to the program. The online youth network was defined by two graphs of data extracted from the information about users who 'friended' the program's *Facebook* and *MySpace* profiles. The authors made use of the *common friends* functionality to create a set of mutual ties among the participants for each platform. The size and density of the network, as well as centrality measures such as centrality-degree, centrality-betweenness and eigen-centrality, were calculated for each network. Also, two homophily measures, based on the percent of ties of the same gender and percent of ties of similar age, were created for each person [24].

A recent study from Zhang and Yang [25] aimed to understand the social support exchange patterns and user behaviors of an online smoking cessation intervention program. The content of the messages was analyzed to identify the types of social support given. Then, the data of the social support givers and receivers was extracted from the forum and the exchange support pattern was analyzed. Network analysis and statistical analysis were used to build user interaction models.

Quit status was considered in the network structure and interactions between users at different quit stages were explored. For each one of the social support types (informational and nurturant support), a directed - from the support giver to the receiver – social network was developed. As usual, users are represented as nodes and ties the connection between two nodes. The value of the tie indicates the number of support exchanges between the two users in different threads. By comparing these two social networks, the exchange patterns of different types of social support were investigated using network exposure and blockmodel based on quit stages. Nodes structural equivalence was also analyzed [25].

Bhattacharya, Srinivasan and Polgreen [26] intended to investigate engagement with health agencies, expressed by retweeting, identifying which *handle-level* and *tweet-level* features could influence levels and time span of retweeting. The analysis of some of the *handle-level* features – "network centrality, tweet count, numbers of followers, following, and favorites" – was done with SNA measures. Thus, each account's network was represented by a directed graph in terms of nodes "following" a node (in-degree), and nodes "followed" by a node (out-degree). Betweenness-centrality was the measure used to calculate the importance of a node in its network. Negative binomial hurdle regression models and Cox proportional hazards models were used to determine if the *handle-* and *tweet-level* features correlate with the number of retweets, and with the time to the first and last retweets, respectively [26].

One other study, conducted by Weitzel, Quaresma and Oliveira [27], aimed to provide a framework to help users evaluate the content of health webpages. Thereby, a measure of *Trust,* calculated by a formula that contemplates *quality indicators value* (the mean of the values assigned to a set of ten quality indicators) and *reputation*, was purposed.

Reputation is calculated with a formula described as a "linear combination of centrality measures - betweenness, closeness, PageRank, and eigen-vector - with associated weights". A formula to the edges weight's calculus is also purposed. As proof of concept of the *reputation* calculus, a RT-network, based on retweeted (RT) posts from users who address health subjects, was modeled as a directed graph, where nodes represent users and edges represent a retweet relationship, from the *retweeter* to the *retweeted.* The purposed formulas were then applied to the network [27].

V. Classification of Articles by Used Metrics

The review of the different articles showed differences when it comes to their methodologies. Table II summarizes the different types of analysis that are usually conducted, the metrics used for that purpose and the papers that do so.

TABLE II. CLASSIFICATION OF PAPERS BY GOALS AND USED METRICS

Objective		Metric/Method Used		Articles
Identify influential/key members		Centrality Measures	Degree centrality (in/out)	[11,12,16, 19-24,26]
			Betweenness centrality	[11,12,21, 22,24,26, 27]
			Closeness centrality	[11,12,21, 27]
		Group centrality measures		[11]
		Coreness		[11]
		Eigen centrality		[24,27]
Measure the extent to which an actor is exposed to neighbors with a specific behavioral attribute		Network exposure analysis - the nodes relations based on attributes		[25]
Measure the extent to which nodes of similar degree cluster together		Degree Assortativity		[18]
Measure the extent to which two nodes are connected to the same other		Structural equivalence		[12,25]
Characterize Network	**Assess basic network characteristics**	Density		[12,16,18, 19,20,22, 24]
		Arc weights, Node weights, Size		[11,12, 16-27]
		Diameter		[19,22]
	Assess more advanced network characteristics	Clustering coefficient		[18]
		Characteristic path length		[18]
		Average neighbors		[18]
Determinate user type/group influence		*Kruskall Wallis* test - Analysis of variance comparing degree centrality by type or group		[11,20,23]
Show different interaction patterns between different groups		Structural Blockmodel		[12,16,17, 23,25]
Evaluate tendency for connection based on social similarity		Variable Homophily model		[23]
Assess other general characteristics		Average length of a thread		[11,20]
		Average number of days the thread is active		[11,20]
		Percentage of unanswered questions		[11,20]
Detect communities, number of clusters, average size clusters, max and min size clusters, modularity, number of top 20 nodes in top		Greedy Optimization		[19,21]
		Affinity Propagation		[21]
		Connected Components Cluster		[21]

Objective	Metric/Method Used	Articles
clusters	Mcode	[21]
Evaluate Community similarity	Jaccard index	[18]
Find a central group of community members	Core-periphery analysis	[11,19,22]

It is possible to observe that most of the studies aim to identify influential/key members. Exploring different interaction patterns between different groups is another frequently explored topic. Also, characterizing the network in terms of nodes and arcs weight, as well as in the density and size of the networks is commonly done. Only a few studies perform a clustering and homophily analysis.

VI. CONCLUSIONS

This literature review showed that there are few studies that make use of the Social Network Analysis (i.e. graph detection) as a singular technique to analyze an online health community. It is typically combined with additional metrics and/or tools that hold higher-level analytics such as content analysis tools.

We found that a common methodology among researchers conducting SNA on online health communities is to analyze the network in terms of density, weight of arcs and nodes, and apply typical centrality measures. This suggests that, in this domain, SNA is mainly applied to identify influential key members, as well as the most active members in terms of posting or answering questions. The exploration of different interaction patterns between different groups of the network, typically applying the structure blockmodel technique, is also a frequent subject of study. In addition, this review summarizes the main goals of the analysis and the main used metrics.

To conclude, it should also be noted that the number of studies regarding Social Network Analysis to communities in the health domain was significantly higher when addressing physical – non-virtual – communities.

ACKNOWLEDGMENT

This work is funded by the project "NORTE-01-0145-FEDER-000016", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

REFERENCES

[1] Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*. 2013;15(4):e85. doi:10.2196/jmir.1933.

[2] Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The Role of Social Network Technologies in Online Health Promotion: A Narrative Review of Theoretical and Empirical Factors Influencing Intervention Effectiveness. *J Med Internet Res*. 2015;17(6):e141. doi:10.2196/jmir.3662.

[3] van der Eijk M, Faber MJ, Aarts JWM, Kremer JAM, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. *J Med Internet Res*. 2013;15(6):e115. doi:10.2196/jmir.2476.

[4] J. GP, Molyn L, Marguerite E, et al. The Effect of Group Psychosocial Support on Survival in Metastatic Breast Cancer. *N Engl J Med*. 2001;345(24):1719-1726. doi:10.1056/NEJMoa011871.

[5] van Berkel JJ, Lambooij MS, Hegger I. Empowerment of patients in online discussions about medicine use. *BMC Med Inform Decis Mak*. 2015;15(1):1-9. doi:10.1186/s12911-015-0146-6.

[6] Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. *J Med Internet Res*. 2011;13(4):e122. doi:10.2196/jmir.1696.

[7] Huh J. Clinical Questions in Online Health Communities: The Case of "See your doctor" Threads. *CSCW Proc Conf Comput Coop Work Conf Comput Coop Work*. 2015;2015:1488-1499. doi:10.1145/2675133.2675259.

[8] Freeman L. What is Social Network Analysis? *INSNA*. http://www.insna.org/. Accessed December 16, 2015.

[9] Chambers D, Wilson P, Thompson C, Harden M. Social Network Analysis in Healthcare Settings: A Systematic Scoping Review. *PLoS One*. 2012;7(8):1-10. doi:10.1371/journal.pone.0041911.

[10] Mizruchi MS, Neuman EJ. The effect of density on the level of bias in the network autocorrelation model. *Soc Networks*. 2008;30(3):190-200. doi:http://dx.doi.org/10.1016/j.socnet.2008.02.002.

[11] Stewart SA, Abidi SSR. Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum. *J Med Internet Res*. 2012;14(6):e170. doi:10.2196/jmir.1982.

[12] Stewart SA, Abidi SSR. Understanding Medicine 2.0 - Social Network Analysis and the VECoN System. In: Traver V, Fred ALN, Filipe J, Gamboa H, eds. *HEALTHINF*. SciTePress; 2011:70-79.

[13] Hanneman RA, Riddle M. *Introduction to Social Network Methods*. Riverside, CA: University of California, Riverside; 2005. http://faculty.ucr.edu/~hanneman/nettext/. Accessed December 30, 2015.

[14] Passmore DL. *Social network analysis: Theory and applications*.; 2011. http://train.ed.psu.edu/WFED-543/SocNet_TheoryApp.pdf. Accessed December 30, 2016.

[15] Buccafurri F, Lax G, Nocera A. A New Form of Assortativity in Online Social Networks. *Int J Hum-Comput Stud*. 2015;80(C):56-65. doi:10.1016/j.ijhcs.2015.03.006.

[16] Curran J, Abidi SSR. Evaluation of a discussion forum for knowledge sharing among emergency practitioners: a social network approach. *Stud Health Technol Inform*. 2006;124:941-946.

[17] Chuang KY, Yang CC. How Do e-Patients Connect Online? A Study of Social Support Roles in Health Social Networking. In: *Proceedings of the 6th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction*. SBP'13. Berlin, Heidelberg: Springer-Verlag; 2013:193-200. doi:10.1007/978-3-642-37210-0_21.

[18] Chomutare T, Årsand E, Hartvigsen G. Characterizing development patterns of health-care social networks. *Netw Model Anal Heal Informatics Bioinforma*. 2013;2(3):147-157. doi:10.1007/s13721-013-0033-y.

[19] Zhang M, Yang CC. The differences of user behavior between forum and Facebook for smoking cessation intervention. *Proc Am Soc Inf Sci Technol*. 2012;49(1):1-3. doi: 10.1002/meet.14504901382

[20] Durant KT, McCray AT, Safran C. Social Network Analysis of an Online Melanoma Discussion Group. *Summit on Translational Bioinformatics*. 2010;2010:6-10.

[21] Dias A, Chomutare T, Botsis T. Exploring the community structure of a diabetes forum. *Stud Health Technol Inform*. 2012;180:833-837.

[22] Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. *Am J Public Health*. 2010;100(7):1282-1289. doi:10.2105/AJPH.2009.165449.

[23] Gruzd A, Haythornthwaite C. Enabling community through social media. *J Med Internet Res*. 2013;15(10):e248. doi:10.2196/jmir.2796.

[24] Rice E, Tulbert E, Cederbaum J, Barman Adhikari A, Milburn NG. Mobilizing homeless youth for HIV prevention: a social network analysis of the acceptability of a face-to-face and online social networking intervention. *Health Educ Res*. 2012;27(2):226-236.

[25] Zhang M, Yang CC. Using content and network analysis to understand the social support exchange patterns and user behaviors of an online smoking cessation intervention program. *J Assoc Inf Sci Technol*. 2015;66(3):564-575. doi:10.1002/asi.23189.

[26] Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on twitter. *PLoS One*. 2014;9(11):e112235. doi:10.1371/journal.pone.0112235.

[27] Weitzel L, Quaresma P, d. Oliveira JPM. Evaluating Quality of Health Information Sources. In: *2012 IEEE 26th International Conference on Advanced Information Networking and Applications*. ; 2012:655-662. doi:10.1109/AINA.2012.41.