Healthcare Social Question Answering: Concept Mapping and Cluster Analysis based on Graph Theory

Mohan John Blooma Tran Duc Huy School of Business IT and Logistics Royal Melbourne Institute of Technology Vietnam Ho Chi Minh City, Viet Nam

Email: blooma.john@rmit.edu.vn, huy.tranduc@rmit.edu.vn

Nilmini Wickramasinghe
Epworth Chair Health Information Management,
Health Innovations Research Institute (HIRi)
School of Business IT and Logistics
Royal Melbourne Institute of Technology
Melbourne, Australia

Email: nilmini.wickramasinghe@rmit.edu.au

Abstract

Healthcare Social Question Answering (SQA) services are dedicated platforms for users to freely ask questions regarding their health related concerns and respond to or rate other users' questions. To have a deeper insight into harnessing the rich data collected in healthcare SQA services, this study aims to investigate the concepts discussed using the intricate web of social relationships among questions, answers, associated askers and answerers by applying graph theory, concept mapping and cluster analysis. We collected 4212 question from Drugs.com, one of the popular healthcare SQA services to visualise concepts using Leximancer and cluster similar questions using quadripartite graph-based cluster analysis. The findings demonstrate the openness demonstrated by users on their weight, sleep and drug related questions. The cluster analysis revealed the possibility of applying graph theory to identify similar questions.

Keywords

Cluster Analysis, Concept Map, Graph Theory, Healthcare, Social Question Answering

INTRODUCTION

Healthcare Social Question Answering (SQA) services are dedicated platforms for users to freely ask questions regarding their health concerns, and respond to or rate other users' questions. Patients, doctors, clinicians, nurses, pharmacists, insurance agents among others use healthcare SQA services to seek medical opinion, share remedy, discuss medical issues, and offer emotional support. Online healthcare websites embed SQA services to improve quality, nurture relationship among users, and provide an open user-centred environment for information sharing and exchanging. Examples of healthcare SQA services are MedHelp, BabyHub, and Drugs.com. Patients can create an open SQA discussion on their topic of interest, for example obesity or diabetes, where patients and professionals contribute to the topic, assist and interact with each other in that community. Thus, healthcare SQA service blends technology and social interaction to deliver health information to people (Denecke and Stewart 2011).

Due to growing activities in online healthcare communities, SQA services play an important role in users' health information seeking (Zhang and Zhao 2013). The Pew Internet Project estimated that at least 72% of US Internet users searched for health information online (September 2012). Professionals are preferred for technical issues related to health, however, non-professionals are preferred for personal coping or quick relief issues. Moreover, 60% of US adults tracked their weight, diet or exercise routine. These activities eventually accumulate an information rich community. The data generated by patients and professionals in SQA services empowers patients (Rozenkranz et al. 2013) and contributes to the overall health outcome in a society. However, the challenge faced by the user and research community is to synthesize the complex health data shared in SQA services to build fundamental knowledge, and make informed and timely decisions. Harnessing the wealth of data to identify relevant information is increasingly challenging due to the following three research gaps.

First, there is a need to explore discussion in healthcare SQA services in various diseases, drugs and health topics. Lu et al. (2013) revealed that these diverse topics included announcement, emotion, symptom, procedure, drug, and complications. It is important to have a deeper understanding of the depth, breadth, and dimensions of topics to make the best of the knowledge shared (Zhang and Zhao 2013). Second, few studies have explored the intricate web of relationships among questions, answers, associated askers and answerers (Blooma and Kurian 2012) in healthcare and related SQA services. These social relationships lead to profound network, which could be mined to identify similar concepts and questions. In particular, analysis of the question and answer relationship helps uncover the richness in health queries and responses of the community. Third, user-generated questions are free form with detailed description. For example, "Hi all, I know someone who is taking phentermine. She has a history of alcohol/drug abuse. (as recent as in the last 6 months) Her BMI is no more than 25 and she isn't fat. She hasn't tried working out. Is this dangerous? Why would a doctor prescribe these to her??". Whereas, an example of a simple question is "What's a good weight loss pill?". These two questions are examples of difference in style, form, content, and complexity. Consequently, automatic extraction of similar concepts and identification of similar questions are complex tasks (Lu et al. 2013).

Hence, to address the research gaps, this study applies graph theory to gain deeper insight into the rich data in healthcare SQA services. Graph theory is the mathematical theory about properties and applications of graphs. Modelling web-based social interactions as a graph is a classical application of graph theory to complex networks (Boccaletti et al. 2006). The social relationships between questions, answers, associated askers and answerers are an example of such complex network, and thus are suitable application of graph theory. Applying graph theory to healthcare SQA services, we aim to answer two research objectives.

The first research objective is to identify concepts discussed in questions and answers, we use concept mapping. Concept mapping assists in revealing hidden and evolving patterns in healthcare SQA services (Zhang and Zhao 2013). This paper visualises important concepts with Leximancer (Smith and Humphreys 2006) to analyse the range of knowledge and emotional experiences in healthcare SQA services. The second research objective aims to adapt graph based cluster analysis to identify similar user-generated content by relating questions, answers, askers and answerers (Blooma et al. 2011). Identifying similar content is significant because they facilitate the retrieval of high-quality answers and reduce the time lag in gaining knowledge (Zhang and Zhao 2013).

Through graph theory, we gauge the relationship of questions, answers, askers and answerers to understand better contribution trends of users in the healthcare community. This paper first gives a review of related theories that address the research gap, and details the methodology of visualising concept maps and clustering similar questions in healthcare SQA services. Then, we discuss the results and ends with conclusion.

LITERATURE REVIEW

We present the review of previous studies in four major strands. The first strand reviews healthcare SQA services and prominence of the research gap addressed in this study. The second strand reviews graph theory and its applications in various web-based networks to understand the applicability of graph theory in healthcare SQA services, and gauge the intrinsic relationship of networks. The third strand reviews principles and tools for concept mapping. The final strand reviews various types of cluster analysis for text clustering.

Related to healthcare SQA services, Denecke and Nejdl (2009) performed a content analysis of health information in medical SOA services, medical weblogs, medical reviews and wikis to gain an overview on the content. The results evidenced substantial differences in the content of various health-related Web resource and users. For example, weblogs and SQA services mainly dealt with diseases and medications while wiki and the encyclopaedia dealt with anatomy and procedures. Moreover, patients and nurses described personal aspects of their life while doctors aimed to present health information. Rozenkranz et al. (2013) provided an overview of research in the field of healthcare social media by analysing the extant literature. The study classified stakeholder groups of healthcare system based on their interconnections as individual (example, patients, physicians, pharmacists, health insurance companies) and organization (medical associations, hospital associations, pharmacy associations). Although the study highlighted patient centeredness as well as the dissemination of different types of health information, one of the limitations was that literature reviewed was restricted in terms of its selection of sources and no medical databases were used. Goeuriot et al. (2012) worked on opinion mining by creating a medical opinion lexicon for drug reviews. Yang et al. (2012) used association mining in MedHelp discussions to determine adverse drug reactions. Tang and Yang (2012) further identified influential users of online healthcare community by incorporating users' reply relationship, conversation content and response immediacy to capture both explicit and implicit interaction between users, and tested this in MedHelp. From the review of recent studies related to healthcare SQA service, there is high potential in exploring the intricate web of relationships among questions, answers, associated askers and answerers to identify popular concepts and similar content.

Applying graph theory to web-based social interactions, two typical features of these networks with different origins are community structure (Arenas et al. 2004) and assortative mixing (Newman 2002), both have a clear analogy in real-life experiences. As a multipartite network representation of web-based social interactions, both interacting agents (humans, users of the Web portals) and subjects of their interactions (music, movies, books, postings) in a social network can be represented by nodes on the graph, and their mutual connections can then be analysed in detail. Studies by Lambiotte and Ausloos (2005; 2006) are examples of social connections related to music, where communities related to music genres have been detected. Quadripartite graph network are composed of four kinds of nodes. For example, Blooma and Kurian (2012) used questions, answers, askers and answerers to identify similar questions. The quadripartite graph network was able to overcome the lexical mismatch problem of short questions by using the network relationship. For this study, we adopt quadripartite graph network to identify similar questions.

Concept mapping is one of the methods to visualise the relationship between key concepts graphically (Novak and Gowan 1984). Concept mapping is widely used in medical domain, Watson et al. (2005) analysed patient records to investigate possible relationships surrounding the concepts of heart and tumour, and suggested comparing the rich result of textual analysis to quantitative facts in order to profile high-risk patients. Hepworth and Paxton (2007) conducted interviews about three aspects of help seeking in patients with eating disorder: problem recognition, barriers to help seeking, and prompts to help seeking. Hewett et al. (2009) explored about intergroup communication between doctors, and interpretatively analysed the major themes. These studies use Leximancer as it reduces human subjectivity during the coding stage, allows rapid analyses of extremely large corpora of text, works in an unsupervised manner without prior information about the text, and produces visual relationship network between concepts for comparative insight at both local and global scope of a document set (Smith and Humphreys 2006; Stockwell et al. 2009b). Hence, for this study, Leximancer is used to reveal hidden patterns evolved during discussion and gain insight on the range of knowledge and emotional feedback shared in healthcare SQA services.

Cluster analysis is the grouping of similar objects into meaningful, mutually exclusive groups based on similarities among the objects. Cluster analysis is used as an analytical tool in IS research for classifying configurations of various entities comprising the IT artefact. For example, Sun (2012) identified four clusters of adaptive system usage behaviours, and three triggering conditions that revealed how people engage in the use of system features. Based on the analysis of 55 IS applications of cluster analysis, Balijepally et al. (2011) reaffirmed that cluster analysis is a valuable tool for IS research, and they highlighted the need to adopt a conservative approach in its application and reporting. However, few studies in IS research focused on clustering similar user-generated content in virtual communities like healthcare SQA services. Although cluster analysis of text documents has been well studied (Joo and Lee 2005), few studies have applied this to SQA services because of the nature of questions. Blooma and Kurian (2012) used graph theory to cluster similar questions using hierarchical clustering algorithm based on the relationship between question, answer, asker and answerer. This study adapts the quadripartite graph based cluster analysis proposed by Blooma and Kurian (2012) to identify similar questions in a healthcare SQA service.

METHODOLOGY

To find a source for rich user-generated data in terms of questions and answers, we selected one website among the popular healthcare SQA services. The two measures for popularity were web award nomination for developer trend (webbyawards.com), and web traffic for user trend (alexa.com). Drugs.com, a comprehensive website about drug information, satisfied both measures. Another reason of the selection was the presence of user profile for meaningful long-term interactions and social presence. Weight and sleep were selected as the topic of discussion for this study for the following reasons. First, weight and sleep represents key societal challenge with far reaching implications for patients and professionals in common. Second, weight and sleep focus on the health and well-being of individuals, and lead to related topics like depression, food and exercise. Asides from investigating about the specific conditions, we were also interested in details not specific to any particular condition, therefore, we collected data from sleep group to compare to weight group for common themes as both groups have relatively the same number of questions.

For data collection, we developed a web crawler using Scrapy framework (scrapy.org) to collect data from the "weight" related support groups in "Q&A" section of Drugs.com (http://www.drugs.com/answers/). For each question, we collected question title, question detail, ID of asker, all answers to the question, rating of each answer, and ID of each answerer. The data collection was performed in June 2014. As each question contained one or more answers, to process one answer for each question, we chose the best answer by selecting answers with highest rating. In total, we have 4212 pairs of questions and their best answers. Table 1 gives the statistics of data collected. The collected data was pre-processed and analysed using graph theory.

Graph theory was applied to model the relationship between questions, answers, askers and answerers as illustrated in Figure 1. There is a deep relationship between the nodes in a Question-Answer-Asker-Answerer network. This assumption is based on the coupled mutual reinforcement principle proposed by Bian et al. (2009). Graph theory leads to questions and its answers as related content, and asker and respective answerer as related users. This study used graph theory in two stages. First stage, we adapted graph theory and considered Question-Answer as a unit to identify common concepts and themes in a question-answer unit (Figure 3). In second stage, we used Question-Answer-Asker-Answerer relationship to cluster similar questions. These two stages are detailed below and the findings are discussed.

Table 1. Data Statistics

Statistic	Value
Number of questions	4212
Average number of words in a question	37.9
Number of answers	4212
Number of distinct answerers	911
Number of distinct askers	3936
Number of members who are both askers and answerers	173

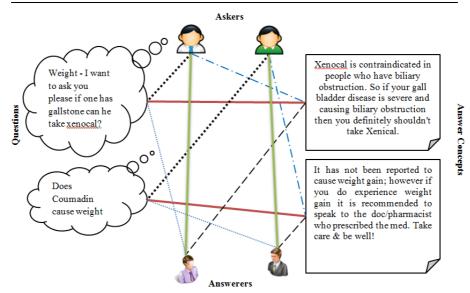


Figure 1: Questioner-Answer-Answerer network relationship using graph theory

Leximancer was used to visualize the concepts discussed in the data (Smith 2003; Smith and Humphreys 2006; Stockwell et al. 2009). Leximancer is very sensitive to the length of context block. Two settings govern this parameter: the number of sentences per block, and break at paragraph boundary. It reflects whether a topic or idea is likely to span across paragraphs, and depends on the writing style of different types of text such as research papers or questions. As context block is the resolution unit in building thesauri and classifying text, the change in block length directly influences the association between terms in a thesaurus and the association between concepts in the result. The concept map encodes concept frequency, total and relative co-occurrence through visual information of proximity, brightness, link intensity, and order of appearance. We used Leximancer to visualise concept maps in three stages. First, we visualised the full data set with default setting as illustrated in Figure 2. Second, we pre-processed the dataset by extracting the noun phrases from the dataset. We set Question-Answer as a unit of analysis based on the content component of quadripartite network as illustrated in figure1. Figure 3 captures the important concepts discussed in each thread of discussion as we used Question-Answer as the unit of analysis based on graph theory. Third analysis was on noun phrases used in questions and in answers separately. This compared the transformation between concepts in questions and answers.

Finally, quadripartite graph based clustering was conducted to identify similar questions from the collected dataset. Noun phrases of questions and answerers were used as content while asker id and answerer id were used to represent answers. Similarity measure used to identify similar question was adapted from Blooma et al. (2011). The result obtained is illustrated in Table 2.

RESULTS & DISCUSSION

In the full data set, other than 'weight', and 'sleep' as the main two themes, four themes in the central of the map are 'drug', 'doctor', 'work', and 'take' as shown in Figure 2. On further analysis, in all themes, people use specific units in their replies. The concept 'pounds' specifies weight gain issue ("I don't know what caused the 15 lb"), weight loss objective ("I have lost 15 lbs"), or fitness exercise level ("I do light weights 5lb to 8lbs a lot of reps"). The concept 'hours' specifies the duration of their sleep ("Now I sleep 2hrs"). Such specific numbers suggest that people tend to provide as much details as possible in their replies as if they are in a real conversation with professional medical staff as highlighted by Bowler et al. (2013). This highlights the openness in healthcare SQA services.

In 'weight', the other concepts are 'appetite', 'diet', 'common', 'cause', 'anyone', and 'water' (Figure 2). The concept 'diet' connects with 'eating' in theme 'food', and the concepts 'common' and 'cause' connect with 'effects' in theme 'drug'. Thus, people discuss about the common side effects of drugs, and if these effects cause the weight issues. On the other hand, people also talk about healthy eating diet to counter balance the weight gain effects of drugs. The theme 'doctor' encapsulates three aspects: problems of understanding doctors (concepts 'problem', 'prescribed', 'sure'), supportive replies ('helps', 'best', 'luck'), and redirection to a doctor when verified expert knowledge is needed ('called', 'talk'). The theme 'work' refers to the effectiveness of the drugs ('work', 'try', 'need', 'pain'), seeking alternative drugs ('better', 'bad', 'symptoms'), and how things work ('severe', 'things'). The theme 'drug' relates to the details, effects, and experience of using the drug. An interesting finding is that the concept 'anyone' suggests people are actively seeking relatable experience from other participants, and do not discriminate the source of information. This was also highlighted in other studies such as Bowler et al. (2013) and Culver et al. (1997).

From the whole dataset, there are 704 occurrences of 'who', 2377 'what', 297 'where', 1731 'when', 501 'why', and 1439 'how'. Overall, people discuss facts ('what') about effects and treatments, and the best way to take drugs or treat a condition ('when' and 'how'). The lower frequency of 'where', 'why' and 'who' indicates less attention to the location, reason, or human agent. People mostly use 'who' as a relative pronoun in the context of asking for or giving reference to other people, such as doctors or users similar to the concept 'anyone' ("I would appreciate hearing form anyone who takes or has taken, or who knows anyone who has taken or takes phentermine..."). It reemphasize that people are actively seeking relatable experience from other participants irrespective of professionals or non-professionals (Bowler et al. 2013).

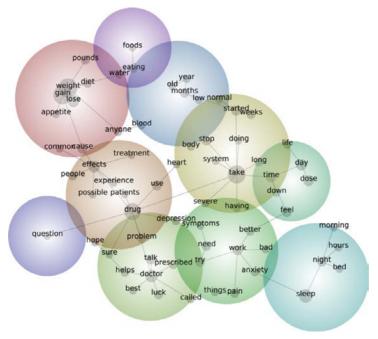


Figure 2: Concept map of original dataset

For further analysis, Figure 3 plots the data set after extracting only noun phrases to gain focus on the important topics discussed and question and answer as a unit. In Figure 3, there are approximately two distinct groups of interconnected themes where there is a dividing boundary between left and right side of the map.

The first group of themes in the left half discusses about 'weight' and its related themes. The major illnesses and conditions associated with weight are diabetes, thyroid, heart and blood pressure, birth, and metabolic syndrome. Some of these weight issues are side effects of drugs ("Weight gain is a common side effect of Geodon"), especially weight gain from medication to type II diabetes. The second group discusses about 'sleep' and its related themes. Sleep issues are associated with withdrawal of drugs due to 'cold turkey' (abrupt halt of medication). Sleep issues are also associated with other disorders and diseases such as depression, insomnia, and fibromyalgia. The symptoms occur in 'head', 'mind', 'muscle', and 'leg'. A common recommendation is exercise.

"Diabetes med usually have the tendency to cause hunger pangs which might increase the weight, you must monitor weight gain as it is not recommended for diabetic patients, but it also could cause weight loss, however I have been on diabetic meds so can tell you all this by experience, which is well over 20 years now. I suggest that you seek medical advice & do some walking at least 3-4 miles 6 days a week. I knocked off around 40kg - about 80-85 pounds of weight in 2 years."

Independent of the 'weight' and 'sleep' issues, people tend to provide emotional support through wishes in their replies such as 'luck' and 'hope', they also provide 'hyperlink' to a better source of information or answers. This agrees with results from Bowler et al. (2013) and Hwang et al. (2010) that people use such site as social and emotional scaffolding, however, the concept 'hyperlink' stands against the conclusion of Bowler et al. that people do not provide answers with credible source. One possible explanation is that the full website in our research has a relevant information archive, thus expert users could link their answers to relevant section in the archive. Yahoo Answer in research of Bowler et al. lacks this readily available source. The following answer to the above quoted question is an example of emotional support:

"we all have been there at one time or another. you got all of our permissions to say how you really feel. so just let it out baby. we all heard them words before in a day or 2 your gonna look at the clock and say "i musta slept for a while, the last time i looked it was an hour ago". so don't you worrie, everythings gonna be alright. ok? keep us informed one of us crazy people will up and online if you wana talk"

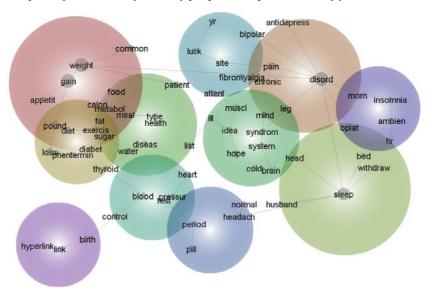


Figure 3: Concept map of dataset using Q&A as a unit

The concept map of dataset using Q&A as a unit (Figure 3) provides a more balanced account of the whole dataset compared to the two maps of all questions (Figure 4) and of all answers (Figure 5). It does not skew towards sleep issues as in question map, or towards weight issues as in answer map. More importantly, the map captures important concepts related to both issues, as well as most of the common concepts. Concepts evolved in questions as illustrated in Figure 4 are about sleep, people indicate how difficult their sleep is through concepts 'morn', 'night', 'bed', and 'hour'. The cause of sleep issues might come from 'pain' in 'muscle', 'leg'. Another cause is dosage of drugs treating Attention Deficit Hyperactivity Disorder (ADHD), depression, anxiety, insomnia, and other disorders. This result agrees with findings of Denecke and Nejdl (2009) that people tend to discuss about drugs and disorders. In question about weight, people associate the issues with effect of drugs for birth control, thyroid, and blood pressure. Common to questions about both issues, people include their syndromes, duration of medication, and ask about effect of drugs ("What is the best anti depressant that doesn't have weight gain as a side effect?"). The two major groups of themes about sleep and weight are on opposite sides, and of the two groups, sleep is the dominant one. The three common themes effect, problem, and time are

in the middle. Thus, common to both conditions, people tend to inquire information about their 'problem' with prescription of 'doctor' and 'syndrome' from taking prescribed 'drugs', and provide information about the 'time' they spent on the drugs and the effect they experienced.

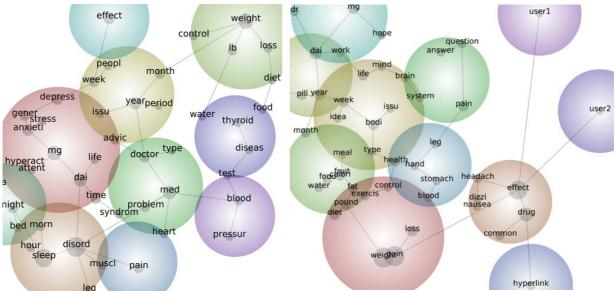


Figure 4: Concept map of all questions

Figure 5: Concept map of all answers

In the answers about weight, people give advice about 'diet', 'meal', 'exercise'. They discuss the specific effect of drugs such as 'dizzi', 'nausea', 'headach', and the affected body parts such as 'stomach', 'leg', 'blood'. The theme 'sleep' is most strongly associated with theme 'mg', this indicates a relationship between the dosage and sleep issues. The two concepts 'user1', and 'user2' represent two very active answerers, however, most of the top active answerers are not present in the map because they do not sign their names in their answers. This indicates an informal and conversational style in their answers. Furthermore, the emergence of specific users as concepts indicates that a relatively small number of users make major contribution to the group. Although this result is from weight and sleep group, it agrees with findings from studies of ex-smokers group (Burri et al. 2006), pain in arm and hand group (Culver et al. 1997) in other websites.

Finally, we applied quadripartite graph-based cluster analysis to identify similar questions based on question-answer-asker-answerer relationship. The sample result is given in Table 2. As similar question lead to similar answers, askers and answerers, we applied the cluster analysis on the data collected. We used the similarity measure (QSim) as proposed by Blooma and Kurian (2012), given in the following equation.

QSim =
$$(1 - \delta)$$
 * (answerSim * α + askerSim * β + answererSim * γ + questionSim * δ (Equation 1) where $\delta = 0.2$, $\alpha = 0.8$, $\beta = 0.1$, and $\gamma = 0.1$

The question content similarity (questionSim), answer (answerSim), asker (askerSim) and answerer (answererSim) similarity are calculated as proposed by Blooma et al. (2011). The similarity is measured with 20% similarity for question content and 80% similarity for question-answer-asker-answerer relationship. Among question question-answer-asker-answerer relationship, the weight of answer similarity is 80%, asker similarity and answerer similarity is 10% each. For an initial analysis on the data collected from healthcare SQA services, we used 70% similarity as the threshold for getting most similar questions. From the 4214 questions, 92 questions were clustered for the 70% threshold.

On further analysis, 23 clusters were found with an average cluster size of four. From Table 2, it was evident that cluster 1 has same asker. The answerer posted similar question and its answer was repeated. The answer even referred to the previous question with the same answer. The question was repeated after two days, which signifies that the asker needed more answers and he kept posting the question. The other answer that he received was recommending him to ask doctor. In general the trend to advice to consult doctor was significantly found in the answers. In one cluster with 28 questions, the question content similarity was much less. However, they were clustered because the answer for all the questions was to refer to doctor. This general answer was misleading to cluster similarity question depending on the general answers. This is evidenced in cluster 2 given in Table 2. Hence, there is a need to emphasis more on the question content and enhance the similarity measure by the use of medical terms. There is also a need to use medical thesaurus to identify similar content as proposed by Zhang and Zhao (2013).

Table 2. Examples of Quadripartite graph-based cluster analysis

Question	Asker	Answer	Answerer
Cluster 1			
Gallbladder Disease - can I take xenical with gallstone? i am 32 years old and i have a gallstone and i want to take xenical to reduce weight	AS1	I don't know how severe your gall bladder condition is it is best to discuss with your doctor whether it is appropriate for you to take Xenical. Xenocal is contraindicated in people who have biliary obstruction. So if your gall bladder disease is severe and causing biliary obstruction then you definitely shouldn't take Xenical.	AN1
Weight - I want to ask you please if one has a gallstone can he take xenical?	AS1	Please refer to the answer to your other question: I don't know how severe your gall bladder condition is it is best to discuss with your doctor whether it is appropriate for you to take Xenical. Xenocal is contraindicated in people who have biliary obstruction. So if your gall bladder disease is severe and causing biliary obstruction then you definitely shouldn't take Xenical.	AN1
Cluster 2			
Does Soma make you gain weight?	AS2	It has not been documented to cause weight gain; would recommend that you talk with the doc/pharmacist who prescribed the med. Take care; be well!	AN2
Does Coumadin cause weight gain?	AS3	It has not been reported to cause weight gain; however if you do experience weight gain it is recommended to speak to the doc/pharmacist who prescribed the med. Take care & be well!	AN2
Cluster 3			
What's the exact mechanism of Metformin's anorexia? I know that it causes anorexia and sometimes used for weight loss. but how!! Thanks	AS4	Metformin is a prescription medication licensed for the treatment of type 2 diabetes. The drug is commonly thought to cause weight loss; although studies have shown mixed results. It is not entirely clear if metformin causes weight loss however; it seems that people taking metformin do not gain weight. It is not known if metformin has any effect on weight in people who do not have diabetes.	AN3
Can Metformin be used with insulin to help control weight brought on by insulin injections?	AS5	Metformin is a prescription medication licensed for the treatment of type 2 diabetes. The drug is commonly thought to cause weight loss; although studies have shown mixed results. It is not entirely clear if metformin causes weight loss however; it seems that people taking metformin do not gain weight.	AN3
Does glucophage help in weight reduction?	AS6	Glucophage (Metformin) is a prescription medication licensed for the treatment of type 2 diabetes. The drug is commonly thought to cause weight loss; although studies have shown mixed results. It is not entirely clear if metformin causes weight loss however; it seems that people taking metformin do not gain weight. It is not known if metformin has any effect on weight in people who do not have diabetes.	AN3

On the other hand, quadripartite graph-based cluster analysis emphasizes not only on identifying similar questions but also verifies the assumption that similar questions are posted by similar users. Cluster 3 is an example of the assumption that similar question leads to similar answers. Askers AS3, AS4 and AS5 posted similar question, and they were answered by the same answerer. This finding will lead to better clustering results after integrating medical thesauri and improving similarity measure.

Thus, based on the findings from concept mapping and cluster analysis, it is evident that there is profound relationship between content and users, using the intricate web of social relationships among questions, answers, associated askers and answerers by applying graph theory. This study also evidence that healthcare SQA services address the society in terms of networking, collaboration and openness. In future, there is a need to measure how

healthcare SQA services address social networking, participation, collaboration, apomediation and openness the five characteristics of Medicine 2.0 as introduced by Eysenbach et al. (2008).

CONCLUSION

This study investigated the concepts discussed in a healthcare SQA service to gain deeper insight into the rich data in healthcare SQA services. We analysed the intricate web of social relationships among questions, answers, associated askers and answerers by applying graph theory, concept mapping and cluster analysis. We visualised concepts using Leximancer and clustered similar questions using quadripartite graph-based cluster analysis. To summarise the findings from this study, people tend to provide as much details as possible in their replies as if they are in a real conversation with professional medical staff. People actively seek relatable experience from other participants, and do not discriminate the source of information. Question-answer-asker- answer relationship leads to similar content as well as users. There is a need to work towards better cluster analysis measures based on the content and user relationship. Furthermore, the concept map reaffirms results of previous studies on the behaviours of users in online medical support groups. Therefore, future research about health data analytic should incorporate patterns of these behaviours in their algorithms to achieve a result closer to real needs of users. Thus, healthcare SQA services are enabling patient empowerment and the better transfer of pertinent information and germane knowledge with the potential end-result being superior healthcare delivery. In many ways, healthcare SQA service has the potential to revolutionize current healthcare delivery practices and/or roles.

REFERENCES

- Arenas, A., Danon, L., D'iaz-Guilera, A., Gleiser, P., and Guimera, R. 2004. "Community analysis in social networks," *European Physical Journal* (38:2), pp 373.
- Balijepally, V., Mangalaraj, G., and Iyengar, K. 2011. "Are We Wielding this Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research," *Journal of the Association for Information Systems* (12:5), pp 375-413.
- Bian, J., Liu, Y., Zhou, D., Agichtein, E., and Zha, H. 2009. "Learning to recognize reliable users and content in social media with coupled mutual reinforcement," in *Proceedings of the 18th International Conference on World Wide Web*, Madrid, Spain, pp 51-60.
- Blooma, M. J., Chua, A. Y. K., and Goh, D. H. 2011. "Quadripartite graph-based clustering of questions," in *Proceedings of the 7th International Conference on Information Technology: New Generations*, Nevada, United States, NY: IEEE Computer Society.
- Blooma, M. J., and Kurian, J. C. 2012. "Clustering Similar Questions In Social Question Answering Services," in *Proceedings of PACIS 2012 Pacific Asia Conference on Information Systems*, pp 160.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. 2006. "Complex networks: Structure and dynamics," *Physics Reports* (424:4), pp 175–308.
- Bowler, L., Mattern, E., Jeng, W., Oh, J. S., and He, D. 2013. "I know what you are going through': Answers to informational questions about eating disorders in Yahoo! Answers: A qualitative study," in *ASIS&T 2013 Association for Information Science and Technology Annual Meeting*, Montreal, Quebec, Canada.
- Burri, M., Baujard, V., and Etter, J.-F. 2006. "A qualitative analysis of an Internet discussion forum for recent ex-smokers," *Nicotine and Tobacco Research* (8: Suppl 1), S13-S19.
- Culver, J. D., Gerr, F., and Frumkin, H. 1997. "Medical Information on the Internet," *Journal of General Internal Medicine* (12), pp 466-470.
- Denecke, K., and Nejdl, W. 2009. "How valuable is medical social media data? Content analysis of the medical web," *Information Sciences* (179), pp 1870-1880.
- Denecke K, and Stewart, A. 2011. "Learning from Medical Social Media Data: Current State and Future Challenges," in *Social Computing in Higher Learning Environments: How Media, Tools, and Platforms Change Learner's Behaviour*, White B, King I and Tsang P (eds), Springer Verlag, Heidelberg.
- Goeuriot, L., Na, J. C., Kyaing, W. Y. M., and Khoo, C. 2012. "Sentiment Lexicons for Health-related Opinion Mining," in *Proceedings of ACM SIGHIT International Health Informatics Symposium*, Miami, Florida, USA
- Hepworth, N., and Paxton, S. J. 2007. "Pathways to Help-Seeking in Bulimia Nervosa and Binge Eating Problems: A Concept Mapping Approach," *International Journal of Eating Disorders* (40), pp 493-504.

- Hewett, D. G., Watson, B. M., Gallois, C., Ward, M., and Leggett, B. A. 2009. "Intergroup communication between hospital doctors: Implications for quality of patient care," *Social Science and Medicine* (69), pp 1732-1740.
- Hwang, K. O., Ottenbacher, A. J., Green, A. P., Cannon-Diehl, M. R., Richardson, O., Bernstam, E. V., and Thomas, E. J. 2010. "Social support in an Internet weight loss community," *International Journal of Medical Informatics* (79), pp 5-13.
- Joo, K. H., and Lee, W. S. 2005. "An incremental document clustering for the large document database," in *Information Retrieval Technology*, Hou, Y., Nie, J.-Y., Sun, L., Wang, B., and Zhang, P. (Eds.), Springer, Berlin, Heidelberg, pp 374-387.
- Lambiotte, R. and Ausloos, M. 2005. "Uncovering collective listening habits and music genres in bipartite networks," *Physics Review E.* (72:6), pp 66-107.
- Lambiotte, R. and Ausloos, M. 2006. "Collaborative tagging as a tripartite network," in *Computational Science ICCS 2006*, Alexandrov, V. N., Albada, G. D., Sloot, P. M. A., and Dongarra, J. (Eds.), Reading, United Kingdom. LNCS 3993, pp 1114–1117.
- Newman, M. E. J. 2002. "Assortative mixing in networks," *Physics Review Letter* (89:20), pp 208.
- Novak, J.D., and Gowan, D.B., 1984. Learning How to Learn, Cambridge University Press, Cambridge.
- Rozenkranz, N., Eckhardt, A., and Rosenkranz, C. 2013. "Health Information on the Internet: State of the Art and Analysis," *Business and Information Systems Engineering* (4), pp 259-274.
- Smith, A. E. 2003. "Automatic Extraction of Semantic Networks from Text using Leximancer," in *HLT-NAACL* 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: Companion volume, Edmonton: ACL.
- Smith, A. E., and Humphreys, M. S. 2006. "Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping," *Behavior Research Methods* (38:2), pp 262-279.
- Stockwell, P., Colomb, R. M., Smith, A. E, and Wiles, J. 2009. "Use of an automatic content analysis tool: A technique for seeing both local and global scope," *International Journal of Human-Computer Studies* (67), pp 424-436.
- Sun, H. 2012. "Understanding user revisions when using information system features: adaptive system use and triggers," *MIS Quarterly* (36:2), pp 453-478.
- Tang, X. and Yang, C. C. 2012. "Ranking user influence in healthcare social media," *ACM Transactions on Intelligent Systems and Technology* (3:4).
- Watson, M., Smith, A., and Watter, S. 2005. "Leximancer Concept Mapping of Patient Case Studies," in *Knowledge-based intelligent information and engineering systems*, Springer Berlin Heidelberg, pp 1232-1238.
- Yang, C. C., Yang, H., Jiang, L., and Zhang, M. 2012. "Social Media Mining for Drug Safety Signal Detection," in *Proceedings of ACM CIKM International Workshop on Smart Health and Wellbeing*, Maui, Hawaii, USA, 2012.
- Zhang, J., and Zhao, Y. 2013. "A user term visualization analysis based on a social question and answer log," *Information Processing and Management* (49:5), pp 1019–1048.

ACKNOWLEDGEMENTS

We are grateful to the two anonymous reviewers who gave invaluable feedbacks on our paper. We would like to acknowledge the RMIT Vietnam Research Grant 2014 No.2.

COPYRIGHT

Mohan John Blooma, Tran Duc Huy, and Nilmini Wickramasinghe © 2014. The authors assign to ACIS and educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to ACIS to publish this document in full in the Conference Papers and Proceedings. Those documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.