INFORMATION PASSING WITHIN SOCIAL NETWORKS

IN HEALTH

BORA GÖKNAR

B.S., Management Information Systems Işık University, 2011

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science
in
Information Technologies

IŞIK UNIVERSITY
2013

IŞIK UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

INFORMATION PASSING WITHIN SOCIAL NETWORKS IN HEALTH

BORA GÖKNAR

APPROVED BY:

Assist. Prof. Dr. Gülay ÜNEL _____
(Thesis Supervisor)

Dr. Cem ÜLGER _____

(Thesis Co-Supervisor)

Assist. Prof. Dr. Sevinç RENDE _____

Assist. Prof. Dr. Cüneyt SEVGİ _____

Assist. Prof. Dr. Ali İNAN _____

APPROVAL DATE:

INFORMATION PASSING WITHIN SOCIAL NETWORKS IN HEALTH

Abstract

In recent years, online social networks have become very popular with the vast usage of Internet connected computers, tablets and mobile devices. Social network is a social structure including individuals or organizations and different types of relationships such as friendship, common interests, work, city, etc. The main factor that separates online social networks and social networks is: in the online case, there is no face to face interaction between the nodes of the network. Since the 1970s, the role of social networks in terms of illness, distress, disease and disability; and in understanding how people identify and respond to these crises, has become a mainstay of social and behavioral science research [3]. In this thesis, we investigated trust relationships and information passing in online social networks that focus on health. There is a significant amount of related work in this area conducted on consumer choice prediction, consumer influence factors and information on eBay, Amazon and other early adapter big e-commerce Web sites all over the world. In this thesis, we particularly focused on the content of a Web site called doktorsitesi.com. This is the largest online social network in Turkey related to health that is managed by professionals. We collected data from this domain and investigated how the work on information passing in social networks can be used in this domain.

SAĞLIK SEKTÖRÜNE AİT ONLINE SOSYAL AĞLARDAKİ BİLGİ AKTARIMI

Özet

Son yıllarda online sosyal ağlar, Internet bağlı bilgisayarlar, tablet ve mobil cihazların yaygın kullanımı ile oldukça popüler olmuştur. Sosyal ağ, bireylerin ve organizasyonlarin içinde bulunduğu sosyal bir yapıdır ve arkadaşlık, ortak ilgi alanları, aynı şehirde yaşamak aynı mesleğe sahip olmak gibi farklı ilişki türlerine sahiptir. Online sosyal ağın herhangi bir sosyal ağdan temel farkı, Online bir sosyal ağda iseniz diğer bireylerle yüzyüze iletişiminizin bulunmamasıdır. 1970lerden bu yana hastalık, sağlık problemleri, sakatlık veya salgın içerikli sosyal ağların görevi kullanıcılar arasında bir dayanak noktası oluşturmak ve birbirleri ile paylaşım yaparak sağlık krizlerini atlatmalarında yardımcı olacak bir alan yaratmak olarak açıklanabilir [3]. Bu tezde sağlık sektörü odaklı Online sosyal ağlardaki bilgi aktarımını ve güven ilişkisini inceledik. Bu alanda kullanılabilecek başta eBay, Amazon ve benzeri büyük elektronik ticaret sitelerinde yapılan ilgili tüketici tercihleri tahmini, tüketicileri etkileyen faktörler gibi önemli araştırmalar bulunmaktadır. Bu tezde özellikle doktorsitesi.com adlı Web sitesindeki içeriğe odaklandık. Profesyoneller tarafından yönetilen ve kullanılan bu site sağlık sektöründe Turkiye'de bulunan en geniş çaplı online sosyal ağdır. Bu çalışmada ilgili siteden almış olduğumuz verileri kullanarak sosyal ağlardaki bilgi aktarımı ilişkilerini bu alana uyarlamayı amaçladık.

Acknowledgements

I am sincerely grateful to my supervisors Assist. Prof. Dr. Gülay Ünel, and Dr. Cem Ülger for their immense knowledge, encouragement and support during my graduate studies and thesis writing. I also acknowledge the support from the doktorsitesi.com by Dr. Erden Asena.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CRO	Contract Research Organizations
OSN	Online Social Network
IDC	International Data Corporation
ALS	Amyotrophic Lateral Sclerosis

Chapter 1

Introduction

The emergence of online social networks (OSNs) in the last years has led to a huge increase in the volume of information about individuals, their activities, connections amongst individuals or groups, and their opinions and thoughts [13].

Social network is a social structure made of individuals (or organizations) that represent "nodes" and they are associated with one or more types of relationships and interdependencies, such as friendship, common interests, work, knowledge, and prestige [2]. Online social networks are more complicated than social networks as a result of their network structure [2]. The primary difference between these networks is: "Social networks are based on personal physical connectivity (the citizens' associations, transplant networks, etc.), whereas Online Social Networks or Global Internet social networks (such as Facebook, Twitter, Skype) do not necessarily require personal or physical connectivity" [2].

According to IDC (International Data Corporation) Report [6] Online social networks become very popular in recent years as a result of the increased usage of the internet enabled devices such as laptops, mobile devices, cell phones and tablets. Worldwide media tablet shipments for 2010, 2011 and 2012 are 19.4 million devices in 2010, 68.7 million devices in 2011 and 106.1 million devices in 2012 [6] [7].

The people that use Facebook, Twitter and LinkedIn have a significant power on the usage of online social networks. Twitter has been accessed 50,000 times per day and has over 200 million users [2]. In March 2013, Facebook had 1.3 billion users who belong to the unique cyber culture. In December 2012, Facebook had 974 million users where more than 275 million users were from Asia [8], LinkedIn had 156 million users and 73 million of them from USA [9] and Twitter had 140 million users and 340 million tweets (Twitter posts) per day [10].

As a result, investors started to create online social networks for specialized purposes according to their needs. One of these specialized areas that use online social network structures is health care. As stated in [4], “Social networks hold considerable potential value for health care organizations because they can be used to reach stakeholders, aggregate information and leverage collaboration”. Various health related social networks include Health Care Forums, and Healthcare Industry Forums whereas medical social networks for non-professionals include DailyStrength, CaringBridge, CarePages, and MyFamilyHealth. There are also social networks supported by professionals such as HealthBoards, Spas and Hope Association of Disabled and diabetic Enurgi [2]. According to [4], “Similar to their growing importance in other industries, social networks are playing an increasingly prominent role in health care. %55 of surveyed Americans gets information about a therapy or condition online”. Hence, social networks have become a stronghold of scientific research in the social behavior in case of large outbreaks of infectious diseases, disability, accidents, and in understanding how people react to crisis situations [2].

In this thesis, we focus on a domain called doktorsitesi.com, which is the largest Turkish online social network in health care domain. This web site connects the patients and the doctors in a social network so that the professionals and the patients can easily contact each other to ask questions about health and to get appointments from the doctors. The primary process we investigate is called information passing. In our study, information passing analyzes the likelihood of a patient to contact a doctor given that s/he is messaged by a friend who contacted the doctor previously.

Remainder of this thesis is organized as follows. In the rest of this chapter we provide general information about health related social networks. In Chapter 2, the related work from the literature for the technical presentation is presented. In Chapter 3, study and design subjects, study variables, and the thesis methodology are presented. In Chapter 4, the analysis and the results of the thesis are explained, and finally the thesis is concluded in Chapter 5.

1.1 Social Networks in Improvement of Health Care

Online community is a public framework made of individuals (or organizations) that symbolizes “nodes” and they are associated with one or more kinds of interdependency, such as relationship, common passions, work, information, reputation, and many other interests. System concept was designed in the seventies targeted at knowing individual actions through public interaction [2]. Since then, public networking sites have become a stronghold of medical research in the public actions in case of large episodes of contagious illnesses, impairment, injuries, and in understanding how people respond to emergencies. Following the previous claims the public networking sites can be separated into several groups, based on relationship methods, field of functions or skills of those who get involved in specific networks [2]:

- Social networks with personal physical connectivity (the citizens’ associations, transplant networks, etc.).
- Global Internet social networks (Facebook, Twitter, Skype)
- Specific Internet health related social network (forums, Health Care Forums, Healthcare Industry Forum).
- Medical social internet network for non-professionals (DailyStrength, CaringBridge, CarePages, MyFamilyHealth).
- Scientific internet social network (BiomedExperts, ResearchGate, iMedExchange).
- Social Internet networks supported by professionals (HealthBoards, Spas and Hope Association of Disabled and diabetic Enurgi).
- Scientific networks in the world’s biomedical literature databases (Current Contents, ISI Web knowledge, PubMed/Medline, ExcerptaMedica/EMBASE, EBSCO, Index Copernicus, etc.).

These public media sites are often non-governmental companies, but rarely can be organized systems of government institutions, such as a national transplant system or the Eurotransplant system. As they have the support of expert companies in healthcare, details published is reliable. Facebook or Myspace, Tweets, MySpace and Sparkle are four famous and popular public media sites. New systems continue to

form, although at the web market there are already too many. Tweets have been utilized 50,000 times per day and have over 200 thousand customers. In Jan 2011, Facebook or Myspace had 600 thousand customers who are part of the unique cyber culture. Social networking today is what the World Wide Web was 20 years ago [2].

In addition to the general social networks there are also such sites for enabling communication, cooperation and details gathering in the field of medical care. More than half (55%) of people in America who investigate wellness issues online get details about the therapy or condition via the Internet, and one third use public media sites. About 60% of doctors in U.S. use public media sites for expert purposes. Using public media sites in the health sector enables: increase in communication and cooperation, where patients exchange information about similar problems, and professionals can share experiences on care or treatment, and as a result, make better health decisions [2]. Using social networks in health care enables the following opportunities:

- Patients can provide reviews about their disease and therapy to their doctors and locate treatments in different medical centers. Such reviews can affect the recommendation of patients to other levels of medical care or other vital processes in making decisions [2].
- Global Internet public social networking sites enable the exchange of details without time and space restrictions. This improves the number of potential users of health services [2].
- Social networking can be used to advertise and improve work ability and knowledge of stakeholders.

In addition to these, as stated in [3], "the role of personal and business relationships form a solid place in our understanding of how the social world impacts and is impacted by healthcare and wellness phenomena across a variety of contemporary socio-medical research concentrating on the epidemiology of wellness issues, the organization and use of official medical care services, and the diffusion of healthcare technology".

An example of social Internet networks supported by professionals is WebMD. Through collaboration with an experienced website, WebMD provides an advanced

client care. The group has about 730,000 associates, with 767,886 subjects and prepared 4,418,155 feedbacks until May 2011 [2].

1.2 Business Use of Social Networks in Health Care

Internet-based public social networks can enable communication, collaboration and information collection and sharing in the health care space. About one-third of Americans who go online to investigate their health currently use social networks to find fellow patients and discuss their conditions and 36 percent of social network users evaluate and leverage other consumers' knowledge before making health care related decisions [4].

Public networking sites hold considerable potential value for medical care organizations because they can be used to reach stakeholders, combined information and make use of cooperation. Public networking sites can also transfer media such as video, web records (blogs), ratings and reviews, podcasts and audio among people who are linked by common attributes, such as loved ones, profession, school, residence and even likes and dislikes [4].

A survey is conducted on the use of social networks in heath care and it is found that %55 of surveyed Americans get information about a therapy or condition online. In addition, "%60 of surveyed physicians are interested in using social networks for professional purposes, approximately one out of every six U.S. physicians are members of Sermo, an online physicians network and %65 of surveyed nurses indicate they are planning to use social networks for professional purposes" [4]. Since both consumers and clinicians are using social networks, health care organizations have an opportunity to leverage their influence across multiple audiences [4].

There are various social networks in health. For example, the website PatientsLikeMe gives consumers a way to track disease progress, access disease information and learn from the real-world experiences of other patients with the same medical condition and to share their findings with patients, health care professionals and industry organizations that are trying to treat the disease. Other consumer-directed sites include MedHelp which, in addition to being a social

network, offers a number of tracking tools for pain, weight and other chronic conditions; CureTogether, which helps people anonymously track and compare health data to better understand their bodies, make more informed treatment decisions and contribute data to research; DailyStrength, which allows patients and caregivers to give and receive support; Inspire, which hosts different communities, some of which are co-sponsored by non-profit foundations, to educate and offer support; and FacetoFace Health, a social network that uses a proprietary algorithm to match people with similar diagnoses [4]. Industry organizations have an opportunity to interact with the members of online health communities and to gather real world data sets to inform new treatments and care pathways as more patients use social networks to track their health conditions and care [4].

1.3 Social Networking Applications in Health Care

Table 1.1 provides examples of various social network applications in the health care area [4].

Table 1.1. Social Networking Applications in Health Care [4]

Health care application	**Participants**	**Impacted organizations**	**Representative social networking applications**
Maintaining health and wellness	Consumers Health coaches	Physicians Health plans Wellness facilities Hospitals Alternative providers/health coaches Employers	WEGO Health offers health activists the chance to advise consumers about their health by reviewing and linking tools on one site. WebMD offers a social networking platform for both consumer- and physician/expert-moderated health conversations.
Disease management	Consumers Physicians Allied health professionals	Physicians Retail clinics Health plans	PatientsLikeMe offers a 24/7 secure login for health plans to enable disease management. Inspire.com offers patients 24/7 access to peer communities.

		Device manufacturers Drug companies Alternative care providers Disease management companies	
Clinical trial recruitment	Consumers Clinical investigators	Academic medicine Drug & biotech companies Contract Research Organizations (CROs) Device manufacturers	Amyotrophic lateral sclerosis (ALS) patients organized their own clinical trial on PatientsLikeMe. Novartis recruited patients from PatientsLikeMe for its FTY720 multiple sclerosis trials. iGUARD, a wholly owned subsidiary of Quintiles, is a drug review site that leverages its network of participants to model clinical trial feasibility.
Personal Health Records (PHRs)	Consumers Health professionals	Drug & device manufacturers CROs Academic medicine Health plans	PatientsLikeMe and MedHelp allow participants to upload detailed information about their condition and receive information from similar patients.
Health professional training	Physicians Advanced practice nurses Allied health professionals	Drug & device manufacturers Licensing organizations Hospitals Schools	radRounds offers radiology professionals a community to collaborate and share cases and receive opinions. Sermo and Ozmosis offer physicians the opportunity to submit cases for community discussion; they fund their sites with manufacturer sponsorships and advertisements.
Public health announcements and campaigns	Consumers Regulators	Regulatory agencies Public health	The CDC collaborated with DailyStrength and Sermo to inform the public and physicians

		agencies Local/state/federal government	about flu outbreaks. In addition, the CDC monitored the conversations and addressed the public on a variety of social networking sites, including Facebook, MySpace, CafeMom, etc.
Treatment, physician or hospital selection	Consumers	Drug & device manufacturers Hospitals Health plans Retail clinics	DailyStrength and FacetoFace Health offer communities and a search engine so that patients can find "matches" to ask advice. The Mayo Clinic offers the "Sharing Mayo Clinic" blog for patients and physicians to share their stories; it also offers the "Medical Edge" for patient information podcasts and news.

1.4 Challenges in Social Networks About Health Care

Health care stakeholders are using social networks on a more widespread and frequent basis. However, the networks' commercial value to specific health care organizations and the industry as a whole remains elusive. Many businesses have not yet developed robust metrics to measure success and return on investment related to social networks as in Internet advertisements and traditional marketing tactics [4].

Challenges in social networks about health care include risk, liability and consumers trust. Risk and liability: Communications between medical service providers and their patients are subject to certain acts imposed by the governments, which limit plans, medical centers and doctors from answering questions on a specific individual's health information across online community [4]. Consumer Trust: The online community location contains a considerable amount of false information, which might offer itself to customer trust issues if key health care industry categories are thought to be "manipulating" the free circulation of information [4].

1.5 Potential Social Network Use by Health Care Stakeholders

Table 1.2 provides examples of potential social network use by health care stakeholders.

Table 1.2. Potential Social Network Use by Health Care Stakeholders [4]

Organization	**Implications/ future directions**
Professional societies	Online networks may replace health care professional societies. For example, Sermo recently represented U.S. physicians as a contributor to the federal Health Care Reform debate.
Nonprofit foundations	Foundations may leverage recreational and health care sites to educate their stakeholders and to mobilize them for advocacy and activism.
Public health organizations	Social networks may enable public health organizations to quickly reach the public and alert them to policy debates and key decisions such as vaccination drives, quarantines and evacuations.
Hospitals	Hospitals may use social media to engage in crisis management, help consumers understand their treatment options and obtain feedback on their services.
Life sciences companies	Pharma and biotechnology companies – may increasingly mine physician data from physician networks. Medical device companies – may increasingly use professional networks to train physicians remotely on their products. CROs – may build their own social networks to help with clinical trial feasibility and recruitment.
Federal regulators	Given the life sciences' industry's interest in social networks, the FDA may eventually set up pharmacovigilance channels through online social networks.
Health plans	Health plans may be able to increase member marketing and education efforts via social networks.

1.6 Health Information Resources

The Internet is an important source of information for most consumers, which can be also demonstrated by the following facts [15]:

- In 2010, more than half of customers (55%) said that they looked for details about treatments on the Internet.
- 1 in 4 reports that they searched online for physician quality-of-care information and 12% used the Internet to find information on provider costs, similar to 27% and 13%, respectively, who reported doing so in 2009.
- 1 in 10 consumers uses web sites to compare hospital treatment options.
- A small percentage of customers (5%) use social media websites to look for details about prescription medications, connect with a doctor (3%) and connect with their health plan (3%).
- Almost half of the consumers believe in safety and efficiency details from healthcare organizations (45%) and educational healthcare facilities (41%); whereas 10% or less believe in wellness plans, companies and producers.
- 35% of consumers trust cost information from healthcare organizations and 32% of them trust educational healthcare facilities; whereas 15% or less trust employers and manufacturers, and 1 in 3 distrusts these sources.

Another important study is presented in Table 1.3 on independent health-related web sites' (e.g., WebMD) information reliability. This study includes nationally representative sample of 4,008 American adults, aged 18 and older who were surveyed between December 28, 2009 and January 5, 2010, using a web-based questionnaire.

Table 1.3. Reliable Information on Online Social Networks about Health [15]

Year	**Trust / Distrust**	**Percentage**
2009	Trust	28%
2009	Distrust	12%
2010	Trust	22%
2010	Distrust	14%

1.7 Use of Technologies to Improve Information Access in Health Care

Following information about the use of technologies in health care is provided in the 2011 Survey of Health Care Consumers Global Report Key Findings, Strategic Implications [16]:

- Interest in using a smart phone or PDA to monitor their health if they are able to access medical records and download information about their medical condition and treatments, varies by generation and country. Consumers in Brazil (43%), China (47%), and Mexico (50%) are very likely to use their phones for health care monitoring. Consumers in France (65%), Germany (64%), and Belgium (63%) are not at all likely to use their phones or PDAs for this purpose.
- Use of social media sites for wellness proper care purposes was consistently low ranging from 4% in Italy to 32% in South America. Chinese suppliers were the exemption with 63%. Use of social media was primarily for sharing personal wellness proper care experiences or for seeking details on pharmaceutical products.
- Less than 1 in 5 consumers maintain a personal wellness record (PHR), excluding customers in Chinese suppliers, where around 1 in 3 have such a record.
- Around half of consumers look online for treatment information. The practice ranges from 26% in Belgium to 54% in Canada.
- Consumers are highly interested in using a medical device that would enable them to check their condition and send information to their doctor electronically through a computer or cell phone via the Internet. Interest ranges from 46% of consumers in Belgium to 79% of consumers in Mexico.

1.8 Content Analysis of Breast Cancer Groups on Facebook

Online communities are convenient venues for exchanging information and support with people in similar circumstances and are increasingly being used for health purposes, particularly by breast cancer survivors [5]. One of the most popular online communities is the social network site Facebook (www.Facebook.com). According to the study in [5] “While young adults are still more likely to use social network sites, the fastest growing demographic of Facebook users is women 55 years and older, which corresponds to the average age of onset of breast cancer”. Although recent studies indicate that Facebook groups are used for health purposes, there is little information about how this resource is used by people affected by breast cancer

[5]. The study presented in [5] aimed to characterize the purpose, use, and creators of Facebook groups related to breast cancer.

As a result of the study [5] it is pointed out that “Facebook groups have become a popular tool for awareness-raising, fundraising, and support-seeking related to breast cancer attracting over one million users. Given their popularity and reach, further research is warranted to explore the implications of social network sites as a health resource across various health conditions, cultures, ages, and socioeconomic groups”.

Chapter 2

Related Work

In this part we provide the related work that is most relevant to our methods and analysis. We first introduce the general research topics in the social networks area in Section 2.1. In Section 2.2 we explain the role of social networks in online shopping, and finally we explain the methods on measuring trust in online social networks that are relevant to our problem in Section 2.3.

2.1 Research Topics in Social Networks

The field social networking has seen rapid improvements recently. A key aspect of many of the public networking sites is that they are rich in information, and provide novel difficulties and opportunities from the viewpoint of knowledge development and information exploration. There are two primary kinds of data, which are often examined in the viewpoint of social networks [11]:

- Linkage-based and Structural Analysis: "In linkage-based and structural analysis, we construct an analysis of the linkage behavior of the network in order to determine important nodes, communities, links, and evolving regions of the network. Such analysis provides a good overview of the global evolution behavior of the underlying network" [11].
- Adding Content-based Analysis: "Many social networks such as Flickr, Message Networks, and Youtube contain a tremendous amount of content which can be leveraged in order to improve the quality of the analysis. For example, a photograph-sharing site such as Flickr contains a tremendous amount of text and image information in the form of user tags and images. Similarly, blog networks email networks and message boards contain text content that are linked to one another. It has been observed that combining content-based analysis with linkage-based analysis provides more effective

results in a wide variety of applications. For example, communities that are designed with text-content are much richer in terms of carrying information about the topical expertise of the underlying community" [11].

Structural research of social networking sites includes methods from mathematical research of systems, community recognition, category, evolution research, privacy-preserving data mining, link inference and creation [11].

The following Table 2.1 [12] presents the main research problems of online social networks.

Table 2.1. Deep Research Questions on Online Social Networks

Question	**Related Questions**
What do social networks look like, on a large scale?	Do most nodes have few connections, with several "hubs" or is the distribution more stable? What sort of clustering behavior occurs?
How do networks behave over time?	Does the structure vary as the network grows? In what fashion do new entities enter a network? Does the network retain certain graph properties as it grows and evolves? Does the graph undergo a "phase transition", in which its behavior suddenly changes?
How do the non-giant weakly connected components behave over time?	One might argue that they grow, as new nodes are being added; and their size would probably remain a fixed fraction of the size of the GCC. Someone else might counter-argue that they shrink, and they eventually get absorbed into the GCC. What is happening, in real graphs?

What distributions and patterns do weighted graphs maintain?	How does the distribution of weights change over time– do we also observe a densification of weights as well as single-edges? How does the distribution of weights relate to the degree distribution? Is the addition of weight burst over time, or is it uniform?

The social sciences identify two important phenomena [13] that can apply to online social networks, which are homophily, and co-citation frequency.

According to [13], "Homophily, also known informally as "birds of a feather", is when a weblink between people (such as relationship or other public connection) is associated with those people being identical in characteristics. For example, buddies often are generally identical in features such as age, public qualifications, and knowledge level". Homophily is not investigated in our case because of the lack of information.

Co-citation regularity is a concept, which holds when similar individuals tend to refer or connect to the same entities [13]. For example, when two people have identical preferences in songs, literary works or fashion, co-citation frequency indicates that they may be identical in other ways or have other common interests [13]. In our case, the only mutual point between the patients is their health situation.

2.2 Role of Social Networks in Online Shopping

The main subjects that [1] focus on are information passing, the price of trust, and consumer choice prediction. Suppose that an individual purchases a product and then messages a friend. In [1], it is claimed that the information passing provides insight about the likelihood that the friend will then purchase the product and about its purchase location. Influence of information passing is directly proportional to message strength, and is inversely proportional to product price, as well as the time between the purchase and the recommendation [1].

Having demonstrated the existence of information passing in the Taobao network, the question that is investigated in [1] is, “What factors influence the success rate of information passing?” How information passing success varies with respect to 4 variables is examined in the study, which are: communication strength, time difference, product price, and product category. These variables can also be mapped to our work. We can use the types of interactions between the users of our network, which are, ask question, thank, follow and get an appointment to measure the communication strength. For investigating the time difference, we can check the timeline of the events that occur between a patient and a doctor. There are no variables we can directly map for product price in our domain but we can use the entities of the doctors such as, videos, answered questions or articles for interpreting the price. Product category can be converted to our domain as the department of the hospital.

In the following table we provide various questions from the research conducted by Guo [1] and adopt these questions to our case.

Table 2.2. Questions about Online Shopping and Doktorsitesi.com

Taobao Questions	**Doktorsitesi Questions**
How do friends influence consumer purchasing decisions and product adoption?	How do doctors influence patients to ask themselves questions?
What factors influence the success of word of-mouth product recommendations?	What factors influence the patients to choose the correct doctor?
How does social influence and reputation affect commercial activity?	How does social influence and reputation affect patient’s choice?
What is the likelihood that the friend will then purchase the product?	Is there any relationship between the patients to decide a doctor to ask question?
Where will s/he purchase it?	-

2.3 How to Calculate Trust in Online Social Networks

In TaboBao [1] case, the main focus was the information passing in buyer-buyer relationships. To analyze trust, the relationships between buyer-sellers were

examined where trust is encapsulated by seller ratings. As a result, price of trust was defined, as the extra amount a buyer is willing to pay for transaction with a highly rated seller.

The result of this work stated that higher rated sellers are able to sell their products with higher prices because sellers think that highly rated sellers may provide better services, such as replying to messages from customers in a timely fashion, or shipping products more frequently. In other words, buyers are willing to pay more to highly rated sellers to minimize transaction risk, thus sellers who maintain good reputations are financially rewarded [1].

Al-Oufi [18] defined trust in online social network environment as follows: “Trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome”. In addition, they adduced that trust in online social networks has three primary characteristics: transitivity, asymmetry, and personalization [18].

Additionally, in [19], the focus is moved toward time-aware trust prediction in evolving online trust networks. In this work the impact of considering the temporary progress of trust networks was examined explicitly in trust prediction tasks by using a supervised learning method [19]. Web site called epinions.com was used that mostly includes the end user reviews and a combination of the Web sites where online users can buy products. The social trust factors in online environment are listed in Table 2.3 below.

Table 2.3. Social Trust Factors in Online Environment [19]

Trust Factor	**Definition**
Knowledge Factor	Refers to the trust building mechanism where individuals get to know each other through interactions and then predict others behaviors based on the information they obtain from this interactive process.
Relationship Factor	Refers to the trust building mechanism, which relies on qualitative assessments based on connections found in social networks and online communities.

Reputation Factor	Refers to the trust building mechanism in which trustee behavior in the whole system affect the amount of his trustworthiness.
Similarity Factor	Refers to a trust building mechanism, which implies that trust is established based on social similarities such as common characteristics the trust or perceives of the trustee including interests, values, and demographic traits which can lead to establish a new trust relationship between two sufficiently similar users.
Personality Factor	Refers to users' individual traits that lead to expectations about the ones' trustworthiness.

Finally, [20] looks particularly at assigning trust in web-based social networks and investigates how trust in information can be mined and incorporated into applications.

We primarily focus on information passing in this study. However a trust model can also be built on top of this work based on information passing and interactions between the users of our domain.

Chapter 3

Study Subjects and Methods

3.1 Study and Design Subjects

In this thesis, we selected a domain called doktorsitesi.com, which is the largest online social network about health in Turkey. The purpose of this Web site is to create connections between the patients and the doctors and increase the vision and also knowledge of patients about their illnesses with the help of doctors. As of December 2012 there were 1.360.000 patients and 13.479 doctors who are members of the site. In this site, patients can freely ask questions to the doctors and get answers from them. They can also get in touch with the doctors or dentists. In addition to these, patients can check the articles or the videos of the doctors, share them online, and can send messages to each other.

Online social networks have two primary components, which are nodes that represent the users and edges that represent the relationships. In our case nodes represent the patients and the doctors, edges represent the relationships that we can divide to three groups each having its own characteristics:

- Patient to Patient,
- Doctor to Doctor,
- Patient to Doctor.

The user interface details and the working principles of the Web site are provided below:

Main Screen: After you login you can easily check for the online doctors and immediately send them questions. On the main pane, you can check the last answered questions online.

Header Menu Topics: In the “Doctors” menu there is a search button and also a filter button. You can easily search for the doctor name or you can select your department, city and also hospital to find doctors and ask questions to them. In addition, you can filter the doctors that are available to get an appointment. In the “Questions” menu there are categories so that you can filter your search and see the results starting from the recently answered questions. In the results, the patient who asked question is anonymous except the ID number of the patient. You can use patient ID to contact the patient. In the “Videos” menu you can search for the videos according to its category and doctors. The “Medical Center” menu only includes videos about certain categories. After you search for a category, a related video appears immediately. In the “Articles” menu you can search for the articles according to its category and doctors. You can share a video or an article online using social networking sites such as Facebook, Twitter, Google+, etc.

3.2 Study Variables

Below you can find the nodes of the doktorsitesi.com, which we can separate into two different groups: patients and doctors.

Patients: If you login as patient you can see 5 different menus. First one is called “My page”, which is the main page of your account. The second one is “Questions” where you can ask new questions, check for the asked questions, trace your answered questions and check for your time-out dated questions. The third one is “Messages” where you can check your incoming and outgoing messages. The fourth one is “Connections” where you can select and follow patients or doctors. The last one is “Settings” where you can update your preferences and connection type, and upload photos.

Doctors: Doctors have related properties such as the amount of answered questions, articles and videos. Patients can ask questions to get an appointment with, follow and thank the doctors. The pages for the doctors include “Summary”, “Resume”, “Answered Questions”, “Articles and Videos”, and “Followers”. “Summary” includes brief information about the doctor; “Resume” is the background information about the doctor; “Answered Questions” shows the questions answered by the doctor; “Articles and Videos” includes information and results that the doctor can

share; and “Followers” includes the information about patients that follow the doctor. Figure 3.1 below represents the use case diagram of doktorsitesi.com.

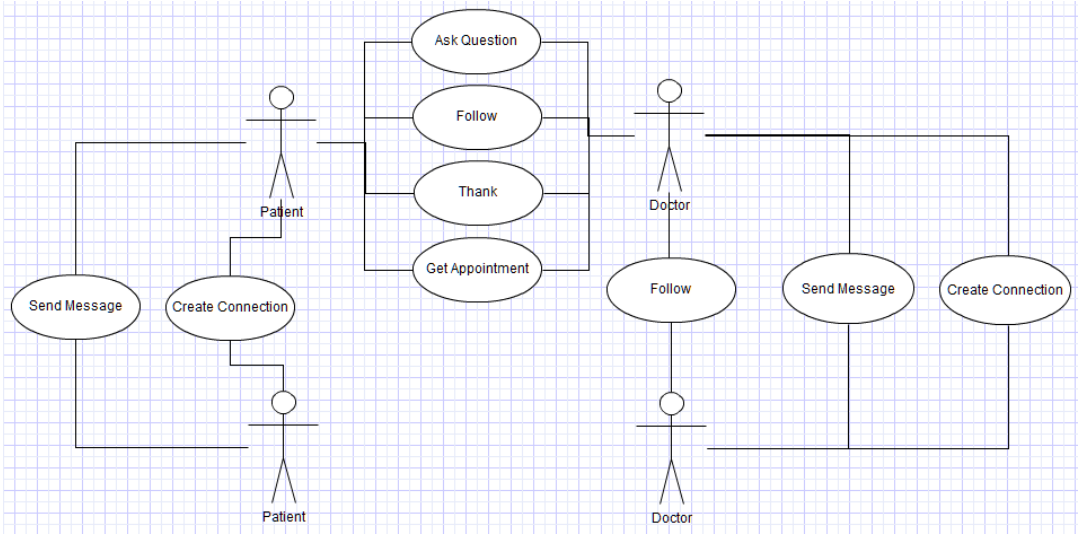

Figure 3.1. Use Case Diagram of Doktorsitesi.com

3.3 Methods

In this work, our aim is to investigate the information passing and trust relationships among the nodes in our network. More formally, if patient P1 asks question to doctor D1 and then send message to another patient P2, will patient P2 then ask question to doctor D1 as well? In the next chapter, we analyze information passing in our domain through the study of Taobao [1].

We conduct our research in 4 steps. Each step is designed in terms of the different actions that can take place between the patients and the doctors to understand and investigate the information passing between the patients. We calculate the information passing success rate among patients for each step. Additionally, we calculate the number of exchanged messages among patients and the time difference between these messages in each step.

Chapter 4

Data Analysis and Results

4.1 Data Analysis

In our analysis, we used data from year 2012, October, November and December. The dataset includes 25.512 unique users who send question(s) to the doctor(s).

In terms of data analysis, we primarily focused on the method used for TaoBao [1] network. Our aim was to investigate the information passing and trust relationships among the nodes in our network. In TaoBao network, the nodes can be buyers and sellers where buyers can share comments about the sellers, and send messages to other buyers. We used the method described in [1] to analyze information passing and trust among patients in our case as illustrated in Figure 4.1. Information passing behavior in this case can be described by the question "if patient P1 asks a question to doctor D1 and then sends a message to patient P2, will patient P2 ask a question to doctor D1 as well?"

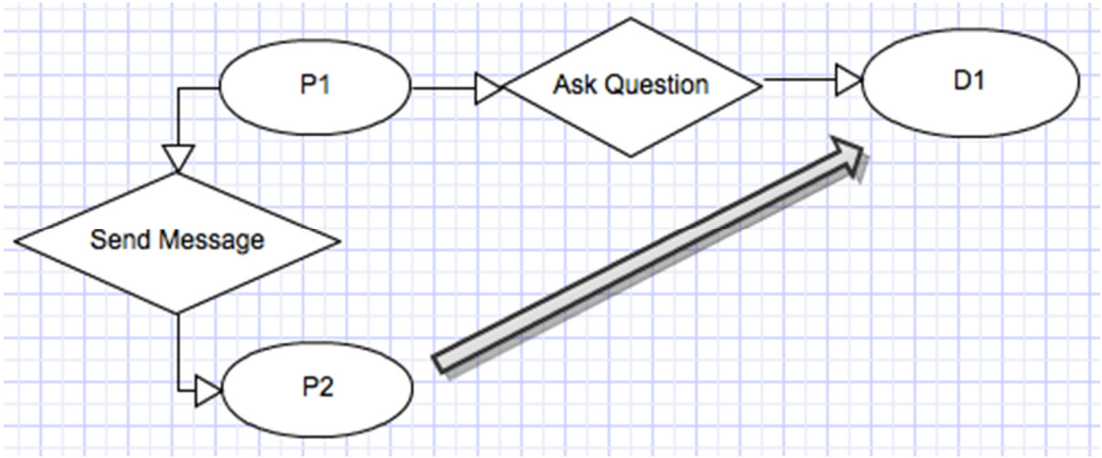

Figure 4.1. An Information Passing Scheme from Doktorsitesi.com

Before we provide the details of our analysis we give statistical details about the doktorsitesi.com and our dataset. In Table 4.1, we provide the number of the users in doktorsitesi.com.

Table 4.1. Total Number of Patients and Doctors on Doktorsitesi.com

Description	**Numbers (December, 2012)**
Patients	1.360.000
Doctors	13.479

In Table 4.2, we provide the general information on the number of questions asked from our dataset.

Table 4.2. Number of Questions Asked on Doktorsitesi.com in Last 3 Months

Description	**Numbers**
Total questions asked in 3 months (10, 11, 12. 2012)	60.690
Patients that asked 3 or more questions in 3 months (10, 11, 12. 2012)	5.282
Patients that asked 5 or more questions in 3 months (10, 11, 12. 2012)	2.312

In Table 4.3, we provide the cardinalities of the entities and relationships in our dataset.

Table 4.3. Cardinalities of Entities and Relationships in Doktorsitesi.com

Description	**Numbers**
Total unique patients	25.512
Total number of messages	7.997
Total number of doctors	2.494
Total number of questions asked	60.690
Total number of answers	46.222
Total count of “follow” actions	3.632
Total count of “thank” actions	6.287
Total count of follow and thank actions (that happen together)	798

Total count of follow and reply actions (that happen together)	3.375
Total count of thank and reply actions (that happen together)	4.963
Total count of follow, thank and reply actions (that happen together)	744

We conducted our analysis in 4 different steps for answering four basic questions:

- Step 1: Given that P1 first asked a question to D1 and received a reply from D1, will P2 ask question to D1?

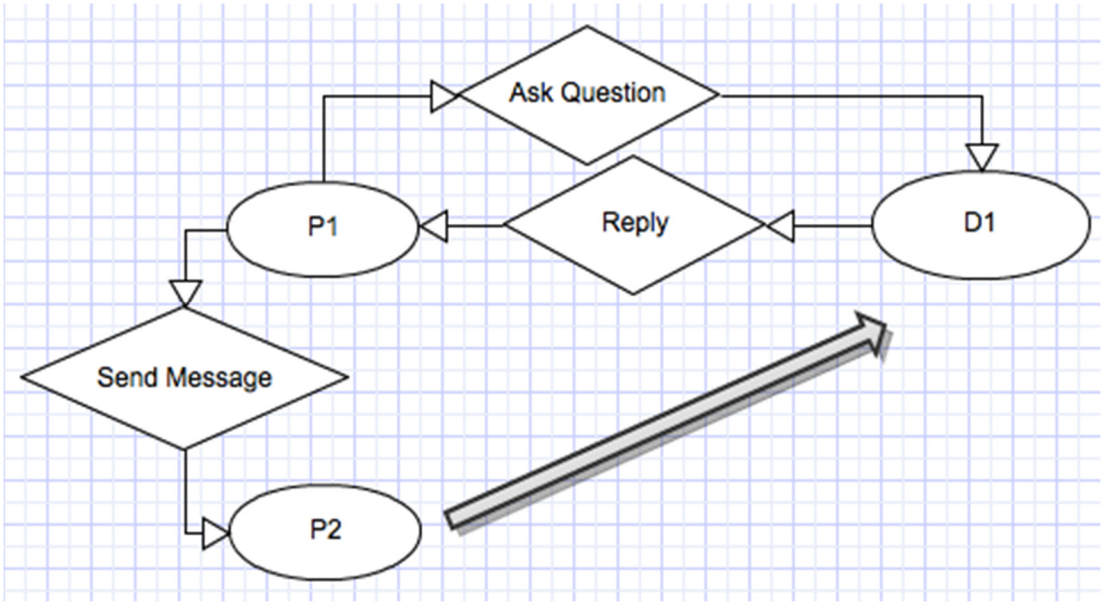

Figure 4.2. Representation of Step 1

- Step 2: Given that D1 is followed by P1, P1 asked a question to D1 and received a reply from D1, will P2 ask question to D1?

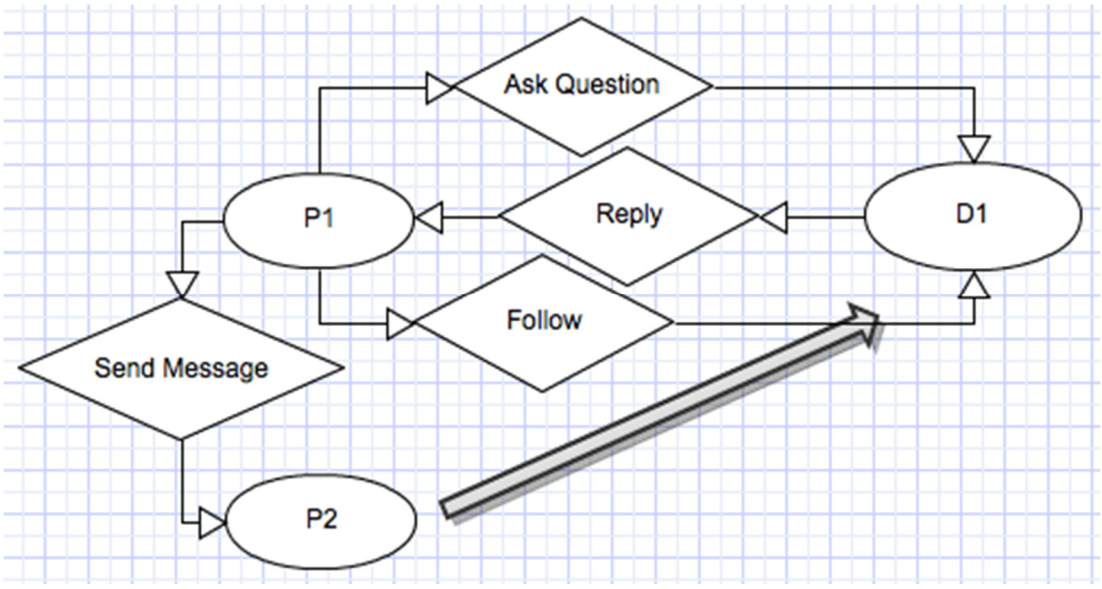

Figure 4.3. Representation of Step 2

- Step 3: Given that P1 asked a question to D1, received a reply from D1 and thanked to D1, will P2 ask question to D1?

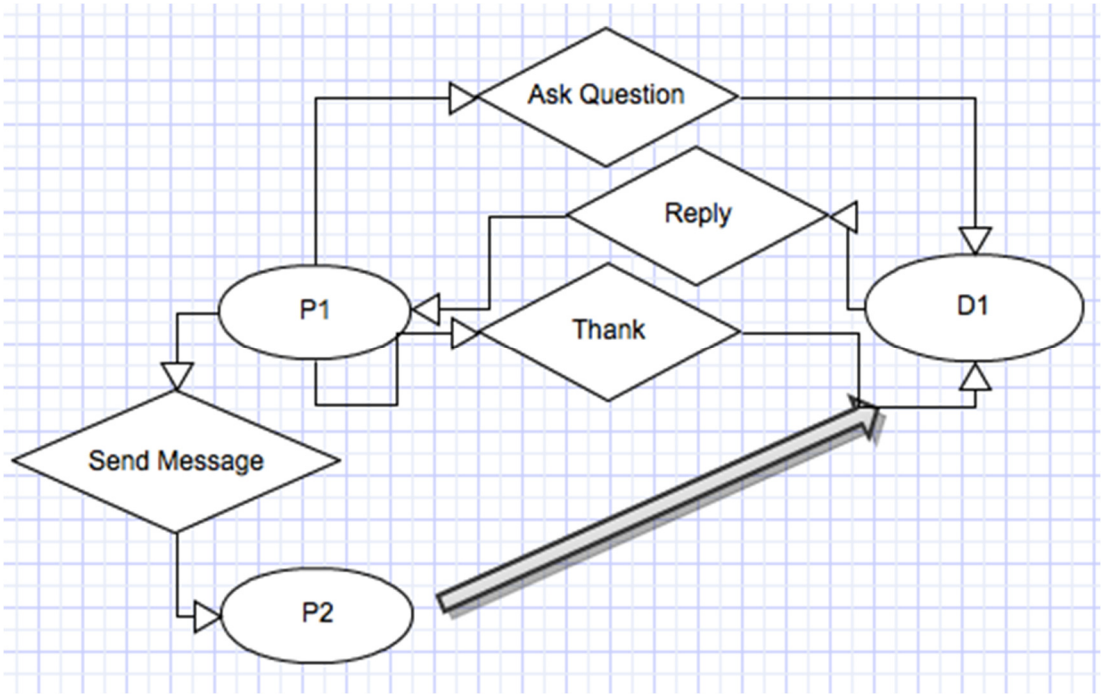

Figure 4.4. Representation of Step 3

- Step 4: Given that D1 is followed by P1, P1 asked a question to D1, received a reply from D1, and thanked to D1, will P2 ask question to D1?

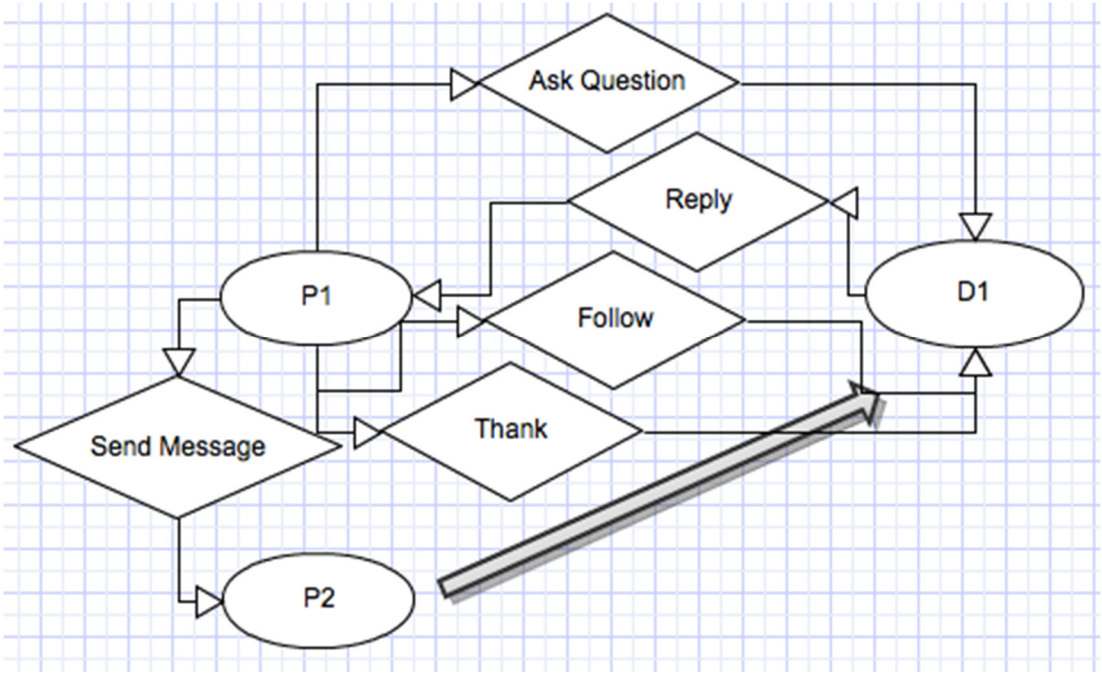

Figure 4.5. Representation of Step 4

Below we provide our basic assumptions for the analysis we conducted:

- In all our steps, we focused on the patients who received answers from the doctors. Hence P1 node in the figures represents the patients who asked question to a doctor, received an answer from the doctor and sent a message to another patient. There are 7.489 unanswered questions in our dataset, and if the doctor did not reply a patient's question, it means that the patient was not able to create a trust connection with another patient. Therefore, we skip the not answered questions.

- For the questions we analyzed, we also used timestamps for the events except follow and thank events. For the question: "Given that P1 first asked a question to D1 and received a reply from D1, will P2 ask question to D1?" the timestamps we used are:

    - T0: P1 asked question to D1

    - T1: D1 answered the question

- T2: P1 sent message to P2
- T3: P2 asked question to D1

In the Table 4.4, we provide information in terms of the number of nodes, edges and average degree for different types of relationships in our network.

Table 4.4. Dataset Statistics

Network	Nodes	Edges	Average Degree
Question&Answer	28.006	60.960	2.17
Message	2.558	11.285	4.45
Connection	1.420	1.768	1.24
Follow	2.575	3.632	1.41

For the calculation of the numbers in Table 4.4, we used the methods described in [12].

4.2 Results

In this section, we provide our results based on TaoBao [1] trust model. We provide the information passing success rate that is also called the triangle probability and details about the influences on information passing for our steps mentioned in the previous chapter. In terms of the variables that influence information passing we investigate communication strength and time difference.

Step 1

The information passing success rate of our network for this step can be defined as Prob(E1|E2) where E1 is P2 asks question to D1 at T3 , E2 is P1 asks question to D1 at T0 and D1 answers P1 at T1 and P1 messages P2 at T2 where T0<T1<T2<T3. There are 467 patients in P2 role that satisfy E2 and 46 patients in P2 role that satisfy E1 given E2, so Prob(E1|E2) = 0.098. We provide the results of the queries we used in our measurements in Table 4.5 and Table 4.6.

For this step, we also computed a base probability value for comparison. We used the probability of a patient asking a question to a doctor on the days that the 46 patients in P2 role asked question as our base of comparison which turned as 0.0037.

Figure 4.6 shows the relationship between the number of messages exchanged between the patients P1 and P2 and the triangle probability. Figure 4.7 gives the relationship between the time difference between reply from the doctor and the initial message sent to P2 and the triangle probability.

Table 4.5. Numbers of the Entities and Relationships on Step 1

Description	**Number**
P1	355
P2 (E2)	467
P2 (E1 and E2)	46
Probability	0,098501

Table 4.6. Numbers of the Messages Sent on Step 1

Number of Messages (Interval)	**Number of P2 satisfying E1 given E2**	**Probability**
1	13	0,02783726
[2,3]	15	0,03211991
>3	18	0,0385439

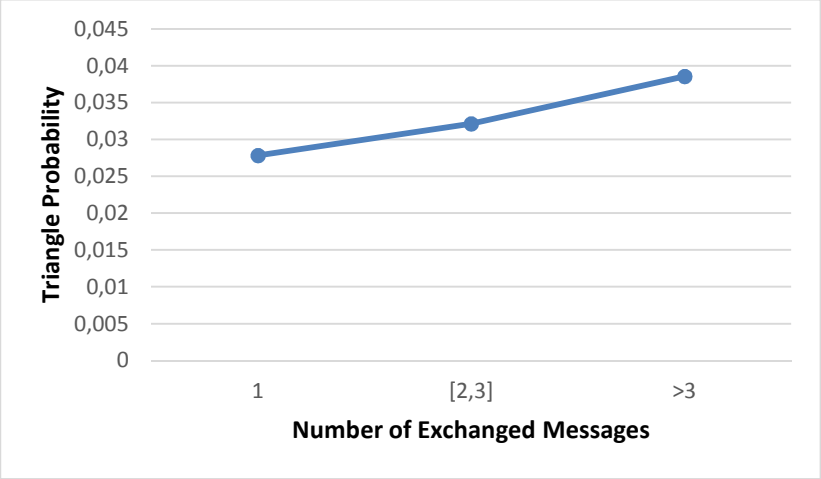

Figure 4.6. Number of Messages vs. Triangle Probability on Step 1

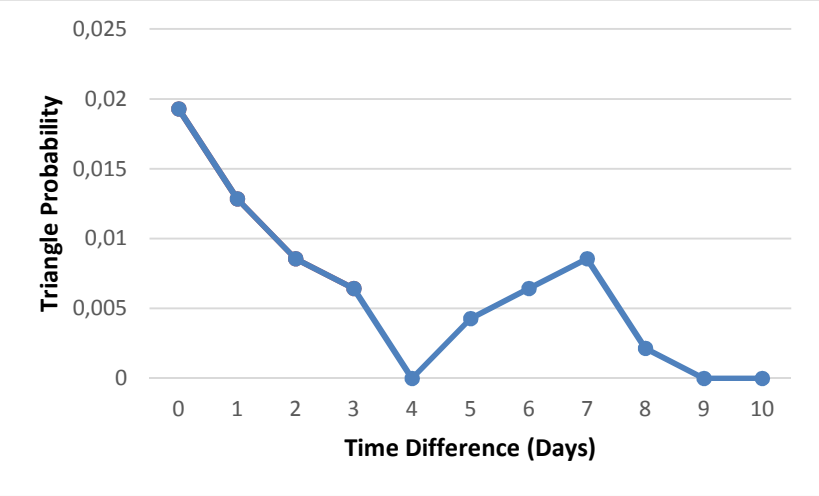

Figure 4.7. Time Difference vs. Triangle Probability on Step 1

Step 2

The information passing success rate of our network for this step can be defined as Prob(E1|E2) where E1 is P2 asks question to D1 at T3 , E2 is P1 follows D1, P1 asks question to D1 at T0 and D1 answers P1 at T1 and P1 messages P2 at T2 where T0<T1<T2<T3. There are 163 patients in P2 role that satisfy E2 and 10 patients in P2 role that satisfy E1 given E2, so Prob(E1|E2) = 0.06. We provide the results of the queries we used in our measurements in Table 4.7 and Table 4.8.

Figure 4.8 shows the relationship between the number of messages exchanged between the patients P1 and P2 and the triangle probability. Figure 4.9 provides the relationship between the time difference between reply from the doctor and the initial message sent to P2 and the triangle probability.

Table 4.7. Numbers of the Entities and Relationships on Step 2

Description	Number
P1	84
P2 (E2)	163
P2 (E1 and E2)	10
Probability	0,06135

Table 4.8. Numbers of the Messages Sent on Step 2

Number of Messages (Interval)	Number of P2 satisfying E1 given E2	Probability
1	3	0,01840491
[2,3]	3	0,01840491
>3	4	0,02453988

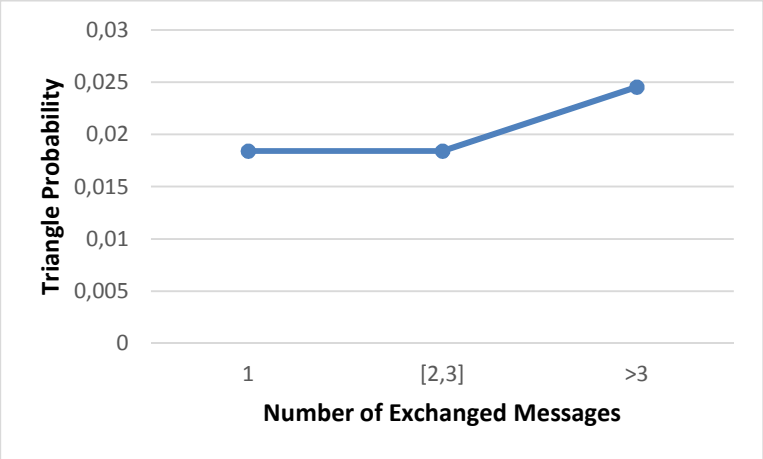

Figure 4.8. Number of Messages vs. Triangle Probability on Step 2

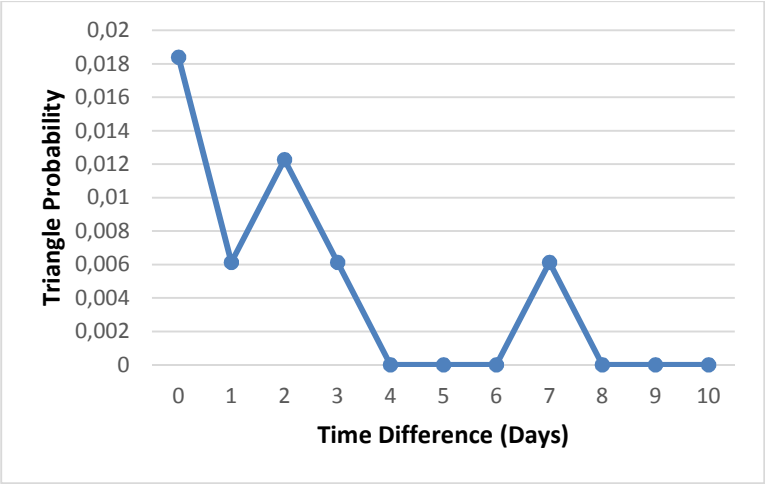

Figure 4.9. Time Difference vs. Triangle Probability on Step 2

Step 3

The information passing success rate of our network for this step can be defined as Prob(E1|E2) where E1 is P2 asks question to D1 at T3 , E2 is P1 asks question to D1 at T0 and D1 answers P1 at T1, P1 thanks to D1, and P1 messages P2 at T2 where T0<T1<T2<T3. There are 170 patients in P2 role that satisfy E2 and 10 patients in P2 role that satisfy E1 given E2, so Prob(E1|E2) = 0.58. We provide the results of the queries we used in our measurements in Table 4.9 and Table 4.10.

Figure 4.10 shows the relationship between the number of messages exchanged between the patients P1 and P2 and the triangle probability. Figure 4.11 provides the relationship between the time difference between reply from the doctor and the initial message sent to P2 and the triangle probability.

Table 4.9. Numbers of the Entities and Relationships on Step 3

Description	Number
P1	111
P2 (E2)	170
P2 (E1 and E2)	10
Probability	0,058824

Table 4.10. Numbers of the Messages Sent on Step 3

Number of Messages (Interval)	Number of P2 satisfying E1 given E2	Probability
1	3	0,01764706
[2,3]	3	0,01764706
>3	4	0,02352941

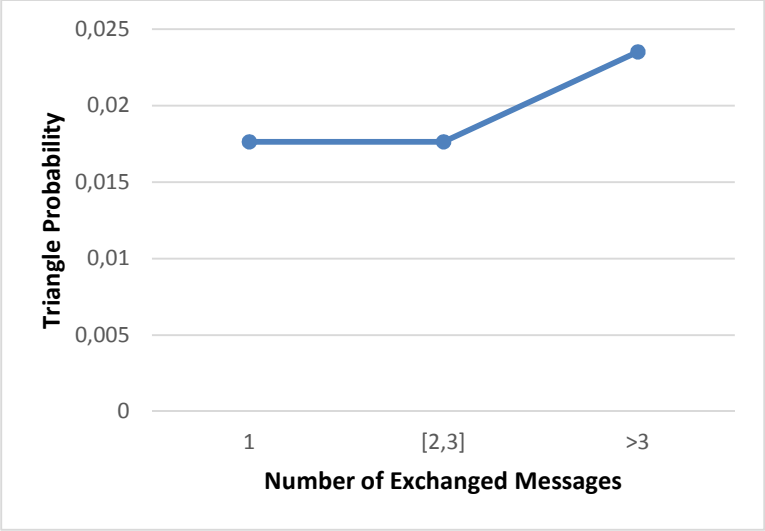

Figure 4.10. Number of Messages vs. Triangle Probability on Step 3

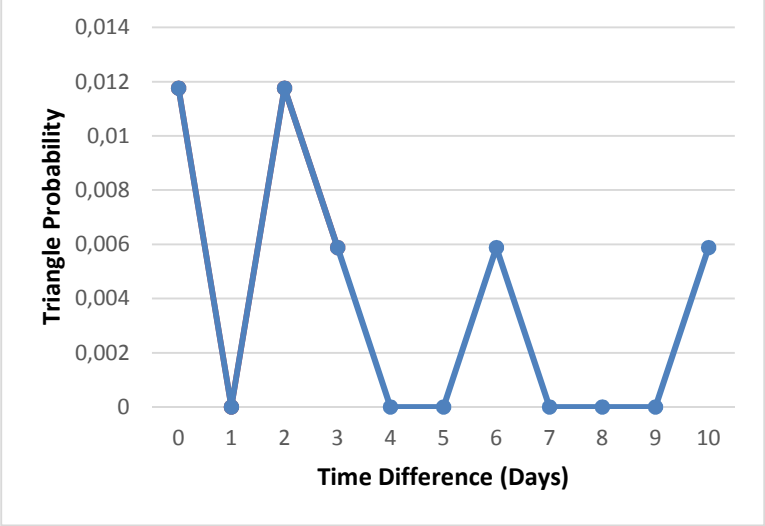

Figure 4.11. Time Difference vs. Triangle Probability on Step 3

Step 4

The information passing success rate of our network for this step can be defined as Prob(E1|E2) where E1 is P2 asks question to D1 at T3 , E2 is P1 follows D1, P1 asks question to D1 at T0 and D1 answers P1 at T1, P1 thanks to D1, and P1 messages P2 at T2 where T0<T1<T2<T3. There are 68 patients in P2 role that satisfy E2 and 2 patients in P2 role that satisfy E1 given E2, so Prob(E1|E2) = 0.02. We provide the results of the queries we used in our measurements in Table 4.11 and Table 4.12.

Figure 4.12 shows the relationship between the number of messages exchanged between the patients P1 and P2 and the triangle probability. Figure 4.13 provides the relationship between the time difference between reply from the doctor and the initial message sent to P2 and the triangle probability.

Table 4.11. Numbers of the Entities and Relationships on Step 4

Description	**Number**
P1	22
P2 (E2)	68
P2 (E1 and E2)	2
Probability	0,029412

Table 4.12. Numbers of the Messages Sent on Step 4

Number of Messages (Interval)	**Number of P2 satisfying E1 given E2**	**Probability**
1	1	0,01470588
[2,3]	1	0,01470588
>3	0	0

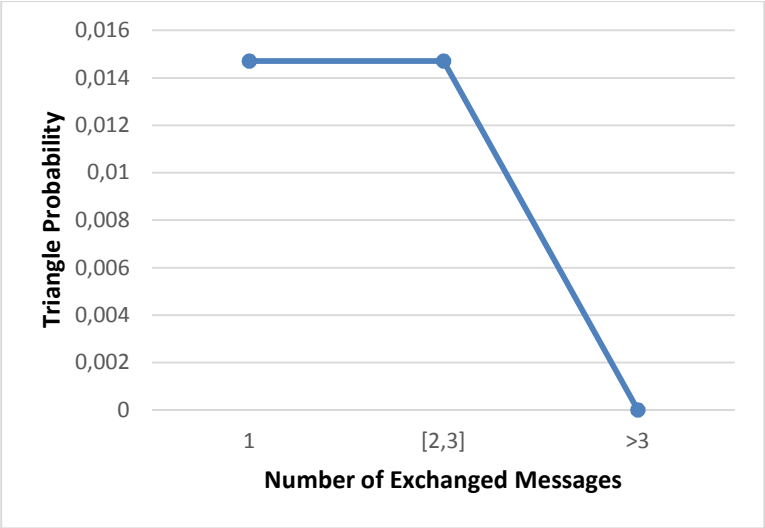

Figure 4.12. Number of Messages vs. Triangle Probability on Step 4

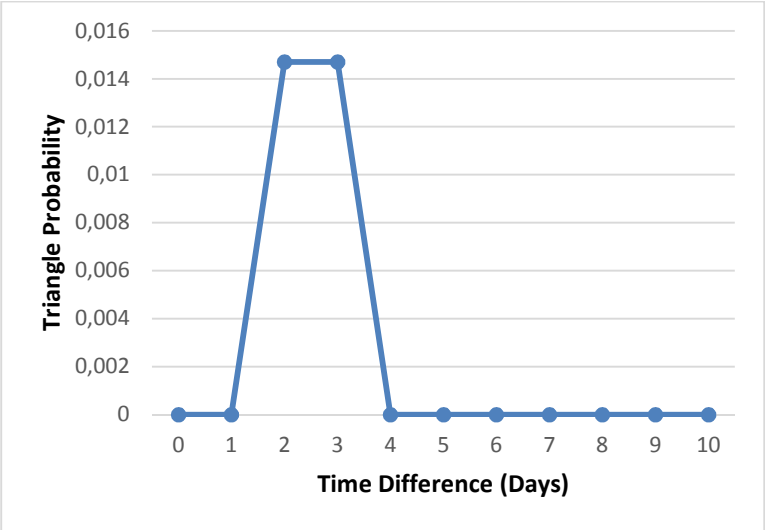

Figure 4.13. Time Difference vs. Triangle Probability on Step 4

Summary of the Analysis Results

We provide the triangle probability results of the steps we used in our measurements in Table 4.13, and the change in the triangle probability between the steps is depicted in Figure 4.14. We see a decrease in the probability from Step 1 to Step 4. One reason for this can be the limited size of our dataset which is more pronounced in

steps 3 and 4 compared to 1 and 2. Even though we had limited number of samples in the last steps we included them anyway for the sake of completeness.

Table 4.13. Analysis of Steps

Steps	Triangle Probability
Step 1	0,098501
Step 2	0,06135
Step 3	0,058824
Step 4	0,029412

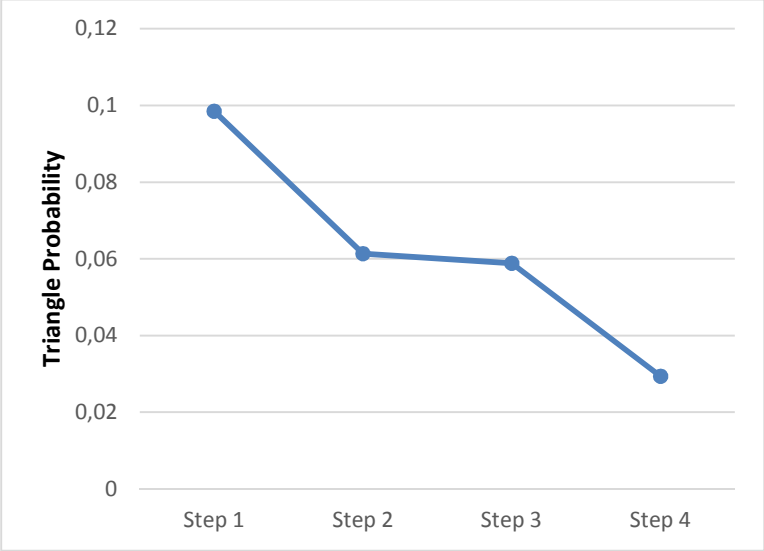

Figure 4.14. Steps vs. Triangle Probability

4.3 Discussion

Following our examination of the results we can argue that our dataset has parallel properties to the TaoBao [1] dataset in terms of information passing. The information passing success rate we computed for our dataset in Step 1 is 0.098 whereas it is 0.00203 for the TaoBao [1] network. One reason for the difference in the rate can be the size difference between the two datasets where ours is much smaller than the TaoBao [1] dataset. Another reason can be related to the internal properties of the two networks such as their use. Since our network is used primarily for getting

advice from professionals and the type of all the relationships are based on messages rather than trades and messages in the TaoBao [1] case, this can also be a reason for the higher rate of information passing in our case.

In addition to the existence of information passing we also investigated the relationship between the factors: communication strength and time difference, and information passing. As also stated in [1], it can be hypothesized that the stronger the communication between the two users, the more likely that information passing will occur. Our results are parallel to this expectation as it was also the case in [1]. In terms of the time difference it is expected that the larger the time difference between the interaction with the doctor and the message from P1 to P2, the lower the influence of the message on the interaction of P2 with the doctor. Our results show that the probability of information passing success generally decreases with time as it was also the case in [1].

Chapter 5

Conclusion and Future Work

Our work analyses the activities of 15.892 unique patients within 3 months of period in doktorsitesi.com. Through the study of Taobao case and using the trust methodology of Taobao, we verified that there is a connection among patients in terms of information passing and trust. Our results show the existence of information passing in our network. In terms of the relationship between information passing and related parameters, namely communication strength and time difference we also show that our results conform to the expectations as the Taobao network.

We hope our study will motivate the future research into online social networks on health. Future areas of related study include: analyzing the trust relationships in doktorsitesi.com among doctors and the success of information passing from the doctors' perspective, analyzing the relationship between the information passing success rate and strength of the relationship among doctors, and investigating the trust relationship in the online social networks supported by professionals and non-professionals.

The work accomplished to date marks a substantial road of progress in understanding the trust relationship on online social networks about health. Yet, it is also suggested that [3] we have at least as many miles to travel ahead. We are glad to be witness and a part of this progress with this thesis.

References

[1] Guo, S., Wang, M., Leskovec, J., “The Role of Social Networks in Online Shopping: Information Passing, Price of Trust, and Consumer Choice”, Stanford University, Stanford, CA, USA, 2011.

[2] Masic, I., Sivic, S., Toromanovic, S., Borojevic, T., Pandza, H., “Social Networks in Improvement of Health Care”, *Materia Socio Medica*, 2012.

[3] Pescosolido, B. A., Levy, J. A., "The Role of Social networks in Health, Illness, Disease and Healing: the accepting present, the forgotten past, and the dangerous potential for a complacent future", *Social Networks and Health Advances in Medical Sociology 8,* 3-25, Emerald Group PublishingLimited, Italy, 2001.

[4] Keckley, P. H., Hoffmann, M., “Social Networks in Health Care: Communication, collaboration and insights”, *Deloitte Center for Health Solutions*, 2010.

[5] Bender, J. L., Jimenez-Marroquin, M. C., Jadad, A. R., “Seeking Support on Facebook: A Content Analysis of Breast Cancer Groups”, *Journal of Medical Internet Research 16*, Canada, 2011.

[6] Media Tablet Shipments Outpace Fourth Quarter Targets; Strong Demand for New iPad and Other Forthcoming Products Leads to Increase in 2012 Forecast, According to IDC-IDC - Press Release, Available at: http://www.idc.com/getdoc.jsp?containerId=prUS23371312#.UNBgxayGfmg, 2012.

[7] Global Mobile Statistics, All the Latest Stats On Mobile Web, Apps, Marketing, Advertising, Subscribers, And Trends, Available at: http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats, 2012.

[8] Facebook Statistics by Country, Available at: http://www.socialbakers.com/facebook-statistics/.

[9] LinkedIn Statistics, Available at: http://www.socialbakers.com/linkedin-statistics/.

[10] What is Twitter? - Twitter Stats, Available at: https://business.twitter.com/basics/what-is-twitter/.

[11] Aggarwal, C. C., "Social Network Data Analytics", IBM T. J. Watson Research Center, NY 10598, 2011.

[12] McGlohon, M., Akoglu, L., Faloutsos, C., "Statistical Properties of Social Networks", *Social Network Data Analytics*, IBM T. J. Watson Research Center, NY 10598, 17-42 2011.

[13] Bhagat, S., Cormode, G., Muthukrishnan, S., "Node Classification in Social Networks", *Social Network Data Analytics*, IBM T. J. Watson Research Center, NY 10598, 115-148 2011.

[14] Dholakia, U. M., Bagozzi, R. P., Pearo, L. K., "A Social Influence Model of Consumer Participation in Network- and Small-group-based Virtual Communities", *Research in Marketing 21*, 241-263 (2004).

[15] Keckley, P. H., Hoffman, M., "2010 Survey of Health Care Consumers Key Findings, Strategic Implications", *Deloitte Center for Health Solutions*, 2010.

[16] Keckley, P. H., Coughlin, S., "2011 Survey of Health Care Consumers Global Report Key Findings, Strategic Implications", *Deloitte Center for Health Solutions*, 2011.

[17] Keckley, P. H., Coughlin, S., "Deloitte 2012 Survey of U.S. Health Care Consumers: The performance of the health care system and health care reform", *Deloitte Center for Health Solutions*, 2012.

[18] Al-Oufi, S., Kim, H., El Saddik, A., "A Group Trust Metric For Identifying People of Trust in Online Social Networks", *Expert Systems with Applications 39*, 2012.

[19] Zolfaghar, K., Aghaie, A., "Evolution of Trust Networks in Social Web Applications Using Supervised Learning", *Procedia Computer Science 3*, 2011.

[20] Golbeck, J., Hendler, J., "Inferring Trust Relationships in Web-based Social Networks", 2005.

Curriculum Vitae

Bora Göknar was born in 14 November 1988, in İstanbul. He received his BS degree in Management Information Systems in 2011 from IŞIK University, Istanbul. He is currently working as Product Manager in an IT company TRN Technology. His research areas include information security, network security, firewall systems, storage systems and storage area networks, online social networks and online health networks.