Telematics and Informatics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Evaluating reliability of question-disease relations in online health forms: A link prediction approach

Buket Kaya[a,*], Esra Gündoğan[b]

[a] Department of Electronics and Automation, Fırat University, Elazığ, Turkey
[b] Department of Computer Engineering, Fırat University, Elazığ, Turkey

ARTICLE INFO

Keywords:
Online health forums
Question-disease relations
Link prediction
Social network analysis

ABSTRACT

The Internet has become an indispensable part of human life in today. People can now easily find answers to questions they are curious about via the internet. The short, effortless and free way that the Internet provides is extremely attractive for people to have an idea in subjects they wonder related to their health. There are many online health forums where people can ask questions answered by health professionals. Every day, people ask thousands of questions on these sites and get answers about which diseases their complaints may be related to. The frequent use of online forum sites by people has led to the selection of these forums as data source for this study, and analysis of reliability. Firstly, in this study, link prediction in bipartite social networks, where intensive works have been done and it is applied on many areas nowadays, is tried to be carried out on question-disease bipartite network constructed with data obtained from analysis of online health forums whose use rate increase substantially. For this purpose, a novel link prediction method called as intensive link prediction is proposed, and prediction success of this method is compared with five of similarity-based link prediction methods. Better results have been obtained with the proposed method than the other methods. Then, the accuracy of the answers given to the users on online health forums which received intense interest are tested. The reliability of online health forums is measured by the accuracy analysis performed.

1. Introduction

Nowadays, the Internet is commonly used in the field of health as it is in every area. People can now easily find answers to questions they are curious about via the Internet. The Internet, through the short, effortless and free way it provides, has become a reference source for people to have ideas about their health problems. People use the Internet for many purposes in health care. They actively employ it to obtain an idea about which disease may be associated with their symptoms when people have certain disease symptoms. There are many online health forum sites where patients and doctors communicate on the Internet. Before people go to the doctor and do some tests, they write their complaints on online forum sites and get information from health professionals. Because of this tendency, forum sites have thousands of questions and answers. The frequent use of online health forums by people has led to the selection of these forums as data source for this study, and analysis of reliability.

We live together with many complex systems in our environment. The Internet, nervous system, protein networks, transport networks are some examples of complex networks we encounter. Complex networks are a structure used to model the complex systems, in which links are relationships, and nodes are persons, objects, etc. Complex networks make it easy to analyze the structure

* Corresponding author.
E-mail address: bkaya@firat.edu.tr (B. Kaya).

https://doi.org/10.1016/j.tele.2018.05.009
Received 21 February 2018; Received in revised form 24 April 2018; Accepted 22 May 2018

of complex systems, their development, and the relationships between the entities they represent. Therefore, the analysis of complex networks has become an important research area in many sciences.

Social network analysis is used in many areas for a variety of purposes. Link prediction among entities in the social network (Al Hasan and Zaki, 2011; Lü and Zhou, 2011), recommendation systems (Li and Chen, 2009), terrorist network analysis (Anil et al., 2015), and community discovery (Xu et al., 2013) are some of done studies on social network analysis. Analysis of enormous quantities of data on social networks provides good, useful information about individuals, communities, organizations (Wasserman and Faust, 1994). Valuable information can be discovered from social networks that will enable a better understanding of some situations and relations and allow improvements to be made in the current situation. Therefore, researchers from different fields have shown interest in social networks. But analyzing social networks is not an easy task. One of the difficulties in social network analysis is the dynamic nature of social networks. Over time, new nodes and links may be added to the network, or existing nodes and links may be lost. One of the most important social network analysis topics used to investigate dynamic structure of the network is link prediction. It estimates potential future connections by using the current state of the network (Liben-Nowell and Kleinberg, 2007).

Link prediction has important applications in many fields. It can be used for purposes such as finding co-authors for academic social networks (Pavlov and Ichise, 2007; Wohlfarth and Ichise, 2008), finding protein-protein interactions, exploring relationships that can occur between private persons in a network that shows the interactions of terrorists (Krebs, 2002), suggesting new friends for people (Aiello et al., 2012), recommending products to customers in online shopping (Akcora et al., 2011), realizing optimal orientation in sensor networks (Jia et al., 2013), finding relationships between diseases (Gül et al., 2016; Kaya and Poyraz, 2014; Kaya and Poyraz, 2015), estimating future citations (Jawed et al., 2015).

Many complex systems in our environment are bipartite network structure. The modeling of many complex systems in a bipartite network structure has increased the orientation to apply link prediction to such networks. Link prediction in networks modeled as bipartite has become an issue where intense work has been done so far. Different methods have been tried for link prediction in bipartite networks. Benchettara et al. (2010) have identified new topological features that can reflect the probability of a connection between two nodes and have used these features in a supervised machine learning approach to link prediction in bipartite networks. Xia et al. (2012) have made a study based on structural holes using weak ties between nodes. Nigam and Chawla (2016) have used a semi-bipartite network model to predict links in a bipartite network consisting of relations between users and health topics that interest them. Thus, it has been tried to be suggested health topics that will interest people by utilizing common features such as demographic characteristics, reading habits of people. Chang and Kao (2012) have formed a bipartite editor-article network with data obtained from Wikipedia. Some features have been extracted from the bipartite network created and these features were used in supervised machine learning algorithm. With this approach, link prediction has been made to suggest articles that authors can edit. Allali et al. (2011) have described a special type of link, called internal links, in bipartite networks. Link prediction has been made based on these links. Gao et al. (2017) have performed a projection-based link prediction. Link prediction is made based on assumption that the greater the sum of weights of patterns covered by a pair of candidate nodes, the higher the probability of a future connection between this candidate node pair. The concept of candidate node pair presented in the study is similar to the concept of internal links defined by Allali et al. (2011). However, there are some significant differences in this study. Computation time of this method has been reduced with applied changes and it has achieved higher speed.

In our previous study, we performed the prediction of symptom-disease links in online health forums (Gündoğan et al., 2017). In this study, we improve that method by considering question-disease links on a larger dataset. Also, as the second goal of this paper, we test the reliability of the information obtained from online health forums. For these goals, this study was carried out in two steps.

In first step,

- The questions on online health forums https://www.icliniq.com/tr/, https://www.drugs.com) were analyzed and user complaints were extracted.
- Symptom information for each disease was obtained from https://www.drugs.com/.
- The symptoms of the diseases and the symptoms expressed in the questions were compared, and found which diseases the questions might be related to.
- A bipartite network consisting of diseases and questions was constructed.
- The intensive link prediction method was applied to bipartite network.
- Five of similarity-based link prediction methods were applied to the network for performance comparison.
- Six different methods were compared in terms of precision, recall and F-measure.
- It was seen that the proposed method has better performance and higher accuracy prediction rate.
- Also, real life data obtained by analyzing online health forums where are thousands of questions every day was used for link prediction.
- In second step,
- A disease-symptom bipartite network was constructed using the disease-symptom dataset created with data extracted from https://www.drugs.com/.
- User complaints in each question were identified.
- By analyzing the answers given by the doctors to every question analyzed, the diseases that the doctor thought the user had, were determined.
- Diseases corresponding to the symptoms of each question analyzed were found using constructed disease-symptom bipartite network.
- For each question, the diseases expressed by the doctor and the diseases found using the actual network were compared.

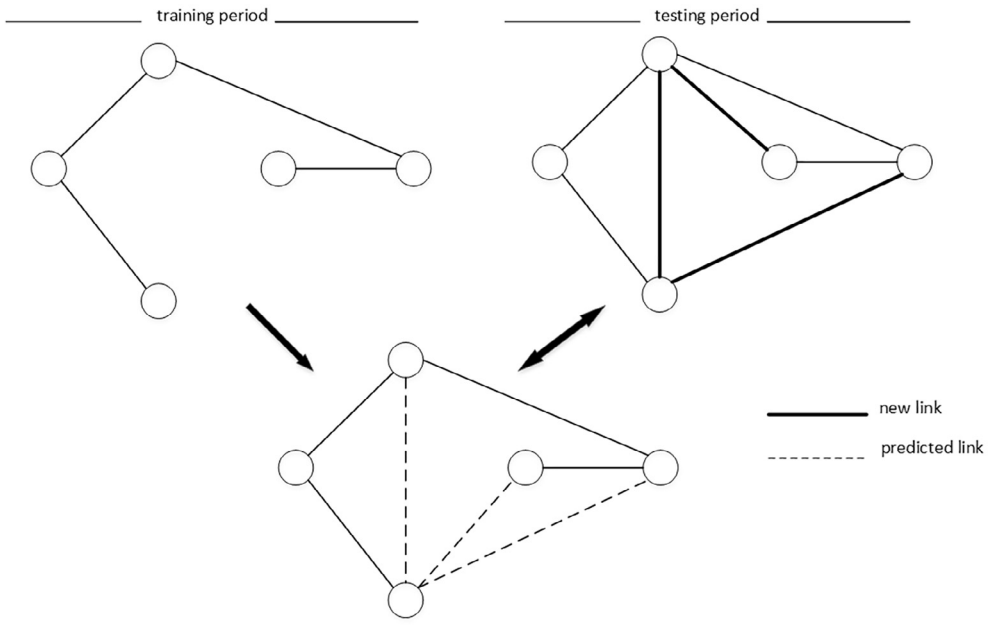

Fig. 1. Link prediction problem.

- It was found how accurate doctors' answers were. It was analyzed reliability of online health forums that people have frequently visited recently.

The rest of the paper is organized as follows. Section 2 includes link prediction problem in bipartite networks and baseline methods. Section 3 presents the proposed intensive link prediction method. Section 4 shows experimental results performed in question-disease bipartite network. Section 5 evaluates the reliability of online forum sites and finally Section 6 gives the conclusions.

2. Link Prediction in Bipartite Networks

2.1. Link prediction problem

The link prediction problem in social networks makes predictions about the future structure of the network. Link prediction methods are developed by using properties of nodes and relationships in social network. Network is defined as a graph. The data in the network are represented with nodes, and the relations are represented with links. The future of non-connected links between node pairs is predicted. Calculations such as likelihood of link formation in the future can be done according to the shortest path or link state between the two nodes.

Fig. 1 shows link prediction process. Link prediction methods calculate the probability of link formation between pairs of nodes that are not connected in the network in training period, based on some measures. If the probability of connecting two nodes is high, it is assumed that the link between the two nodes is more likely to occur in the future. According to this assumption, high-valued links are predicted. In the test period, the predicted links are compared to the new links. It is determined how many links are correctly predicted.

Bipartite graph is a special kind of graphs. Bipartite graph nodes are separated into two mutually independent sets as U and V. Each edge connects one cluster node to another cluster node. There is no link between the nodes in the same set, and links are only between nodes U and V sets. Many real social networks are bipartite networks, such as the scientists-papers cooperation network (Li et al., 2014), disease-gene network (Zhang et al., 2014), RNA-protein interaction network (Ge et al., 2016), drug-disease network (Gündoğan and Kaya, 2017). Link prediction in bipartite networks are used for many purposes, such as recommendation (Wang et al., 2016; Zhao et al., 2016), community discovery (Luo et al., 2014), prediction of drug side effects (Zhou et al., 2007).

Link prediction methods are generally applied to for single-mode networks. Since bipartite networks have two types of nodes unlike single-mode networks, these methods cannot be applied directly to bipartite networks. So, it is required to transform bipartite networks to single-mode networks. The single mode network obtained as a result of this transformation is called projection network. When creating a projection network, the links between the nodes are looked at. For two nodes in the same cluster to be connected in the projection network, these nodes must have at least one common neighbor from the other node set. Since the bipartite networks have two different sets of nodes, two different projection networks as called U-projection and V-projection are created for each set of nodes (Newman, 2001). Thus, bipartite graph is converted to single-mode graph.

Projection of bipartite network includes less information than bipartite networks. When creating projection network, at least one

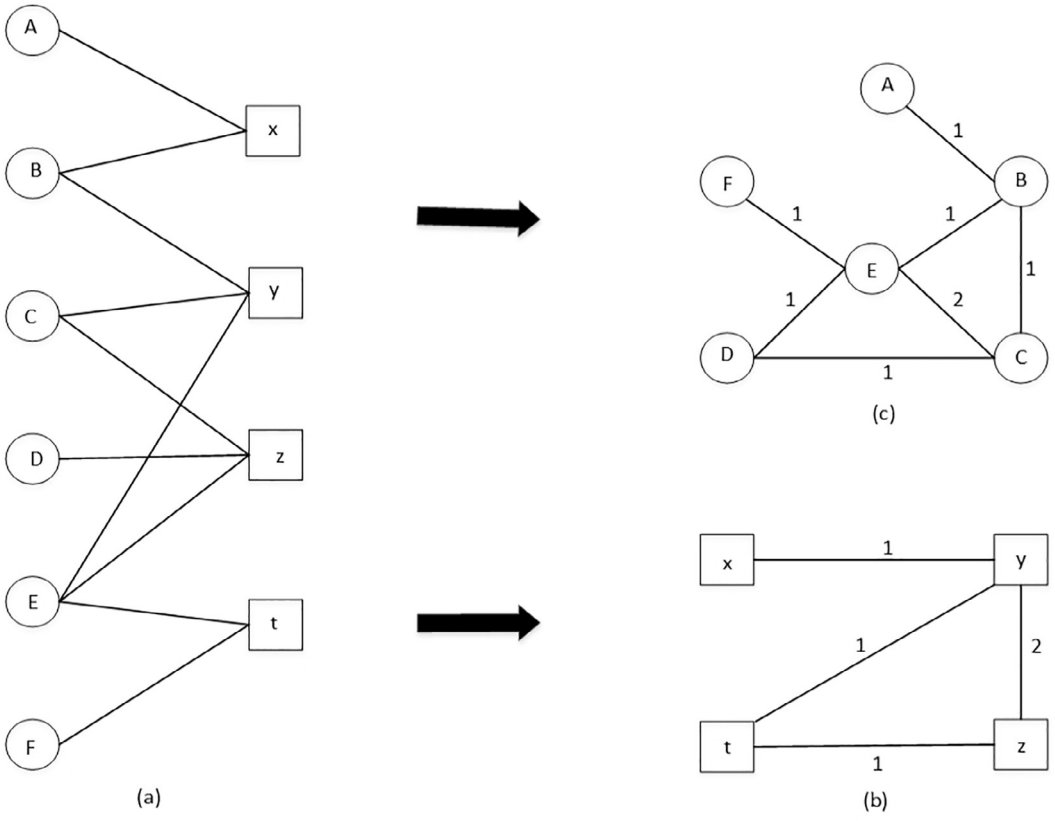

Fig. 2. (a) A bipartite graph example G, (b) U-projection, (c) V-projection.

common neighbor between the two nodes in bipartite network is sufficient to create a link between these nodes in projection network. However, there can be more than one common neighbor between the two nodes connected in projection network. The number of common neighbors indicates how similar these nodes are, and this similarity is one of the most important criteria used in link prediction. Projection network contains less information than original bipartite network. So, it is necessary to construct projection network as weighted. The weight of link (Wang et al., 2015) between two nodes on projection network is determined by the number of common neighbors that these two nodes have in the bipartite network.

Fig. 2(a)–(c) show a bipartite network example, U-projection network and V-projection network, respectively. The links in U and V projections has been weighted according to their common neighbors. For example; in Fig. 2(c), the weight of the link between C and E nodes has calculated as 2, because C and E are connected to two common nodes y and z in the bipartite network.

2.2. Baseline methods

Link prediction become a popular topic for many research areas in recent years. Two basic approaches have been proposed for link prediction by researchers (Malviya and Gupta, 2015). These are similarity-based and learning-based approaches. In a similarity-based approach, a score is calculated for each node pair that is not connected. The higher the calculated score, the higher the likelihood that the pair of nodes will be linked in the future. In learning-based approaches, link prediction is performed as a binary classification. The properties extracted from similarity-based approaches and social network are used to train the classifier in learning-based approaches. By using these approaches, you can have an idea about the structure and future of a social network.

Commonly used criteria for link prediction are similarity-based. Similarity is an important measure that reinforces predictions about the future of the network in the problem of link prediction. The two nodes, which are not connected to each other, are more similar to each other, and the link between these two nodes is more likely to occur in the future. For example, the more common friends of two non-friends on Facebook, the more likely they are friends in the future. The basic idea in similarity-based measures is that the more common the features of $\Gamma(x)$ and $\Gamma(y)$ of the neighbors of nodes x and y are, the higher the probability of linking them in the future (Tan, 2006).

Similarity-based approaches that used for performance comparison of the proposed method in this study are as follows.

Common Neighbors (CN): It is one of the most used metrics. It describes the number of common neighbors for x and y nodes. The greater number of common neighbors indicates that the similarity between the two nodes is high. Two nodes with high similarities are more likely to be connected in the future. Equation is defined as (Wang et al., 2015).

$$CN(x, y) = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

Jaccard Coefficient (JC): This metric indicates the ratio of the number of common features of nodes x and y to the total number of features x and y. Mathematical equation is given below (Adamic and Adar, 2003).

$$JC(x, y) = |\Gamma(x) \cap \Gamma(y)|/|\Gamma(x) \cup \Gamma(y)| \quad (2)$$

Adamic/Adar (AA): This measure gives a higher weight to rarer features. Thus, common features are counted easier. A node with a low weight has more influence on predicting a link between the x and y nodes. Equation is given below (Barabâsi et al., 2002).

$$AA(x,y)= \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{\log(z)} \quad (3)$$

Preferential Attachment (PA): The possibility of connecting the x and y nodes is proportional to the number of neighbors. Nodes with a large number of neighbors are more likely to create new links. Its definition as follows (Zhou et al., 2009).

$$PA(x, y) = |\Gamma(x)|\cdot|\Gamma(y)| \quad (4)$$

Resource Allocation Index (RA): It measures the likelihood of link formation between pairs of nodes that are not directly connected to each other. Each node has a resource unit and distributes its resource equally to all its neighbors. Nodes that have no connection between them receive their source through their common neighbors. The similarity between the nodes is determined by the resources they receive. Equation is given below (Zhou et al., 2009).

$$RA(x,y)= \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{k(z)} \quad (5)$$

3. Intensive link Predıctıon (ılp) method

In this work, a bipartite network was created with questions and diseases. Questions were obtained from online health forums https://www.icliniq.com/tr/, https://www.drugs.com/, and diseases and their symptoms were obtained from https://www.drugs.com/. Link prediction was done with six different methods and the results were compared.

Users write thousands of questions a day on online health forums to get information about their disease by specifying their complaints. In this study, firstly, it was determined disease symptoms that users express in their questions. The ElasticSearch data analysis tool was used for this. Every person expresses their complaints differently. This makes it difficult to obtain uniform symptoms from the questions asked. In order to remove this problem, ICD-10 codes presenting a standard form of disease and symptoms were used. Thus, in each question, the disease symptoms expressed in different forms were obtained in a uniform, complete and accurate form. Firstly, the ICD-10 codes of symptoms were indexed to ElasticSearch to obtain disease symptoms from the questions. Each question was analyzed with ElasticSearch, and it was found which indexed symptoms were in the question with similarity scores. The symptoms with high similarity scores were accepted as the disease symptoms that the users expressed in the questions. After disease symptoms has been identified that people have, it is necessary to find out which diseases these symptoms may be related to. For this, symptoms of each disease were found by using https://www.drugs.com/. This site provides information about causes, symptoms and treatment of diseases. It also includes questions/answers section. Users can get answers from registered doctors by writing their symptoms in this section. So, we also used this site as question data source. The symptoms of the diseases and the symptoms expressed in the questions were compared. Diseases that may be associated with the symptoms in the question were found. In the bipartite network links between question and disease nodes were formed according to these relationships. To construct bipartite network, ICD-10 codes of diseases and symptoms was used.

The constructed bipartite network contains two different types of nodes as questions and diseases. In the question node set, each question analyzed from online forum sites is shown as a node. In the disease node set, each disease in the database that created with data obtained from drugs.com web site is shown as a node. Firstly, the question-disease bipartite network was converted to weighted single-mode projection network since link prediction methods were performed on single-mode networks. Common neighbor weight function was used for weighting.

Internal link based intensive link prediction method (ILP) is proposed. This method defines links called internal links. Possible links are predicted using these links. Internal links are links that are not in the bipartite network but that do not change the projection of the network when it is added to the network. The bipartite question-disease network is transformed into single-mode network to determine internal links. Pairs of nodes that are not connected in the bipartite network are determined. By creating a link between the pairs of nodes, it is checked whether the projection network has changed or not. If the projection network does not change, the added link is defined as an internal link (Allali et al., 2011). All links specified as internal links are not considered in the link prediction. Unnecessary links can reduce the success of prediction. Therefore, a threshold value is determined. Links in the projection that have a weight larger than or equal to this threshold are accepted as internal links.

In Fig. 3 (Allali et al., 2011); *(B, l)* is an internal link. Neighbors of *l* node are *C, D* and *E* nodes. These nodes have at least one common neighbor with *B* node in the bipartite network (*C* with *i* and *j*, *D* with *j*, *E* with *k*). Therefore, they are connected with *B* node in the projection network. The projection network does not change when the *(B, l)* link is added to a bipartite network because they are connected.

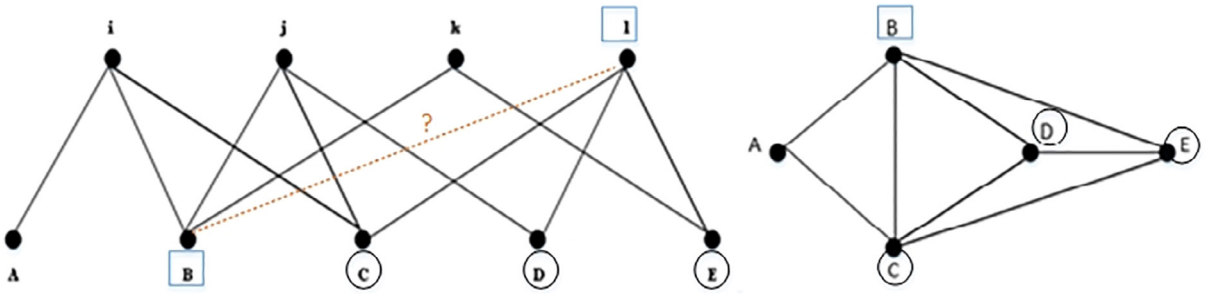

Fig. 3. Prediction of the proposed links.

4. Experimental results

In this section, performance evaluation of the proposed method is performed in question-disease bipartite network. The data used on the network were obtained from online health forums https://www.icliniq.com/tr/, https://www.drugs.com/ and https://www.drugs.com/. The bipartite network has 3352 nodes and 15,575 links. Firstly, to test the prediction success of the proposed method, 1000 links were selected from the bipartite network randomly. These links between questions and diseases were removed from the bipartite network. The proposed method was applied to the network constructed with 14,575 links. Links that will be internal links were determined. Weight of these links was compared threshold value. Links in the projection that have a weight larger than or equal to this threshold were determined as internal links. Then, the links found by our method were sorted according to their weight. The top-1000 high weighted links were selected. The predicted links with the proposed method were compared with the removed links. Success of the prediction method was obtained.

To evaluate performance of the proposed method, five of similarity-based methods were used. The selected methods are traditionally used for link prediction. There are many link prediction studies based on these methods. Our aim is to show the superiority of the proposed method against these commonly used methods. *Common Neighbors(CN), Jaccard Coefficient(JC), Preferential Attachment (PA), Resource Allocation(RA)* and *Adamic/Adar(AA)* methods were applied to the network constructed with 14,575 links. A *score (x,y)* function was defined that computes the probability of link between x and y nodes. This function shows the similarity of two nodes. The high function value indicates that the two nodes are similar to each other and are more likely to be connected in the future. This function was calculated for five similarity-based methods. For each method, the predicted links with this function were ranked according to their scores. The top-1000 high scored links were selected. The selected links for each method were compared to the links initially removed from the network. The comparison results were evaluated using three performance evaluation criteria precision, recall and F-measure. Using the calculated values for these criteria, the proposed method (ILP) was compared with other methods and the prediction success was analyzed.

Figs. 4–6 shows respectively the comparison of precision, recall and F-measure criteria for six different methods. As shown in Figs. 4–6, ILP method has better performance than the other methods. It has the highest values for the three performance measures. After ILP method, the highest scores are found by Adamic/Adar approach. In order to test link prediction methods, the extracted links from the bipartite network are more predicted by the proposed method. Fig. 7 shows performance comparison of ILP method for different number of test links. As can be seen in Fig. 7, as the number of test links increases, performance measure values raise as well. As you can see, the proposed method has predicted the removed links in a high rate. This success of prediction shows that the proposed ILP method can achieve high accuracy results in the prediction of the future state of a network.

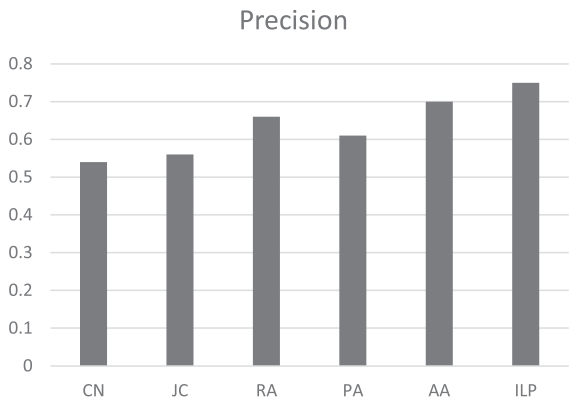

Fig. 4. Precision values comparison of six methods.

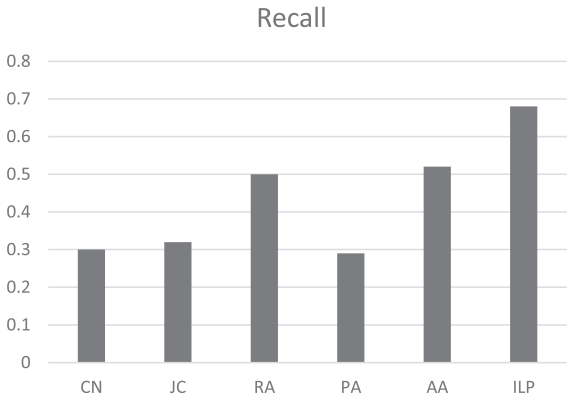

Fig. 5. Recall values comparison of six methods.

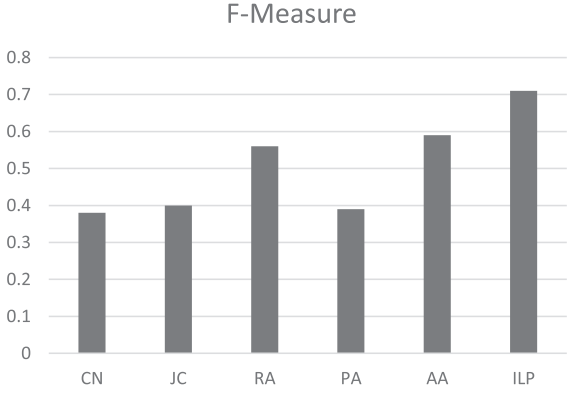

Fig. 6. F-measure values comparison of six methods.

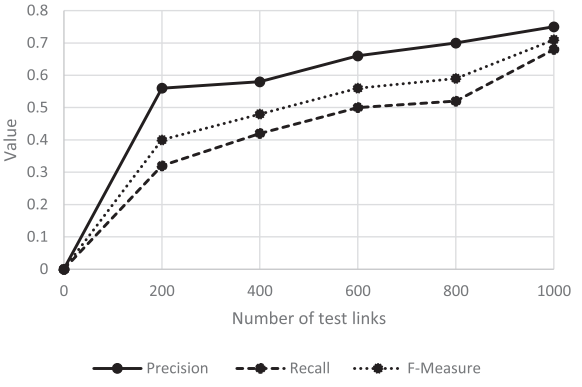

Fig. 7. Performance comparison of ILP method.

5. Relıabılıty of onlıne health forums

Today, the widespread use of online health forums is a reason for doing this work. Online health forums for people is a reference source to obtain information on health issues. The frequent use of these forums by humans reminds the question of how accurate the information received from these forums is. In this section, it was tested the accuracy of the answers given by health professionals to questions of the users in online health forums. The reliability of online health forums is measured by analysis made.

In online health forums, users ask questions about which diseases their symptoms may be related to, by writing their complaints. Registered health professionals respond to the questions of the users about the diseases that these symptoms may be related to. In order to test the reliability of doctors' answers respect to questions in online health forums, the actual knowledge of what diseases may be related to symptoms that the users express in their questions is needed. Therefore, the disease-symptom data set containing data obtained from https://www.drugs.com/ website has been used. In this data set, the main symptoms of each disease are found. A

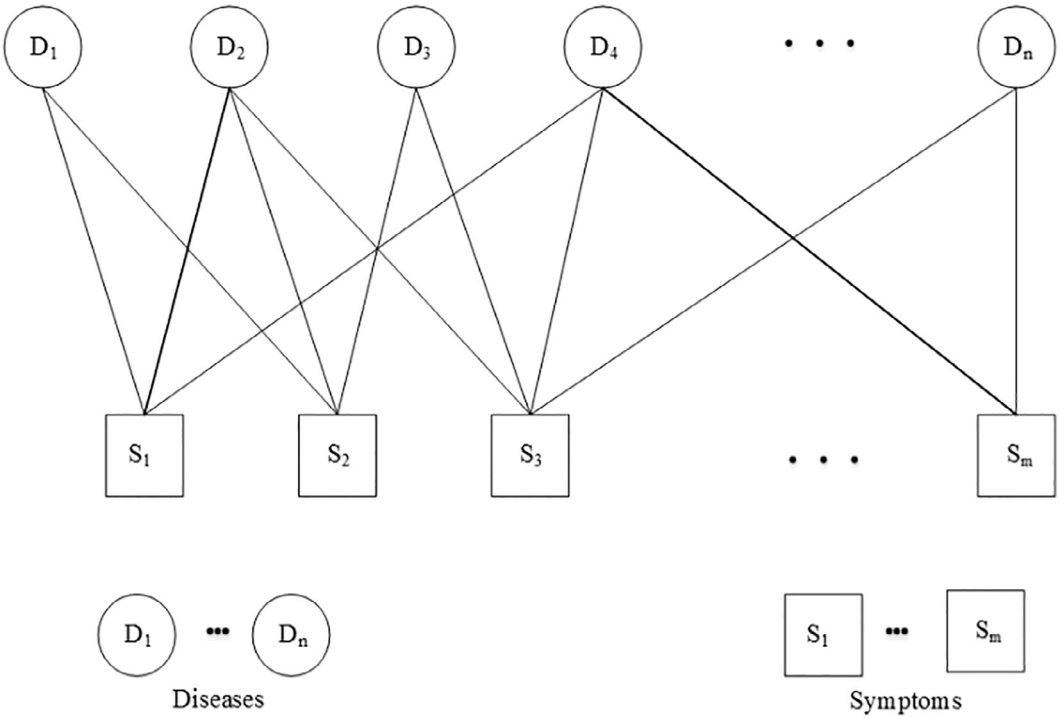

Fig. 8. Representation of disease-symptom bipartite network.

bipartite network has been modeled by using the disease-symptom data set. Clusters of nodes in a bipartite network consist of diseases and symptoms. The relationships between the disease and symptom nodes on the network are constructed based on the symptom information of each disease.

Fig. 8 is a representation of the disease-symptom bipartite network. Each disease node is linked to symptom nodes associated with that disease. Relationships that are defined on the modeled network show symptoms of each disease.

After determining of the diseases corresponding to symptoms, questions written by the users on the online health forums were analyzed. Symptoms expressed by the users in each question were obtained. Each question was kept with the symptom information. Then the doctor answers written to these questions were analyzed. The answers written by doctors respect to the questions were analyzed, and disease information was obtained which is considered to be related to the symptoms expressed by the user. Each question was kept with the disease information found.

After disease symptoms expressed in the questions, and diseases written by doctors in response to these symptoms have been identified, in order to test accuracy of the doctor's answers, it is necessary to determine which diseases actually associated with the disease symptoms obtained from the questions. For this, the modeled disease-symptom network was used. Relationships on this network show which disease has which symptoms. If symptoms expressed in the question have most connections with which diseases, these diseases were accepted as diseases which is related to the symptoms expressed by the user. In this way, it is found that the symptoms expressed in each question is actually related to which diseases by using disease-symptom network.

In this study, 4000 questions and answers written to these questions were analyzed. It has been found that which diseases the symptoms expressed by the users in the questions are related to, with answers of health professionals existing on the forum sites and by using the disease-symptom network constructed. For each question, the diseases expressed by the doctor and the diseases found using the constructed disease-symptom bipartite network were compared. As shown in Table 1, different numbers of questions were

Table 1
Reliabılıty rate for different number of questions.

Number of Questions	Reliability Rate (%)
500	68
1000	78
1500	81
2000	84
2500	85
3000	85
3500	86
4000	86

analyzed and reliability rates of doctors' answers were found. As the number of analyzed questions increases, reliability rate rise too. But after a certain step, the reliability rate is 86% despite the increase in the number of questions. According to the results, it was found that answers given by health professionals on online health forums have 86% reliability.

Today, some diseases have become widespread due to reasons such as living conditions, environmental factors, unhealthy nutrition. The fact that these diseases have specific symptoms and their frequent occurrence in people makes it easier to diagnose. Therefore, the proposed method has been successful in predicting such diseases over time. Comorbidity relationships also has an impact on the prediction success. A comorbidity relationship between two diseases exists whenever they appear simultaneously in a patient more than chance alone. This relationship has facilitated the interdependent prediction of diseases. Therefore, the reliability of the proposed method has increased over time.

6. Conclusıon

In this study, link prediction in a bipartite network was performed with an application in the health field. As data source for this, online health forums, which are a consultation area for people through rapid development of the Internet, have been used. These forums have attracted great interest from people in recent times. Before people go to the doctor, they try to get information about their problems from health professionals on these forums. It has been decided that data should be obtained from these forum sites by considering this tendency of people. Firstly, a question-disease network was created, and links are created between questions and diseases by converting questions to symptoms. Internal links method which is performing well in the studies that have been applied has been proposed for link prediction in this network. Firstly, the ILP method was applied to the projection of the bipartite network. Then, *Common Neighbor, Jaccard Coefficient, Adamic/Adar*, *Preferential Attachment* and *Resource Allocation* methods are used for performance comparison. When the results obtained are compared, it is seen that the proposed ILP method has better performance and predicts links with higher accuracy. After, reliability of online forum sites, which are highly popular, has been tested. It has been determined that well-known English online health forums have relatively good accuracy in a work realized by using constructed disease-symptom network.

In this study, link prediction is done on limited data. In the future we consider making predictions with larger data. In addition, in this study, it was predicted which diseases were associated with their symptoms, regardless of the medical history of the users. However, their medical histories play a significant role in link prediction. Some diseases may bring other diseases over time. Link prediction success can be increased taking account this relationship. Pre-diagnosis is also important for initial treatment. Therefore, a model can be developed to predict future disease risks based on the medical histories of people. This model provides both a high success in link prediction and a useful recommendation for people.

References

Al Hasan, M., Zaki, M.J., 2011. A survey of link prediction in social networks. In: Social Network Data Analytics. Springer US, pp. 243–275.

Lü, L., Zhou, T., 2011. Link prediction in complex networks: a survey. Phys. A Stat. Mech. Appl. 390 (6), 1150–1170.

Li, X., Chen, H., 2009. Recommendation as link prediction: a graph kernel-based machine learning approach. In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, pp. 213–216.

Anil, A., Kumar, D., Sharma, S., Singha, R., Sarmah, R., Bhattacharya, N., Singh, S.R., 2015. Link prediction using social network analysis over heterogeneous terrorist network. In: Smart City/SocialCom/SustainCom (SmartCity), 2015 IEEE International Conference on. IEEE, pp. 267–272.

Xu, Y., Chen, L., Zou, S., 2013. Community detection from bipartite networks. In: Web Information System and Application Conference (WISA), 2013 10th. IEEE, pp. 249–254.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge university press.

Liben-Nowell, D., Kleinberg, J., 2007. The link-prediction problem for social networks. J. Assoc. Inform. Sci. Technol. 58 (7), 1019–1031.

Pavlov, M., and Ichise, R., 2007. Finding experts by link prediction in co-authorship networks. In: Proceedings of the 2nd International Conference on Finding Experts on the Web with Semantics-Volume 290 (pp. 42-55). CEUR-WS. org.

Wohlfarth, T., and Ichise, R., 2008. Semantic and Event-Based Approach for Link Prediction. In PAKM (pp. 50–61).

Krebs, V.E., 2002. Mapping networks of terrorist cells. Connections 24 (3), 43–52.

Aiello, L.M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., Menczer, F., 2012. Friendship prediction and homophily in social media. ACM Trans. Web (TWEB) 6 (2), 9.

Akcora, C.G., Carminati, B., Ferrari, E., 2011. Network and profile based measures for user similarities on social networks. In: Information Reuse and Integration (IRI), 2011 IEEE International Conference on. IEEE, pp. 292–298.

Jia, X., Xin, F., Chuan, W.R., 2013. Adaptive spray routing for opportunistic networks. Int. J. Smart Sens. Intell. Syst. 6 (1).

Gül, S., Kaya, M., Kaya, B., 2016. Predicting links in weighted disease networks. In: Computer and Information Sciences (ICCOINS), 2016 3rd International Conference. IEEE, pp. 77–81.

Kaya, B., Poyraz, M., 2014. Supervised link prediction in symptom networks with evolving case. Measurement 56, 231–238.

Kaya, B. and Poyraz, M., 2015. Finding relations between diseases by age-series based supervised link prediction, 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1097–1103, Paris.

Jawed, M., Kaya, M., Alhajj, R., 2015. Time frame based link prediction in directed citation networks. In: Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference. IEEE, pp. 1162–1168.

Benchettara, N., Kanawati, R., Rouveirol, C., 2010. Supervised machine learning applied to link prediction in bipartite social networks. In: Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on. IEEE, pp. 326–330.

Xia, S., Dai, B., Lim, E.P., Zhang, Y., Xing, C., 2012. Link prediction for bipartite social networks: the role of structural holes. In: Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference. IEEE, pp. 153–157.

Nigam, A., Chawla, N.V., 2016. Link prediction in a semi-bipartite network for recommendation. In: Asian Conference on Intelligent Information and Database Systems. Springer, Berlin, Heidelberg, pp. 127–135.

Chang, Y.J., Kao, H.Y., 2012. Link prediction in a bipartite network using Wikipedia revision information. In: Technologies and Applications of Artificial Intelligence (TAAI), 2012 Conference. IEEE, pp. 50–55.

Allali, O., Magnien, C., Latapy, M., 2011. Link prediction in bipartite graphs using internal links and weighted projection. In: Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference. IEEE, pp. 936–941.

Gao, M., Chen, L., Li, B., Li, Y., Liu, W., Xu, Y.C., 2017. Projection-based link prediction in a bipartite network. Inf. Sci. 376, 158–171.
Gündogan, E., Kaya, B., Kaya, M., 2017. Prediction of symptom-disease links in online health forums. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. ACM, pp. 876–880.
https://www.icliniq.com/tr/.
https://www.drugs.com/.
Li, Y., Wen, A., Lin, Q., Li, R., Lu, Z., 2014. Name disambiguation in scientific cooperation network by exploiting user feedback. Artif. Intell. Rev. 41 (4), 563–578.
Zhang, Y., Tao, C., Jiang, G., Nair, A.A., Su, J., Chute, C.G., Liu, H., 2014. Network-based analysis reveals distinct association patterns in a semantic MEDLINE-based drug-disease-gene network. J. Biomed. Seman. 5 (1), 33.
Ge, M., Li, A., Wang, M., 2016. A bipartite network-based method for prediction of long non-coding RNA–protein interactions. Genom. Proteom. Bioinformat. 14 (1), 62–71.
Gündogan, E., Kaya, B., 2017. A recommendation method based on link prediction in drug-disease bipartite network. In: Advanced Information and Communication Technologies (AICT), 2017 2nd International Conference. IEEE, pp. 125–128.
Wang, X., Liu, Y., Xiong, F., 2016. Improved personalized recommendation based on a similarity network. Phys. A 456, 271–280.
Zhao, D., Zhang, L., Zhao, W., 2016. Genre-based link prediction in bipartite graph for music recommendation. Proc. Comput. Sci. 91, 959–965.
Luo, Y., Liu, Q., Wu, W., Li, F., Bo, X., 2014. Predicting drug side effects based on link prediction in bipartite network. In: Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference. IEEE, pp. 729–733.
Zhou, T., Ren, J., Medo, M., Zhang, Y.C., 2007. Bipartite network projection and personal recommendation. Phys. Rev. E 76 (4), 046115.
Newman, M.E., 2001. Clustering and preferential attachment in growing networks. Phys. Rev. E 64 (2), 025102.
Wang, P., Xu, B., Wu, Y., Zhou, X., 2015. Link prediction in social networks: the state-of-the-art. Sci. China Inform. Sci. 58 (1), 1–38.
Malviya, V., Gupta, G.P., 2015. Performance evaluation of similarity-based link prediction schemes for social network. In: Next Generation Computing Technologies (NGCT), 2015 1st International Conference. IEEE, pp. 654–659.
Tan, P.N., 2006. Introduction to Data Mining. Pearson Education India.
Adamic, L.A., Adar, E., 2003. Friends and neighbors on the web. Soc. Netw. 25 (3), 211–230.
Barabâsi, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Phys. A Statis. Mech. Appl. 311 (3), 590–614.
Zhou, T., Lü, L., Zhang, Y.C., 2009. Predicting missing links via local information. Eur. Phys. J. B Condensed Matter Complex Syst. 71 (4), 623–630.