



ELSEVIER

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Age-series based link prediction in evolving disease networks



Buket Kaya*, Mustafa Poyraz

Department of Electrical and Electronics Engineering, Firat University, Elazığ, Turkey

ARTICLE INFO

Article history:

Received 14 November 2014

Accepted 4 May 2015

Keywords:

Link prediction
Social networks
Proximity measures
Medical informatics

ABSTRACT

Recently, several research efforts based on social network analysis and methods have been made for medical care information. One of these efforts is to extract the relationships between diseases by using social network modeling. However, all of previous works used the relationships in a simple way in a network consisting of diseases regardless of time or age factors. In this paper, we predict the onset of future diseases on the basis of the current health status of patients by considering age factor. The problem of predicting the relations between diseases is a really difficult and, at the same time, an important task. For this purpose, this paper first constructs a weighted disease network and then, it proposes a novel link prediction method, to identify the connections between diseases, building the evolving structure of the disease network with respect to patients' ages. To the best of our knowledge, this is the first attempt in predicting the connections between diseases according to patients' ages. Experiments on a real network demonstrate that the proposed approach can reveal disease correlations accurately and perform well at capturing future disease risks.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Many complex systems, such as social, information, and biological systems, can be modeled as networks [1,2], where nodes represent individuals in the network, while links show the relations or interactions between the nodes [3]. Network is a useful tool in analyzing a wide range of complex systems. Social networks are social structures composed by entities and relationships among them. In this structure, new connections and entities appear or disappear overtime. This case makes them highly dynamic and complex systems. Analysis of social network includes broad field of research concerning method and strategies for the study of social networks [3]. Link prediction is an important task treated by social network analysis. This task is related to the problem of predicting the existence of a connection between two entities and the interactions already present in the network. Many efforts have been made to understand the structure, evolution, and function of networks [3–5]. In this meaning, link prediction tries to infer the likelihood of existence of a link between two nodes, which has important theoretical and practical value [6].

In many real world problems, link prediction can be employed to expect future behavior or to identify plausible relationships that are difficult or expensive to observe directly [7–9]. One of these scenarios is related to medical care research area. Medical care area should become more proactive than reactive in identifying the beginning of disease and risk [10,11]. At present time, doctors

generally utilize laboratory results to evaluate the patient's stage of health. However, these evaluations are generally based on only a few medical parameters and are extremely dependent on a particular doctor's experience, memory, and time. Hence, present medical care mostly steps in after a disease has emerged. Such an assessment does not treat disease at the earliest signs.

As a remedy to the above mentioned problems, in this paper, we propose a predictor to determine the risk of individuals to develop disease, and to undertake the correct actions at the earliest signs of illness. To this purpose, we first construct a weighted disease network which indicates the relationships between diseases. In this network, nodes are the diseases and edges connect those pairs of diseases appearing in the same patient. Then, we propose an age-series based link prediction method, ASLIP, based on the evolution of the constructed disease network in order to identify the relations in between diseases. Differently from the traditional link prediction methods, in this paper, the disease network is split into several age-sliced snapshots. The more often the interaction of two diseases with each other in each age-sliced snapshot, the stronger is the connection between these diseases. Also, the larger the number of common neighbors between two diseases, the higher is their probability to be connected in the future. The disease prediction problem can thus be naturally formulated as a link prediction problem if objects are diseases and a predicted link (x,y) between two diseases x, y means that a patient affected by the illness x (y resp.) is likely to be effected also by disease y (x resp.) after a specified age interval. In other words, we see hidden relationships strengthening between diseases instead of extracting associations between very common sicknesses. The search for these missing links is performed by applying similarity-based algorithms

* Corresponding author.

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

(see Section 3) that assign a score to the pair (x,y) that measures the proximity between x and y . Finally, we test the proposed method on the disease network constructed with laboratory results of patients more than 210,000. The experimental results demonstrate the accuracy of our method on unsupervised prediction and encourage us for further analysis.

The rest of the paper is organized as follows. Section 2 explains the construction of disease network. Section 3 briefly discusses the link prediction problem and gives some related work. Section 4 presents our age-series based link prediction method, proposing the evolving structure of disease network and the evolving cases in this network. The conducted experiments and the results achieved for the disease network are discussed in Section 5. Conclusions and future research directions are included in Section 6.

2. The construction of disease network

In disease network, nodes are the disease and edges connect these pairs of diseases appearing in the same patient. Thus, a relationship between two nodes exists whenever they appear in even different times in the same patient. In this paper, we first construct an evolving disease network based on age-series. For this purpose, the network is split into several age-sliced snapshots, which represent states of the network at different age intervals. After that, age-series are built by grouping sequential snapshots. These series are at the same time called as frames. We then exploit the connections between diseases to build a prediction model. In this model, we propose the utilization of link prediction methods to generate the predictive model of illnesses.

Fig. 1 shows a snippet of the disease network. The patient records contain an important view regarding the co-occurrences of illnesses affecting the same patient. In order to understand better the correlations among the diseases contained in our dataset, a disease network whose nodes are the diseases and a link between two nodes occurs every time those diseases are appeared in even different times in the same patient, i.e., when the couple of illnesses affects at least one patient, is built. The edges are labeled with the number of patients indicating both the illnesses.

3. Link prediction problem

A classic definition of the link prediction problem is expressed by: “Given a snapshot of a social network at time t , we seek to

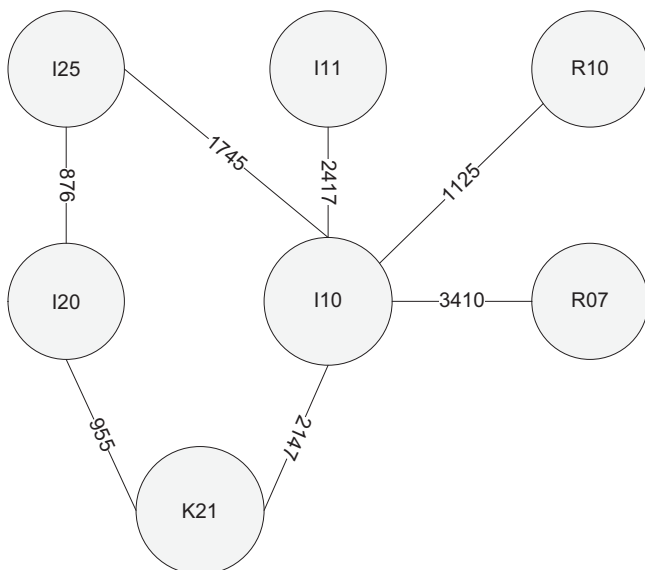


Fig. 1. A snippet of the disease network with ICD-10 code.

accurately predict the edges that will be added to the network during the interval from time t to a given future time t' [12]. So far, many approaches have been proposed in order to handle link prediction task. All of these methods are based on the measures indicating the proximity between nodes. These measures proposed in the literature are generally categorized into semantic or topological/structural measures [6]. In semantic measures, the content of the nodes is taken into account to evaluate proximity. For example, the similarity between topics discovered from titles can be used to predict future connections among the authors in a co-authorship network. As apart from the semantic ones, the topological measures consider the network structure to obtain the proximity values. Topological measures are more widely adopted because they are more general and do not require detailed information related to the content of node. Moreover, this content may not be present for every time at the social network considered. Topological measures are categorized into neighborhood-based or path-based measures [13]. The neighborhood-based measures consider the immediate neighbors of the nodes. According to this measure, if the neighbors of two nodes have a large overlap, they will probably form a link [6]. The most widely used measures among them are Common Neighbors [14], Preferential Attachment [14,15], Adamic–Adar Index [16] and Jaccard’s Coefficient [17]. The path-based measures generate a proximity value considering the paths between the related nodes. According to this idea, it is probable that two nodes will form a link if short paths are present in between the related nodes [18]. However, in order to predict new relations at future time, many of studies are established on the applications of proximity measures to non-connected pairs of nodes in the network at present time. Proximity measures showing the similarity between pairs of nodes can be utilized either by unsupervised [12,19,20], or supervised link prediction [21–23]. In unsupervised methods, a proximity measure is selected and utilized to rank node pairs in the network. The node pairs having good score are predicted to be linked. In the supervised approach, the link prediction problem is handled as a classification task and attributes of network are used as predictor attributes by a classification method. A classifier employs these attributes to implement a binary classification to find whether the link will be occurred or not in the future. In all of the above mentioned studies, the proximity values are calculated without considering the evolution of the network. In other words, the proximity measures are calculated using all network data up to the current network state without taking into account when links are created. Therefore, hidden and potentially valuable source of information is not adequately considered for link prediction.

There are specific cases occurred between a pair of nodes in sequential time intervals. These cases deal with creation or removal of link between nodes. For instance, a new link can be formed between two nodes in the next interval, while they are not connected in a specific time interval. This situation represents the evolution of the related social network. Some studies, recently, have been done concerning the use of time intervals for link prediction. Tylenda et al. [24] proposed a network model in which link weight was showing the age of the most recent specific cases between the related nodes. Briggmann et al. [25] proposed a network with temporal information for discovering association rules explaining the network evolution. The other approaches proposed for link prediction handles the task as a time series forecasting problem [26–28]. For this purpose, time series were first built for each pair of nodes. Here, each series observation is the frequency of occurrence of links between the nodes during a specific interval [28]. All of the studies done in this area consider time concept as interval. For instance, by using time series based link prediction in co-author network the collaborations between authors after several time series can be predicted. In this paper, as apart from the previous works, we handle age concept for intervals in the series. Our aim is

Download English Version:

<https://daneshyari.com/en/article/505213>

Download Persian Version:

<https://daneshyari.com/article/505213>

[Daneshyari.com](https://daneshyari.com)