Physica A 455 (2016) 120–135

Contents lists available at ScienceDirect

Physica A

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

Evolving networks—Using past structure to predict the future

Ke-ke Shang [a,b], Wei-sheng Yan [a], Michael Small [b,c,*]

[a] School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, PR China

[b] School of Mathematics and Statistics, The University of Western Australia, Crawley, Western Australia, Australia

[c] Mineral Resources, CSIRO, Kensington, Western Australia, Australia

HIGHLIGHTS

- We can use the links between a pair of nodes to predict their common neighbors.
- We find that the link weight have significant influence on our prediction.
- The rules of weighted networks which are dominated by human differ from other networks.
- The location and weight both have significant influence on the transport network.
- The structure of engineering networks has both best predictability and robustness.

ARTICLE INFO

Article history:
Received 23 September 2015
Received in revised form 25 January 2016
Available online 24 March 2016

Keywords:
Link prediction
Evolving network
Unweighted network
Weighted network
Network structure

ABSTRACT

Many previous studies on link prediction have focused on using common neighbors to predict the existence of links between pairs of nodes. More broadly, research into the structural properties of evolving temporal networks and temporal link prediction methods have recently attracted increasing attention. In this study, for the first time, we examine the use of links between a pair of nodes to predict their common neighbors and analyze the relationship between the weight and the structure in static networks, evolving networks, and in the corresponding randomized networks. We propose both new unweighted and weighted prediction methods and use six kinds of real networks to test our algorithms. In unweighted networks, we find that if a pair of nodes connect to each other in the current network, they will have a higher probability to connect common nodes both in the current and the future networks—and the probability will decrease with the increase of the number of neighbors. Furthermore, we find that the original networks have their particular structure and statistical characteristics which benefit link prediction. In weighted networks, the prediction algorithm performance of networks which are dominated by human factors decrease with the decrease of weight and are in general better in static networks. Furthermore, we find that geographical position and link weight both have significant influence on the transport network. Moreover, the evolving financial network has the lowest predictability. In addition, we find that the structure of non-social networks has more robustness than social networks. The structure of engineering networks has both best predictability and also robustness.

* Corresponding author at: School of Mathematics and Statistics, The University of Western Australia, Crawley, Western Australia, Australia.
E-mail address: michael.small@uwa.edu.au (M. Small).

http://dx.doi.org/10.1016/j.physa.2016.02.067
0378-4371/

1. Introduction

Link prediction is the key problem of predicting the location of unknown links from uncertain structural information for a network. Link prediction algorithms can aid in identifying unknown interactions to reduce the cost of experiment [1,2]. For example, in biology, the vast majority of interactions among the proteins are unknown [3–5], researchers have to spend significant expensive to recover these unknown interactions. Link prediction algorithms also could help us analyze the evolution of social networks [6], such as we could use the history structure of the online social network to predict which pair of friends will comment on each other in the future network. On the other side, link prediction algorithms are helpful for the algorithm design of recommendation [7] and the spurious links detection problem [8]. More examples including the reconstruction of networks and the classification of partially labeled networks [9].

Moreover, almost all experimental or real-world complex networks have significant temporal dynamics [10], the links, even the nodes of the network are not continuously active. However, the static network or topology, which links and nodes once established will not disappear or added anymore, could not accurately reflect temporal interactions of the real complex networks. For example, in the online social network, the friend relationships once established, then the links and nodes will not disappear. However, two users could comment on each other before a time point and not comment on each other over the long time period after the time point, even a few users will not login their online social accounts anymore. The static friend relationships could not reflect temporal interactions of the online social network. In the co-author network, we also could not only consider the static relationships among the scholars. Two scholars will usually cooperate with each other when they are working in the same research institution, however, they could not cooperate with each other after one of them leave the original research institution. In the stock market, we also can not only consider the relationships among the stocks in one period of time. Two stocks could have a strong correlation in the last year, but they could have a weaker correlation in this year. Research into the structural properties of evolving temporal networks has recently received increasing attention [11,10,12]. Meanwhile, temporal link prediction methods have also recently began to attract increasing attention [13–15]. Traditional link prediction methods pay more attention to use the common neighbors to predict the link between a pair of nodes—positing that nodes with many common neighbors are more likely to be neighbors themselves. Conversely, we use the existence of a link between a pair of nodes to infer the existence of common neighbors[1](some pair of links which connect to one same node) and analyzing the consequential variation of structure for evolving network.

Complex networks often exhibit stationary degree distributions despite the incessant creation and deletion of connections on broadly distributed time scales [16]. A consequence of this is that link prediction may be used to analyze the evolving network structure. However, each kind of network has its own particular structure—link prediction is, at its heart, a heuristic guess at unseen and unknowable network structure. For link prediction, the other issue is that of the different kinds of structure it is unclear whether they will exhibit stationarity or temporal robustness. It is unclear whether the relationship between weight and structure will have any influence on link prediction—an indeed this may only hold for certain types of networks. We will use our new method to predict the common neighbors and analyze the role of weight and structure. In addition, we also not only use our new unweighted prediction method to analyze the static structure and evolving network structure, but also use our new weighted prediction methods to analyze the static structure and evolving network structure.[2]

In this study, we use six kinds of data to construct six classes of networks: the online social network, the co-author network, the transport network, the terror news network, the financial network and the router network. First, we test our new unweighted prediction method and analyze the structure properties in static and evolving networks. We find that if a pair of nodes connect to each other in the current networks, then they will have a higher probability to connect common nodes in current and future networks and the probability will decrease with the increase of the number of neighbors. In addition, compared to the BA scale-free networks, we find that the original networks have their particular structure and statistic characteristics which benefit link prediction. Further, we find that our prediction measures have similar performance in the networks which are dominated by human behavior[3] (the online social network, the co-author network, the transport network and the terror news network). Moreover, compared to static networks, our prediction measure have their most unstable performance in the evolving financial network. Conversely, the most stable performance is in the evolving engineering network. We suggest that the structure of the financial network is more unpredictable than other kinds of networks, whereas the structure of engineering networks has more predictability than other kinds of networks.

Second, we use our new weighted prediction method to analyze the relationship between the weight and the performance of link prediction method, and the relationship between the weight and the structural properties in static and evolving networks. We also find that our prediction measures have similar rules in networks which are dominated by human

[1] If the node A and the node B connect the node C at the same time, node C is one of the common neighbors or nodes of the node A and the node B.

[2] Despite the fact that we are framing this paper in terms of "prediction", this property is highly system dependent. Our ability to "predict" with these methods depends entirely on the regularity and correlation of links within the network. The performance of link prediction is really a measure of this network property.

[3] Here, we make a distinction between the networks which are dominated by human behavior and social networks *per se*. Networks which are dominated by human behavior are the networks for which node interaction is an entirely human driven process, social networks are the networks in-which connection between nodes merely indicates communication.