Strength-Based Link Prediction in Scientific Bibliographic Networks

Djamila Mohdeb, Department of Computer Science, Bordj Bou Arreridj University, El Anasser, Algeria
Abdelhak Boubetra, Department of Computer Science, Bordj Bou Arreridj University, El Anasser, Algeria
Mourad Charikhi, Department of Computer Science, Bordj Bou Arreridj University, El Anasser, Algeria

ABSTRACT

Network analysis literature counts plenty of models of different paradigms designed for solving the link prediction problem in complex information networks. However, fewer studies that have exploited link strength-related social theories for this purpose even in a social context. In this paper, the authors introduce a new approach to solve the link prediction problem in scientific bibliographic networks. The aim is to predict future collaboration relations between scientists relying upon the “strength of strong ties” hypothesis. The proposed model estimates the strength of a relation between two scientists using a set of efficient link strength indicators. The importance of the relation is then validated according to the scientists’ expected collaboration strategies. The prediction process is performed in a heterogeneous context where the types of the nodes and the links are considered. Experiments on the DBLP real-world scientific bibliographic network, show higher performance of our model in comparison with the link prediction baseline methods.

KEYWORDS

Heterogeneous Information Network, Link Prediction, Link Strength, Network Evolution

INTRODUCTION

Scientific bibliographic network is a complex network describing the social structure of science in a scientific discipline across the links that relate scientific entities. These links generally take the form of collaborations to produce new scientific knowledge. As science is a social institution where advances depend crucially on these forms of interactions (Katz & Martin, 1997), many studies have invested in identifying the diverse factors that control their dynamic and their evolution in the bibliographic network (Newman, 2001; Velden et al., 2010; Franceschet, 2011). Indeed, studying the evolution of a scientific network presupposes using effective techniques in order to understand how a link appears or disappears between the nodes. Link prediction (Liben-Nowell & Kleinberg, 2003) is one of the well-known techniques that attempt to answer this question. In the literature, it refers to predicting a probable association between two given nodes in a given time interval in the future relying on their behavior in a given time interval in the past. The relevance of this technique reveals in its widespread applications in various domains. For instance, in online social networks it is used to help actors to form new social relationships, in e-commerce it is used to build efficient recommender systems, in...
security it can help to find suspicious interactions between criminal groups, and in bioinformatics and biomedicine it can help to identify probable interactions between proteins or between diseases and genes. As a result, the extensive study of this problem has produced several models ranging from supervised and unsupervised methods to probabilistic methods, relational models, linear algebraic models... etc. For more information, the reader may find a detailed survey of these models in (Hasan & Zaki, 2011).

Link prediction in scientific bibliographic network seeks for predicting a future formal collaboration between two researchers relying on their collaboration patterns in the past. Early works in this field focused on the network structure where the structural distribution of links and nodes plays an important role in the construction of the network, or on their attributes that contain rich information susceptible to provide expectations that are more reasonable. However, the majority of them uses only co-authorship networks in which the structure is homogeneous i.e. the nodes and the links belong to the same node type (author) and the same link type respectively (co-authorship). A co-authorship network is formed from a scientific bibliographic network by connecting coauthors contributing in the publication of one or more scientific papers. It can best capture formal interactions between scientists (e.g. contributing in a certain part of research) but it does not really reflect the whole collaboration process (Katz & Martin, 1997) due to its incapability to express the informal forms of collaboration (e.g. discussion or exchange of expertise between two scientists on a certain topic). Therefore, since a co-authorship network reflects a partial view of the entire scientific bibliographic network, link prediction traditional models designed for bibliographic networks mostly have the weakness of ignoring and neglecting rich information of the entire network that may have a considerable effect on the link formation between two authors. Fortunately, latest advances in this area (Taskar et al., 2003; Davis et al., 2013; Sun et al., 2011; Sun et al., 2012; Sun & Han, 2013; Yang et al., 2012) have carried interest to the heterogeneous aspect in networks and have taken into account the different types to which a node or a link may belong respectively. Consequently, these studies have proposed new models to tackle the limitations of homogeneous models. Most of them have proved a significant improvement and opened new perspectives that can be exploited to develop more performant link prediction methods. As a contribution in this research axis, this paper proposes that studying the link strength is a good technique that help to find solutions for the link prediction problem in bibliographic networks since it can measure the importance of both the existent relations and the absent relations in the network. The well-known definition of link strength is that proposed by Granovetter in his famous “strength of weak ties” hypothesis (1973). He wrote, “The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding) and reciprocal services which characterize the tie”. The tie strength indicators proposed in this definition or in other later studies (Granovetter, 1974; Lin et al., 1981; Marsden & Campbell, 1984; Mitchell, 1987; Perlman & Fehr, 1987; Blumstein & Kollok, 1988; Mathews et al., 1998; Benassi et al., 1999) may be efficiently used to realize the difference between a connection and another in a scientific social network. This differentiation becomes clearer with observing the structure, the type, and the attributes of the link and its two extremities in addition to benefiting from the heterogeneity nature of the network. Our work attempts to apply this idea in order to predict future formal collaboration links between two authors in a scientific bibliographic network. In summary, our main contributions include:

1. We proposed a new approach to solve the link prediction problem in scientific bibliographic networks based on the “strength of strong ties” hypothesis, the opposite version of the famous “strength of weak ties” hypothesis;
Balancing Theoretical and Practical Goals in the Delivery of a University-Level Data Communications Program


[www.igi-global.com/article/balancing-theoretical-practical-goals-delivery/44548?camid=4v1a](www.igi-global.com/article/balancing-theoretical-practical-goals-delivery/44548?camid=4v1a)