Knowledge-Based Systems 000 (2017) 1–12

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A new knowledge-based link recommendation approach using a non-parametric multilayer model of dynamic complex networks

Yasser Yasami

Department of Computer Engineering and Information Technology, Payame Noor University (PNU), P.O. BOX 19395-3697, Tehran, Iran

ARTICLE INFO

Article history:
Received 19 November 2016
Revised 5 December 2017
Accepted 7 December 2017
Available online xxx

Keywords:
Dynamic complex networks
Social networks
Collaborative networks
Multilayer networks
Link recommendation

ABSTRACT

Traditionally, research on network theory focused on studying graphs with equivalent entities failing to deliberate the useful supplementary information related to the dynamic properties of the complex network interactions. This paper tries to study the evolution process of dynamic complex networks from a multilayer perspective by analyzing the properties of naturally multilayered web-based directed complex social networks of Google+ and Twitter, and undirected collaborative networks of DBLP and ASTRO-PH, thereby proposing a new non-parametric knowledge-based multilayer link recommendation approach. The paper investigates the layers' evolution throughout the network evolution, inspects the evolution of each node's membership in different layers by an Infinite Factorial Hidden Markov Model, and finally formulates the intra-layer and inter-layer link generation process. Some Markov Chain Monte Carlo sampling strategies are driven to simulate parameters of the proposed multilayer model, using certain synthetic and real complex network datasets. Experimental results indicate great improvements in the performance of the proposed multilayer link recommendation approach in terms of certain analyzed performance measures.

1. Introduction

Many real world phenomena like human interactions are the consequence of dynamical structure of systems consisting of a large number of substantial entities with complex interactions. One of the main research efforts is then providing proper representations of such systems. Such research efforts in recent years have led to the development of a research field in science referred to as complex networks theory. However, the traditional monolayer theory for complex networks representation deals with all of the network's constituents, equivalently. This major restriction has directed the recent studies in the complex network theory to enhance the traditional monolayer network theory to contain multilayer networks, i.e. graphs with several layers of connectivity, defined to exactly describe the interactions between network fundamental entities [1].

On the other hand, the unprecedented growth of complex networks makes predicting behavior of such systems too complicated for development of recommendation systems. Nowadays, recommendation systems are used in various categories of applications [2]. Recommendation systems are developed for e-government applications to facilitate difficulties in locating the accurate information for the proper citizens and businesses. E-business and e-commerce applications of recommendation systems focus on recommendations issued to individual customers and business users, respectively. Digital library applications benefit from e-library recommendation systems to guide users find satisfying information and knowledge resources. Applications in the category of e-learning recommendation systems typically help e-learning students to acquire their favorite learning materials and select appropriate learning activities. E-tourism recommendation systems [3] are designed to provide suggestions about attractions, destinations, transportation, restaurants, and accommodations for tourists. E-resources such as webpages, TV programs, and documents can be suggested to users by applications developed in the category of e-resources recommendation systems.

However, almost all of the traditional recommendation systems are unable to provide recommendations to groups. Accordingly, group recommendation systems are designated to incorporate and equilibrate the individual favorites and expectations of group members to appropriately recommend to the group instead of the individual group members. The recommendation systems in all of the mentioned categories of applications also apply different techniques from traditional methods such as collaborative filtering-based, content-based, knowledge-based, and hybrid methods, to certain modern methods, such as computational intelligence-based, social network-based, trust-based, context awareness-based, and group recommendation methods to make proper suggestions for right users [2].

E-mail address: yasami@pnu.ac.ir

https://doi.org/10.1016/j.knosys.2017.12.005

Link recommendation in complex networks as one of the main applications of the recommendation systems consists of predicting new relationships, given a history of a network at previous time and can be broadly categorized into similarity-based, maximum likelihood, and probabilistic model-based approaches [4–6]. The similarity-based strategies are based on the simple idea of the more similar the node pair, the more probable it is to observe a link between them [7–11]. The maximum likelihood approaches presuppose some organizing principles of the network structure, with detailed principles and specific parameters acquired by maximizing the likelihood of the observed structure [12]. However, the main goal of probabilistic models is extracting the underlying structure of the network, and then applying a learned model for the link recommendation. The extracted models of dynamic complex networks can also be useful for link recommendation [13–25], community detection [26,27], forecasting future relationships [28,29], group growth and longevity prediction [30], and anomaly detection [31,32]. The knowledge-based link recommendation approach proposed in this paper provides modeling of node features and the relationships between network entities using a probabilistic approach, and so can be categorized as a probabilistic model-based method.

While link recommendation in complex networks has been addressed in a lot of literature, almost all of the previous work ignores to consider the multilayer aspect of dynamic complex networks. In addition, lots of recent monolayer probabilistic model-based structures of dynamic complex networks fail to straightly apply information captured from past network observations to model the future hidden structure. This is insufficient for social and collaborative complex networks, as the human social relationships certainly influence the personal interests as well as the future social interactions, a phenomenon that is called feature cascade. Furthermore, like many other natural phenomena, complex networks can be decomposed to multiple simultaneously active layers of features and pairwise interactions of nodes that can influence the network evolution. However, it is impossible to apply common dimensionality reduction methods to infer the exact number of layers and identify hidden structures [33] in the observed data, since these methods typically require a fixed number of layers, in advance.

As one of the main contributions, the model proposed in this paper considers the multilayer aspect of directed/undirected dynamic complex networks. The model applies the observed social relationships between entities in a common active layer as well as the entities in different active layers at each time step to model distributions over the hidden structures of the network at future time steps. As another contribution, the proposed approach applies the Infinite Factorial Hidden Markov Model (IFHMM) [38] to capture dynamics of individual nodes taking and giving up features with respect to feature cascade. The proposed IFHMM-based model benefits from the major advantage that each network entity under this model can potentially have any arbitrary number of features without some features of an entity limiting that entity to have other features. This is possible in the proposed model through simultaneous membership of each entity in multiple layers. Despite the related traditional probabilistic approaches [28,29,34,35] facing with some ambiguities due to the lack of explicit modeling of the features' dynamics, this paper also provides a strong approach based on the Indian Buffet Process (IBP) [36,37] for layers' dynamics, as another contribution. By explicitly modeling the layers' dynamics, feature cascade, and layer connectivity structure greater modeling flexibility is achieved. This leads to improvement in the performance of the proposed multilayer link recommendation approach based on the defined criteria, as well as an increase in the interpretability of obtained results.

After this introduction, the paper is organized as follows: Section 2 reviews the background and related work of this research field. The proposed non-parametric knowledge-based multilayer model is provided in Section 3, including the process of layers' evolution, dynamics of nodes' membership in various layers and its relationship with the network link generation, as well as the model inference. Section 4 includes description of datasets and performance measures, experimental set up, results of experiments conducted on synthetic and real complex networks, and experimental results analyses. Finally, Section 5 contains conclusion and future work.

2. Background and related work

There is a large body of literature on the important problem of link recommendation in complex networks. Similarity-based approaches in despite of their simplicity are faced with some challenges. The definition of node similarity is a significant challenge because of the heuristic nature of this problem. Existing similarity indices can be very simple or even very complicated, and can be categorized into node-based and structural similarity indices. Each of the similarity indices may perform particularly promising for some complex networks, while really fails for some others [4–6]. There are structural similarity-based approaches based only on the network structure that do not consider the knowledge of microscopic features of nodes [5]. Neighbor-based [8,15–18,39], path-based [40], and random walk-based [13,14,23] methods are the examples of these approaches.

From the outcome of recent researches in the field, it can be concluded that deploying only structural similarity indices cannot provide the expected performance for the link recommendation. Yet, many of the existing link recommendation approaches [8,13–18,23,39–43] just apply the structural similarity indices. There are also some approaches based on only the structural similarity indices trying to overcome the prohibitive computational cost of the applied complicated indices [41,42]. Moreover, the approaches that incorporate both of the node-based and structural similarity indices mainly consider a limited number of node-based similarity indices, while neglect the effect of many other individual node features on the network evolution. For example, the approach recently proposed in [7] just takes into account four node-based similarities that do not appear to be sufficient for link recommendation in complex networks.

There are also maximum likelihood approaches [26] being categorized as another group of link recommendation approaches. From the practical perspective, an explicit disadvantage of these methods is that they are very time consuming, such that a strong approach is capable to cope with networks with up to a few thousand nodes in a reasonable time, but will definitely fail to deal with the huge online networks including millions of nodes. Moreover, these approaches are probably not among the high performance ones. However, the maximum likelihood methods have the main advantage of providing very valuable insights into the network organization, which cannot be gained from the other categories of approaches [4]. There are approaches based on social theory metrics applying social interaction information [11,44–46]. Somewhat different to the previous metrics, which only use node-based and structural similarity indices, the link prediction metrics based on social theory can improve the performance by extracting useful supplementary social interaction information, not possible to be captured by other approaches, especially for large scale complex networks.

According to the information extracted from the aforesaid link recommendation metrics, there are many recently proposed learning-based link prediction methods that can be divided into feature-based classification, meta heuristic-based, matrix factor-