Accepted Manuscript

A link prediction method for heterogeneous networks based on BP neural network

Ji-chao Li, Dan-ling Zhao, Bing-Feng Ge, Ke-Wei Yang, Ying-Wu Chen

PII: S0378-4371(17)31264-5
DOI: https://doi.org/10.1016/j.physa.2017.12.018
Reference: PHYSA 18948

To appear in: *Physica A*

Received date : 2 May 2017
Revised date : 14 October 2017

Please cite this article as: J.-c. Li, D. Zhao, B. Ge, K. Yang, Y. Chen, A link prediction method for heterogeneous networks based on BP neural network, *Physica A* (2017), https://doi.org/10.1016/j.physa.2017.12.018

A Link Prediction Method for Heterogeneous Networks Based on BP Neural Network

Ji-chao LI, Dan-ling ZHAO, Bing-Feng GE*, Ke-Wei YANG, Ying-Wu CHEN

College of Information System and Management, National University of Defense Technology, Changsha, Hunan 410073, P. R. China.

* Corresponding author. Tel.:+86 73184573592. *E-mail address*: bingfengge@nudt.edu.cn (B. Ge).

Abstract: Most real-world systems, composed of different types of objects connected via many interconnections, can be abstracted as various complex heterogeneous networks. Link prediction for heterogeneous networks is of great significance for mining missing links and reconfiguring networks according to observed information, with considerable applications in, for example, friend and location recommendations and disease-gene candidate detection. In this paper, we put forward a novel integrated framework, called MPBP (Meta-Path feature-based BP neural network model), to predict multiple types of links for heterogeneous networks. More specifically, the concept of meta-path is introduced, followed by the extraction of meta-path features for heterogeneous networks. Next, based on the extracted meta-path features, a supervised link prediction model is built with a three-layer BP neural network. Then, the solution algorithm of the proposed link prediction model is put forward to obtain predicted results by iteratively training the network. Last, numerical experiments on the dataset of real examples of a gene-disease network and a combat network are conducted to verify the e□ectiveness and feasibility of the proposed MPBP. It shows that the MPBP with very good performance is superior to the baseline methods.

Keywords: Link Prediction; Heterogeneous Networks; Meta-Path; BP Neural Network

1 Introduction

Link prediction, an essential problem in link mining, aims to assess the existence probability of links in the future or predict missing links according to observed information [1]. In traditional link prediction problems, research generally focuses on predicting links of homogeneous networks that involves the same type of entities and links [2]. Related methods are generally twofold: unsupervised and supervised [3]. For the unsupervised methods, such as Jaccard's coefficient, common neighbors, Adamic-Adar index, preferential attachment index, random walk with restart, and Katz measure, most of them are only based on the structural similarity of nodes [4]. Supervised algorithms take advantage of predefined labeled information and attribute data of nodes and edges for predicting links [5-10].

However, most real-world systems, whether in social or military fields (e.g., human disease-gene networks or combat networks), are composed of multiple types of interacting components and should be modeled as heterogeneous networks. For example, disease-gene networks contain genes and diseases, with different types of links: gene-gene pairs, disease-disease pairs and gene-disease pairs [11]. The rich semantic meanings brought about by