Title: A Generalized Tree Augmented Naive Bayes Link Prediction Model

Author: Jiehua Wu

PII: S1877-7503(17)30140-0
DOI: https://doi.org/doi:10.1016/j.jocs.2018.04.006
Reference: JOCS 855

To appear in:

Received date: 6-2-2017
Revised date: 18-1-2018
Accepted date: 4-4-2018

Please cite this article as: Jiehua Wu, A Generalized Tree Augmented Naive Bayes Link Prediction Model, *<![CDATA[Journal of Computational Science]]>* (2018), https://doi.org/10.1016/j.jocs.2018.04.006

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Computer Science 00 (2018) 1–18

Procedia Computer Science

www.elsevier.com/locate/procedia

A Generalized Tree Augmented Naive Bayes Link Prediction Model

JIEHUA WU[a,b,]

[a]*Department of Computing, Guangdong Polytechnic of Industry and Commerce, Guangzhou, China*
[b]*School of Computer Sci. and Eng., South China University of Technology, Guangzhou, China*

Abstract

This paper studies link prediction, a recently emerged hot topic with many important applications, noticeably in complex network analysis. We propose a novel similarity-based approach which improves the well-known naive bayes method by introducing a new Tree Augmented Naive (TAN) Bayes probabilistic model. It makes better link predictions since the model alleviates the strong independency hypothesis among shared **common** neighbors to match the real-world situation. To obtain the latent correlation among common neighbors, we exploits mutual information to quantify the influence from neighbors' neighborhood. This yields a better performance than those methods which employing more local link/triangle structure information. In addition, the TAN **model** are easily adopted to other common neighbors-based methods such as AA and RA. Experimental results on synthetic and real-world networks show that our algorithms outperform the baseline methods, in terms of both effectiveness and efficiency.

Keywords:
Link Prediction, Complex Network, Common Neighbors, Tree Naive Bayes Model

1. INTRODUCTION

Complex networks are popular tools to understand the structure and evolution of various real world complex systems, such as the air traffic networks, protein-protein interaction networks and social networks. One core issue to address while studying complex network lies in evolutional regulation of relations among entities. Link prediction is a fundamental data mining and machine learning technique for studying and understanding such relations. The task of link prediction [1] is to infer the missing links between nodes in a given network. It not only offers insights into complex network analysis techniques such as community detection [2], but also play an essential role in the analysis of network diffusion [3]. Traditionally, an effective model for link prediction is to assign a similarity score to the potential node pair. The more similar the pair of nodes are, the more likely they are linked. Most similarity methods that make prediction can be characterized in terms of three basic categories: local node-based link prediction (e.g., Common Neighbors (CN) [1], Adamic Adar (AA) [4] Resource Allocation (RA) [5]), global path-based link prediction (e.g. Katz [1] , LP (Local Path) [5]), and random walk-based link prediction [6] (e.g., Superposed Random Walk (SRW) , Local Random Walk (LRW)).

Email address: wu.jiehua@mail.scut.edu.cn (JIEHUA WU)