A Novel Ego-Centered Academic Community Detection Approach via Factor Graph Model

Yusheng Jia, Yang Gao, Wanqi Yang, Jing Huo, and Yinghuan Shi

State Key Laboratory for Novel Software Technology, Nanjing University, China, 210023

Abstract. Extracting ego-centered community from large social networks is a practical problem in many real applications. Previous work for ego-centered community detection usually focuses on topological properties of networks, which ignore the actual social attributes between nodes. In this paper, we formalize the ego-centered community detection problem in a unified factor graph model and employ a parameter learning algorithm to estimate the topic-level social influence, the social relationship strength between nodes as well as community structures of networks. Based on the unified model we can obtain more meaningful ego community compared with traditional methods. Experimental results on co-author network demonstrate the effectiveness and efficiency of the proposed approach.

Keywords: Ego-centered Community, Community Detection, Factor Graph Model, Social Networks.

1 Introduction

Recently, quite a few methods [1][2][3] have been proposed for the macro-level community detection. In a practical application, the selection of experts for evaluating technology projects usually requires the community information according to one specific expert [4]. Therefore, the community detection task is not only to be modeled as a community detection problem globally from the aspect of the whole network. Ego network models [5] are utilized to represent the social network from ego perspective. Obviously, different experts have the specific research areas as well as the evaluating criteria. Thus, it is necessary to measure the social influence from different research areas. In this paper, we try to address the two major issues which are related to the social influence analysis, including 1) how to quantify the strength of social influence, and 2) how to differentiate the social influence from different research areas. The key point in ego-centered community detection is how to well quantify the influence among researchers by simultaneously leveraging relationship factor, social influence and community structures.
Fig. 1. An illustration of the ego-centered community detection using the co-author networks

Fig.1 illustrates the goal of our work. The input: a co-author network and an expert, (Ada is the ego in Fig.1). As shown in the left of the figure, we can first obtain the research areas distribution of the different experts, and then use ego-centered community detection method to measure the degree that the ego expert is influenced by other experts on each research topic as well as social relationship strength between them. The right of the figure shows the output, where the result reflects not only the closest relationship, but also the mutual influences between them on different research topics.

Our proposed approach for community detection has the following pipeline: (1) we first we get the research areas information by using Latent Dirichlet Allocation (LDA) [8][9] model to analysis the co-author network. (2) Ego Community Factor Graph Model (ECFG) combines the influential factors with the community structure to model the relationship of the networks. By finding the most relevant nodes which connected with the ego node based on the quantitative analysis, we can find out the ego community from the large social networks.

2 Related Work

Recent studies show that there has been a great interest in community detection. Pas- sarella et al. [5] proposed a constructive algorithm to generate the ego networks. The algorithm complemented with an analytical model which considered both the structural properties in the anthropology literature and the properties of their contact process.

Clauset [10] proposed an algorithm to find the local optimal community structures in large networks starting from ego vertex. The method used the greedy strategy to explore the network one vertex at a time where the added vertex maximizes the local community value.

Luo et al. [11] optimized and extended the local community definition, which introduced the connection density measurement to avoid the phenomena of outliers. A greedy strategy and two local search approaches (KL-like and the add-all) are also employed to improve the precision.

However, the discovered communities extracted with traditional methods contain many outliers and with high recall and low accuracy. The main problem is that the discovered communities can’t reflect the real state of community structures accurately