Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 349 – 358

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

An Analysis of Overlapping Community Detection Algorithms in Social Networks

J. Chitra Devi* and E. Poovammal

SRM University, Kattankulathur, Chennai 603 203, India

Abstract

In the field of research, Social Network Analysis is prevalent domain which pulls the attention of many data mining experts. Social network analysis is the specific field of sociology and anthropology. It shares a number of characteristics common to real network. Some real networks like Facebook, Twitter exhibit the concept of community structure within the network. Social network is represented as a network graph. Detecting the communities involves finding the densely connected nodes. Overlapping communities are possible if a node is a member of more than one community. This paper discusses various modularity based approaches on detecting the overlapping communities in the social networks. This work aims in providing the characteristics and limitations of modularity based overlapping community detection algorithms.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: Community Detection; Modularity; Overlapping Community; Social Network; Social Network Mining.

1. Introduction

As real world scenario is dynamic and evolving, Social network is used intensively in wide range of applications and represented as a network graph with nodes and edges. Nodes represent the individual users/actors/items/resources whereas edge represents the link/flow of interaction/relationship among the users. The arrangement of nodes and edges in a graph is coined as topology. Some common techniques involved in social network includes Recommendation Systems, Link Analysis, Expert Identification, Influence Propagation, Trust & Distrust Relationship Prediction, Opinion Mining, Mood Analysis and Community Detection[1–3]. Recommendation systems analyze the available data and suggest something the actor might be interested in. As a result a new link is introduced in the network. Recommending friends or recommending resources happens through collaborative filtering and Content-based filtering. Each node represents entity and the analyzes of links among the nodes gives the behavior pattern of different activities. Link analysis helps in finding this behavior pattern of a social network. Strategy of finding an expert in a required domain by analyzing the social network is coined as Expert Identification. With the help of previous log information, Influence Propagation between the nodes is identified. Two models are used in Influence Propagation namely Independent Cascade Model and Linear Threshold Model. Based on attributes of a node, a relationship or a link in a social network is identified as Trustable or not. This study is termed as Prediction of Trust or Distrust relationship.

*Corresponding author. Tel.: +91 -9940638001.
E-mail address: chitradevi21@gmail.com

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.082

Opinion mining involves building a system to collect and categorize opinions about a product or a person. It uses ideas of machine learning, text mining and natural language processing. The levels of opinion mining are document level, sentence level and phrase level. To track the mood of public about a particular product, mood analysis is performed. Set of actors interacting with each other frequently forms a Community. Detecting the community involves clustering of similar actors in social network graph. Different types of communities are node-centric community, link-centric community, network-centric community and hierarchy-centric community. Various community detection approaches have been proposed namely graph-based approach[4], link-based approach[5], agent and dynamic based approach[25,26], fuzzy-based approach[20–22] and modularity scoring based approach[6–19,23,24]. This paper analyzes the modularity based algorithms in detail.

2. Related Works

Social network is a network graph holding nodes with edges connected to it. Community detection aims to identify the community in the graph. Community is a module containing the set of nodes with major activities/interaction/ similarity among them. Overlapping communities are possible if any of the nodes has participated in the formation of different modules. Each node's strength or membership value in different modules varies accordingly. Various metrics are utilized to measure the strength of the community. Most popular of it is modularity measure. This modularity is the strength of partition of network as a community. Section 3 explains the modularity based community detection approach in detail.

3. Modularity Scoring based Clustering Approach

A community is said to be strong if it has more internal interaction within the community. To qualify the strength of a community, parameters like quality functions, density, conductance and modularity are used. This work reviews on the adaptation of the modularity measure in a undirected or directed, weighted or unweighted graph. Modularity defines the strength of partition of network into communities. It quantifies the association between a node and a community based on some parameters. In this paper, different version of belonging factor in modularity measure is considered. Belonging factor is a measure explaining the degree to which a node u belongs to a community C.

3.1 Modularity based disjoint community detection

Newman's modularity[6] measure considers the intra connection and inter connection between the nodes. As like other properties of communities say, node degree, betweenness, centrality or clustering coefficients, modularity measures the strength of formation of communities in a social network. This modularity is defined in equation (1)

$$Q = \frac{1}{2m} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{2m} \right] \delta(C_u, C_v) \tag{1}$$

where A_{uv} represents the adjacency matrix with row and column as node u and v respectively A_{uv} value is 1, if there is an edge between u and v and 0, otherwise. K_u, K_v is the degree of node u and v respectively, m is total number of edges in the graph $\delta(C_u, C_v)$ is represented as equation (2)

$$\delta(C_u, C_v) = \begin{cases} 1, & (u, v) \in C \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

At any particular time t, it can produce only two communities in a network. Hierarchical partitioning with two sub communities and further partitioned into two smaller communities. Thus only binary partition is possible. It limits its application only to an undirected and unweighted graph in a social network.

In order to overcome the limitations of Newman's modularity, Leicht *et al.*[7] defines the modularity measure for directed graph. The extended definition is the null-model directed graph with P_{uv}. This P_{uv} is the probability of having the link starting from node u to node v. Generally in directed graph, link from node u to node v is different from the