

Community Detection Using Random Walk Label Propagation Algorithm and PageRank Algorithm over Social Network

¹Monika Kasondra, ²Prof. Kamal Sutaria,

¹M.E. Student, ²Assistent Professor ,

¹Computer Engineering Department,

¹V.V.P Engineering Collage, Rajkot, India

Abstract— Now a days, human being become more social, because it is allows people to extend their social life in unprecedented ways. People become more friendly with each other on social sites like Facebook, Twitter, etc, based on their interest, likes etc. In social network, having same interest have the same community. So, detecting and evaluating the community structure is an essential task in social network analysis. In this paper, we are going to discover hybrid approach of Random Walk Label Propagation algorithm and PageRank algorithm to defined best community over social network. Main issues to define community time complexity, scalability, heterogeneity, to improve this issues, hybrid approach has been proposed. Here detect the communities based on seed node

Index Terms— community detection, seed node, social network

I. INTRODUCTION

There are verities of the social networks, such as Facebook, Twitter, Amazone, LinkedIn, etc. We are growing with the information age. Now a day most of the people connected with this type of social network. With the development of the smartphones, more and more people log into their social networks through their smartphones and share text and multimedia information with their friends online.

The social network is usually modeled as graphs. A social network consists of a set of nodes along with edges connecting the nodes. The nodes represent the object in social networks, such as people, commodities, etc. The edges represent the relationships between objects.

The community is an important structure in social networks. Basically, communities are set of nodes with higher edge density than the null model. In recent year, find seed nodes and how to use the seed nodes for community detection has become a hot topic in social network analysis and field of data mining.

In social network analysis, finding seed nodes have a wide range of applications. E.g., if we can dig out the most influential customers in marketing, then through the community structured by the seed nodes, a product brand can be rapidly promoted.

In this paper, we show the different methods, which describe how can detect communities based on the seed nodes over the social network.

II. COMMUNITY DETECTION

Detecting and evaluating the community structure of real-world graphs constitutes an essential task in the area of graph mining and social network analysis. The network contains many structures and tightly connected groups. Also referred as community, clusters, modules, etc.

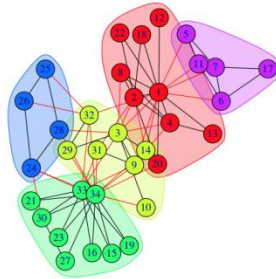
The task of community detection is to find sets of nodes with lots of connections inside the sets and few edges outside the set.

In figure 1, depicted the different communities over the social network. Different color defines different communities and number specify node id.

There are many available approaches for Community Detection^[5]:

- 1) Group-based approach
 - 2) Network-based approach
 - 3) Propagation-based approach
 - 4) Hierarchy-based approach
- Seed-centric approach

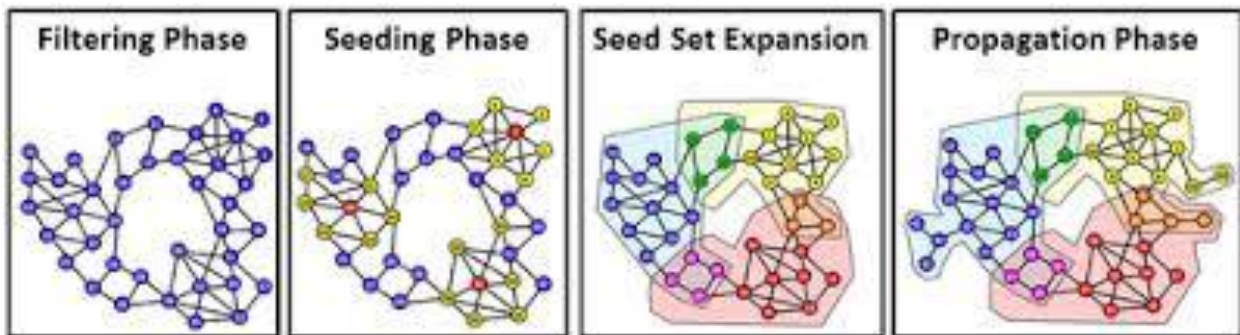
In this paper, we adopt a seed-centric approach to detect different communities over the social network.

Figure, 1. Communities in social network ^[4]

III. RELATED WORK

Seeding phase ^[2]

In seeding phase, which consists of four phases: filtering phase, seeding phase, seed set expansion phase and propagation phase.

Figure 2. Seeding phase ^[1]

Filtering phase

In this phase, to remove regions of the graph that are trivially separable from the rest of the graph, so will not participate in overlapping.

The output of the filtering phase is the biconnected core graph where whiskers (subgraphs connected to the biconnected core) are filtered out. It removes regions of the graph that are clearly partitionable from the remainder. More importantly, there is no overlap between any of whiskers. This indicates that there is no need to apply overlapping community detection algorithm on the detached regions.

Seeding phase

After we get biconnected core graph, we find seeds in filtering phase. The goal of seeding strategy is to identify a diversity of vertices that lie within a cluster. The output of this phase is seed nodes over a graph.

Seed set expansion

In this phase, expand the seed set using a personalized PageRank clustering scheme. A personalized PageRank vector, followed by a sweep over all cuts induced by a vector, will identify a set of good conductance within the graph. The set identified via this procedure has a conductance that isn't too far away from the best conductance of any set containing that vertex.

Propagation phase

After seeding phase, propagation phase will apply. In this phase, once we get the personalized PageRank communities on biconnected core graph, we further expand each of the communities to the regions that we detached in the filtering phase. For each detached whisker connected via a bridge, we add that piece to all of the clusters that utilize the other vertex in the bridge. The output of this final phase of seed is different communities of the graph.

In the Random Walk Label Propagation algorithm ^[2], first this algorithm does not required number of node as well as size of community, so it is random not static. To find the important node over social network it uses the position probability of the random walker. This algorithm following the depicted steps:

- 1) Calculate the importance of each node
- 2) Initialize core node and select seed node from it
- 3) Do the label propagation and detect the community

Same as in PageRank algorithm^[3], is a useful algorithm for finding communities over social network. To obtain the overall community structure of the network, personalized PageRank should be executed amounts of times. A personalized PageRank vector is the stationary distribution of a random walk that follows an edge of the graph with probability α . The essence of the induced community is that an inexact personalized PageRank vector, computed via an algorithm that “pushes” rank round the graph, will identify good bottlenecks nearby a seed vertex.

IV. PROPOSED WORK

To find community here proposed method is depicted.

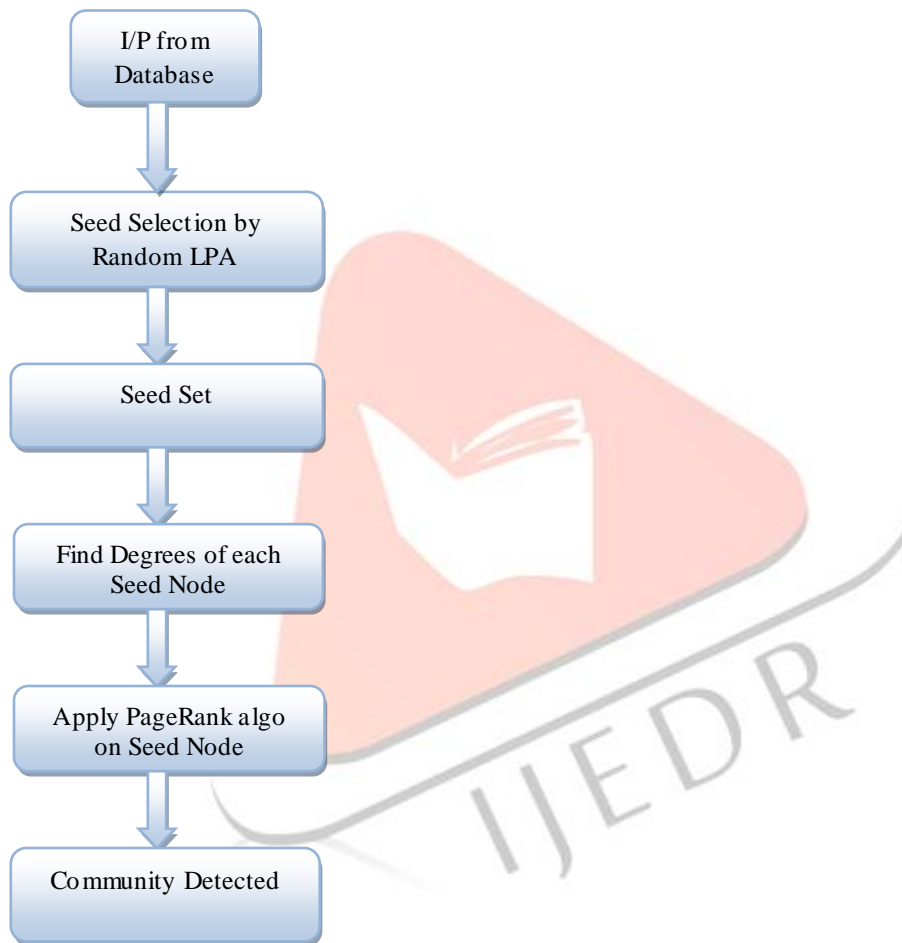


Figure.3 Proposed System

Here, as per depicted figure, first take input from the database dataset, which is in form of graph. Then apply Random Walk Label Propagation Algorithm by conductance parameter. Find all node which having maximum conductance, declare all node as a seed node. Now find the InDegree and OutDegree of all the seed node which are lies in seed set. Then apply PageRank Algorithm over seed set. It will chose maximum InDegree or OutDegree of each seed node which are best seed node and there is high probability having best community. And at last explore the neighborhood and find the communities.

The experiment had done in different three dataset. We measure here the value of conductance and then based on this value find the community.

Dataset	Community Detected	Conductance
---------	--------------------	-------------

Karate Club Network	4	0.4188
American Football Network	10	0.6043
Dolphin Network	6	0.5277

Table 1. Conductance value with dataset

Based on this value here is a graphical representation.

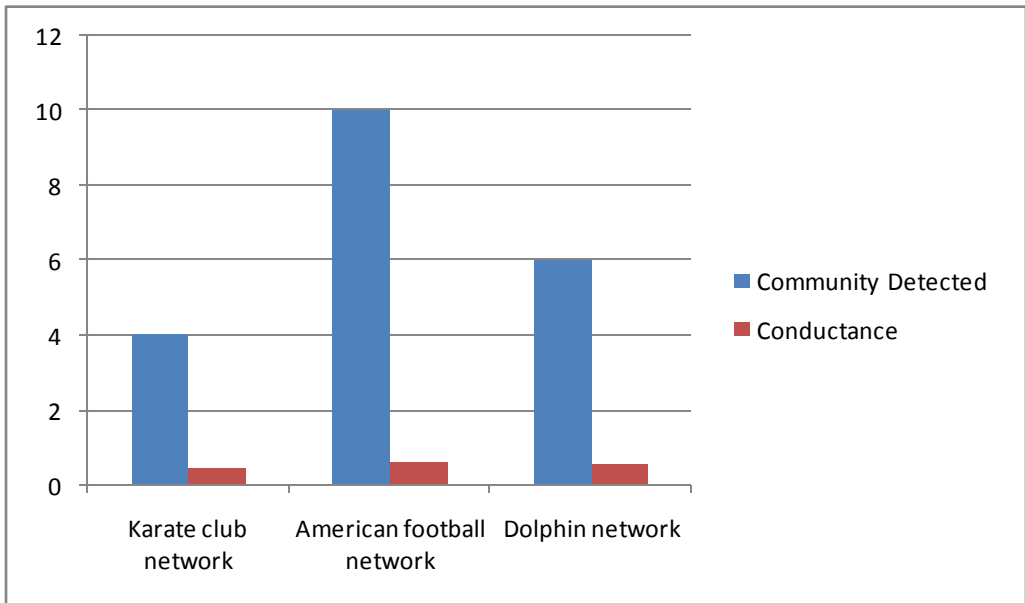


Figure. 5 Graphical representation

We can see that in Table 1, there is a Karate Club Network having conductance value and detect 4 communities which is shown as bellow.

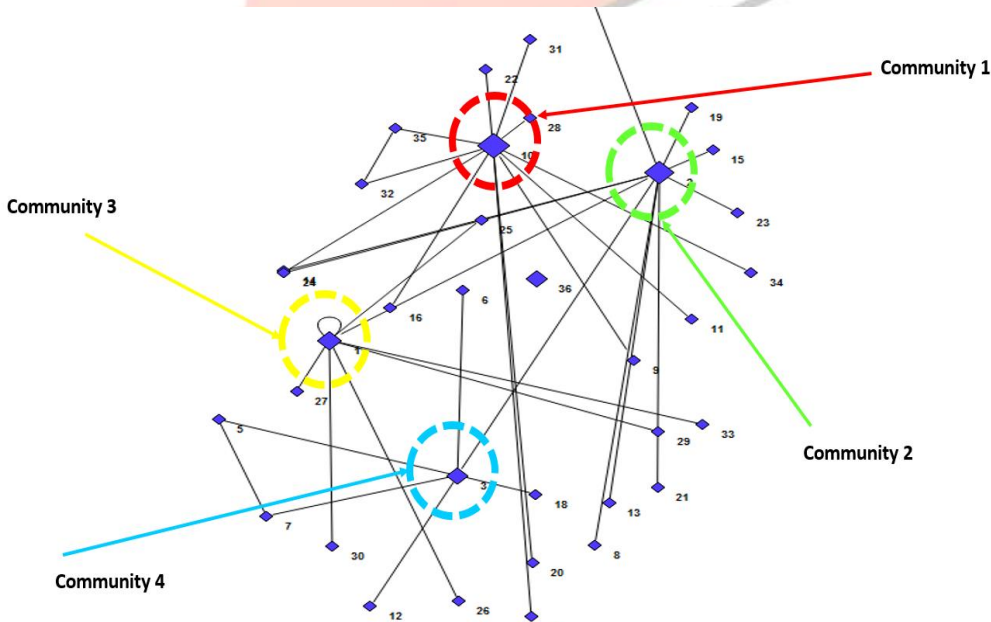


Figure. 6 Community Detected

V. CONCLUSION

In this paper we conclude that, having maximum conductance value we gain best community over social network. Here we tested on three different dataset and find the communities. So this combine approach of Random Walk LPA and PageRank Algorithm gives the best communities.

VI. REFERENCE

- [1] <https://www.google.co.in/search?q=seed+phase+in+social+network>
- [2] Su Chang, Jia Xiaotao, Xie Xianzhong, Yu Yue, "A new random-walk based label propagation community detection algorithm", IEEE – International conferences on web intelligence and intelligent agent technology 2015.
- [3] Fei Jiang, Yang Yang, Shuyuan Jin, Jin Xu, "Fast Search to Detect Communities by Truncated Inverse PageRank in Social Networks, IEEE International conference on Mobile service – 2015.
- [4] <https://www.google.co.in/search?q=social+network>
- [5] <https://www.google.co.in/search?q=community+detection>

