ISSN: 2321-7782 (Online)
Volume 2, Issue 1, January 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

A Survey on Community Detection in Weighted Social Network

Hitesh R. Gor[1]
M.Tech. Comuter Engineering
R.K. University
Rajkot - India

Maulik V. Dhamecha[2]
Computer Engineering Department
R.K. University
Rajkot - India

Abstract: Social Network is an indivisible part of our society whether it may be related with humans or any group of animals or creatures. By analysing them, one can get idea how they are related to each other. When it comes on society, relation will come and generally social networks are with weight. One more thing is that in social network community and its detection is also with same importance. So the combination of community in weighted networks and its detection is one of the major challenges in SNA (Social Network Analysis). So here we have included a survey on recent algorithms which works on Community detection especially on weighted networks. By doing the survey, we have tried to find the quality of methods and their performance issues.

Keywords: Social Network; Social Network Analysis (SNA); Community; Community Detection; Modularity; Weighted Network;

I. Introduction

Social Network is a collection of nodes that may represent individuals, organizations, communities, employees, in computational field – web pages, computer systems, servers etc. and those nodes are connected with the help of edges like in mathematical graph. Such edges may be called as ties, lines, connections. Those edges connected from one node in the network to the other may represent relation, kinship, friendship, social relation, blood relation, internet connection, telephone calling links, common interests, profession etc.

In short, social network is simply combination of mathematical graph and non-separable aspects of our society. It represents how we are related with our environment, with our day to day life, our society. The same relation and connection concept can be applied on computational relation between different nodes, server connections, hyperlinks, bus connections in computer system, community connections on social portals. Though the application of these fundamentals onto the computational systems is not so old, we can see this idea as one of the oldest concept of our i.e. human history. From when in the early past, we had started our families and community or society structure, we can consider the Social Networking as a rising and undividable concept for our human community.

With the advantage of ability to represent the social network as a simple graph, we can analyse all the previous discussed topics with the help of implementing those aspects on to this graph representation. This whole concept is known as Social Network Analysis.

II. Social Network Analysis

We think that Social Network is the concept which interacts only with human society. But this is not the real thing. Social Network comes into the picture whenever it comes on relation whether it would be for humanity or non-human groups [1]. In every SN (Social Network) there would be same fundamentals as we apply in our case. Connection with each other, types of relations, weight of those relations, according to that community structures and so on.

A. Basic Definitions

As a combination of graph theory and social network, there are different terms or definitions which can be used to understand or illustrate structure or characteristics of a network. This gives idea about importance of a vertex in the graph or network. For example how important a person is in network, how much a road is being used or what is the important of the department in an organization? Such measures or definitions are

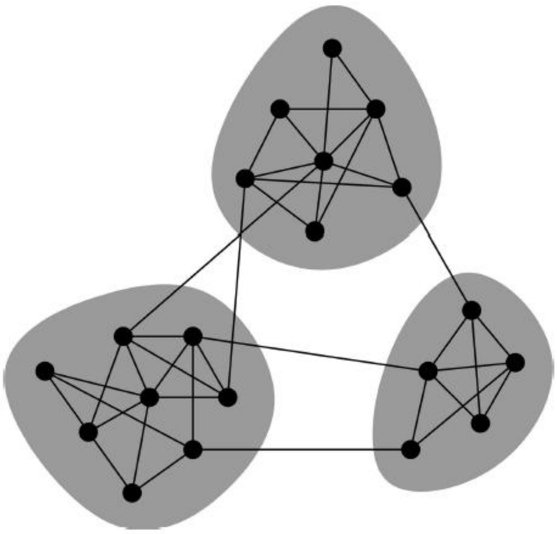


Fig. 1 Example of a network with its communities

Degree Centrality:

It is defined as the number of links incident upon a vertex which means the number of edges a vertex has.

For a graph $G := (V, E)$ with n vertices, the degree centrality $C_D(v)$ for vertex is:

$$C_D(v) = deg(v) / (n-1)$$

Betweenness Centrality:

Vertices that occur on many shortest paths between other vertices have higher betweenness than those that do not.

For a graph $G := (V, E)$ with n vertices, the betweenness $C_B(v)$ for vertex is computed as follows:

1. For each pair of vertices *(s, t)*, compute all shortest paths between them.
2. For each pair of vertices *(s, t)*, determine the fraction of shortest paths that pass through the vertex in question (here, vertex v).
3. Sum this fraction over all pairs of vertices *(s, t)*.

Closeness Centrality:

In graph theory closeness is a centrality measure of a vertex within a graph. Vertices that are 'shallow' to other vertices (that is, those that tend to have short geodesic distances to other vertices within the graph) have higher closeness. Closeness is preferred in network analysis to mean shortest-path length, as it gives higher values to more central vertices, and so is usually positively associated with other measures such as degree.

In the network theory, closeness is a sophisticated measure of centrality. It is defined as the mean geodesic distance (i.e., the shortest path) between a vertex v and all other vertices reachable from it:

$$(\sum dG(v, t)) / (n-1), \text{ where } n > 2.$$

B. Community Structure

For a SN if we want to analyse its properties and characteristics, we need to analyse its structure. How it has been evolved, how and why each node is connected with the other etc. So for this reason analysis of the structure of a SN is important [12].

Community generally gives the organized structure of the Social Network. It includes how many numbers of people like the match of hockey on TV. How like vegetarian food as a regular diet. It gives hints about special relationships between people or web pages or any other who symbolize the nodes. Through such information, we can get the main information very easily [3]. Biochemical pathways in metabolic networks, the way of relation between web pages and same topic orientation of those web pages such information can also be obtained from analysis of communities on Social Network [13].

Fortunato [14], *Newman and Girvan* [2] give really good definition of community. They say that, Complex systems are usually organized in compartments, which have their own role and/or function. In the network representation, such compartments appear as sets of nodes with a high density of internal links, whereas links between compartments have a comparatively lower density. These sub-graphs are called communities, or modules, and occur in a wide variety of networked systems [2, 3]. Due to such importance, community detection is very important task to tackle. But there are some problems with community detection. The next remainder portion is the collection of understandings of community, its detection and problems related to it. And in the remaining sections are full of analysis and observation of a few methods or algorithms for community detection for weighted networks [11].

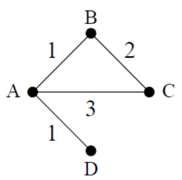
To detect such community structure there are various methods starting from traditional to the moderns [2, 12]. Basically we can divide traditional methods into two types: one for sociology and the other is for computer science. Especially for computer science approach, we can put *spectral bisection method* [3] and *the Kernighan-Lin algorithm* [4]. For sociological approach it is hierarchical clustering. In hierarchical clustering there are more two approaches as *single linkage* and *complete linkage* [5].

III. Methods for Community Detection in Weighted Networks

In most of the networks it is assumed that the edges which connect vertices with each other are weightless or more precisely we can say with equal weight age. But this may not in the real world applications. The former assumption says:

$$A_{ij} = \begin{cases} 1 & \text{if i and j is connected,} \\ 0 & \text{otherwise.} \end{cases}$$

But in the case of a weighted graph, its matrix can be described as a collection of numbers which gives value of weight for the relation between i and j vertices. For example:



$$\equiv \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 3 & 1 \\ 1 & 0 & 2 & 0 \\ 3 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \end{array}$$

So now we come on the main topic. How we can find community structure or communities in the given graph if the graph is weighted (here each edge has some predefined weight age as described earlier). There are different methods which try to detect community structure in weighted networks [11].

A. Center-based Community Detection for Weighted Networks [3]

In any network there must be some important nodes. On the basis of this idea initially we can start from random selection of centers. Then we need to rearrange (reallocate) them and after it, we will go for community detection. The reallocation of the selected centers can be done on the basis of predefined criteria. They have used three main steps in the algorithm. They are:

Steps:

1) Detect centers: we use both weighted information and structural information to find centers of the network, which are connected with many nodes (high degree) and have a high intensity (large weight). The nodes are often the core of a certain community.
2) Adjust centers: according to some measure, we pre-assign some eligible non-central nodes to a community represented by a center, and then adjust centers' distribution according to a criterion.
3) Detect community structure: after digging out a reasonable distribution of the centers, we assign all the nodes to the corresponding community according to same measure in step 2), thus containing the final community structure.

To perform these steps they have given three different algorithms.

1) PageRankCentrality()
2) NodeSort()
3) CenterAdjust()

Performance of these algorithms is given in the comparison table.

B. Community Detection for Weighted Networks using Intra and Inter Centrality [4]

Here Intra and Inter Centrality is introduced by the authors to distinguish Community detected and the rest of the nodes of the network

Intra-Centrality:

The Intra-centrality of a node is defined as the number of shortest paths between pairs of nodes in the same community that go through it.

Inter-Centrality:

The Inter-centrality of a node for two communities is defined as the number of shortest paths between two nodes in these two communities that go through it.

Based on these two measures data forwarding in Delay Tolerant Networks (DTNs) and worm containment in Online Social Networks (OSNs) has been tested. As a result it has been proved that the proposed I^2C algorithm gives better performance with compare to other existing algorithms.

C. Weighted Newman Fast Algorithm [5]

They have used very well known and famous algorithm for community detection i.e. The Newman Fast algorithm. The modification from their side is that they have extended the algorithm up to the weighted networks. Using various undirected graph (Computer generated and real world networks also), they have used *Modularity* [7, 10] that is on what the Newman Fast algorithm works.

"Modularity" Q is a function that measures how much a particular partition $V = \cup V_q$ of the set of vertices V of the network is meaningful. Q is defined as follows. Let e_{ii} be the fraction of the edges in the network that connects vertices inside V_i and for $j \neq i$ let $e_{ij} = e_{ji}$ be half of the fraction of the edges in the network that interconnect vertices of V_i and V_j. Thus, $e_{ij} + e_{ji}$ is the total fraction of *these intra-cluster edges*. For the testing purpose, they have used *Girvan - Newman* [8] computer generated networks. The rest of the information is given in the comparison table.

D. Arc Label Propagation (ALP) to detect Community for Weighted Networks [9]

In this paper, they have proposed the notion of triangular random walk as a way to unveil arc-community structure in social graphs: a triangular walk is a random process that insists differently on arcs that close a triangle. They have proved that triangular walks can be used effectively, by translating them into a standard weighted random walk on the line graph; their experiments show that the weights so defined are in fact very helpful in determining the similarity between arcs and yield high-quality clustering.

Even if the proposed technique gives a weighting scheme on the line graph and can be combined with any node-clustering method in the final phase, to make their approach more scalable they also propose an algorithm *(Arc Label Propagation - ALP)* that produces the clustering directly without the need to build the weighted line graph explicitly. The ALP is base on *Label Propagation* suggested by *Raghavan et.al.* [9]. Their experiments show that ALP, besides providing the largest accuracy, it is also the fastest and most scalable among all arc-clustering algorithms we are aware of.

IV. Analysis

As an analysis we have compared all the four methods with each other. Their pros and cons are given in the *Table I*. Performance of all the algorithms are given for the sake of performance comparison.

TABLE I
Comparison of different Algorithms

Sr. No.	Name	Ref.	Description	+ve Points	-ve Points	Performance
1	3 in 1 Algorithms	[3]	• Steps: 1) Detects Centre 2) Adjust Centre 3) Detect Structure	✓ Stable for Weighted & Large Networks	✓ Like *K-Means* number of communities has to be known / preset	$O(n^2)$
2	I^2C	[4]	• Works on: 1) Intra Centrality 2) Inter Centrality • Tested on OSNs (Online Social Networks) and DTNs (Delay Tolerant Network)	✓ Better performance for OSNs. ✓ Low infection rate for DTNs.	✓ Slightly high delivery time.	O(mn) here, m = number of edges n = number of nodes
3	WNF	[5]	• Uses Newman Fast Algorithm	✓ Low computation time ✓ Good for large networks	✓ For directed networks it becomes complex	O(mn) here, m = number of edges n = number of nodes
4	ALP	[9]	• Uses: 1) Arc - clustering 2) Triangular Random Walks • Follows Label Propagation Algorithm [9]	✓ No need to build weighted line graph explicitly. ✓ Works well for dense graph.	✓ Complex algorithm to implement	O(mn) here, m = number of edges n = number of nodes

[3] is stable for weighted networks but because of having similar characteristics like *k-means* if we don't know the number of communities, it is worthless for us. I^2C is also good with respect to OSNs and DTNs if it come on performance and infection rate respectively but its delivery time is high for DTNs. WNF is one of the best algorithm with low computation time and better performance for large networks. But for directed networks it becomes poor. For ALP it is best for dense graph but for implementation purpose it is complex.

V. Conclusion

After comparing all the algorithms we can have a clear idea about methods used in those algorithms and performance of all the algorithms. As no one is perfect in this world, every algorithm has good as well as bad qualities. Performances of all the algorithms are very similar but according to their good and bad qualities one need to select them for one's purpose. All the algorithms are best with respect to various aspects but we need to select them according to the point where they fall to perform (i.e. their disadvantages).

References

1. David Lusseau and M. E. J. Newman, *"Identifying the role that animals play in their social networks"*, Proceedings of The Royal Society, Biological Sciences, 2004.
2. M. E. J. Newman, *"Detecting community structure in networks",* Eur. Phys. J. B 38, pp. 321-330, 2004.
3. Jie Jin, Lei Pan, Chongjun Wang, Junyuan Xie, *"A Center-based Community Detection Method in Weighted Networks"*, 23rd IEEE International Conference on Tools with Artificial Intelligence, pp. 513-518, 2011. //// 3 in 1 algos
4. Zongqing Lu, Yonggang Wen, Guohong Cao, "*Community Detection in Weighted Networks: Algorithms and Applications"*, 2013 IEEE International Conference on Pervasive Computing and Communications (PerCom), San Diego, 2013.
5. Alireza Khadivi and Martin Hasler, *"A Weighting Scheme for Enhancing Community Detection in Networks",* Communications (ICC), IEEE International Conference , pp. 1-4, 2010
6. M. E. J. Newman, and M. Girvan, *"Finding and evaluating community structure in networks"*, Physical Review E, vol. 69, no. 2, pp. 091-113, 2004.
7. M. Girvan, and M. E. J. Newman, *"Community structure in social and biological networks",* Proceedings of the National Academy of Sciences, vol. 99, no. 12, pp. 7821-7826, 2002.
8. Usha N. Raghavan, Réka Albert, and Soundar Kumara, *"Near linear time algorithm to detect community structures in large-scale networks"*, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 76(3), 2007.
9. Paolo Boldi and Marco Rosa, *"Arc-Community Detection via Triangular Random Walks"*, 8th Latin American Web Congress, pp. 48-56, 2012
10. M. E. J. Newman, *"Modularity and community structure in networks"*, PNAS, vol. 103, no. 23, pp. 8577-8582, 2006.
11. M. E. J. Newman, *"Analysis of weighted networks"*, Phys. Rev. E 70, 056131, 2004.
12. Aixiang Cui, Duanbing Chen, Yan Fu, *"Community detection based on weighted networks",* Network and Parallel Computing (NPC), IFIP International Conference, pp. 273-280, 2008.
13. Yachana Bhawsar, Dr. G.S.Thakur, *"Community Detection in Social Networking"*, Journal of Information Engineering and Applications, Vol.3, No.6,pp. 51-52, 2013.
14. Andrea Lancichinetti and Santo Fortunato, *"Community detection algorithms: A comparative analysis"*, Phys. Rev. E 80, 056117, pages 11, 2009.

AUTHOR(S) PROFILE

Hitesh R. Gor received the B.E. degree in Computer Engineering in 2010 from V.V.P. Engineering College, Rajkot, Gujart, India. Currently he is pursuing M.Tech. in Computer Engineering from School of Engineering, R.K. University, Rajkot,Gujarat, India. His areas of interest are Social Network Analysis, Community Detection and Outlier Analysis. Currently his research area includes Overlapped Community Detection in Weighted and Directional Social Network.

Maulik V. Dhamecha has received his B.E degree in Computer Engineering from VVP Engg. College, Saurastra University, Gujarat, India and master Degree from Dharmsinh Desai University, Gujarat, India. He has been working at Department of Computer Engineering, School of Engineering, RK. University, Rajkot, Gujarat, where he is currently working as an Assistant Professor. His current research interest includes Multiple Classifier System, Sequence Pattern Mining, clustering.