

EIJO Journal of Engineering, Technology And Innovative Research (EIJO–JETIR) Einstein International Journal Organization (EIJO) Available Online at: www.eijo.in

Volume – 2, Issue – 3, May - June 2017, Page No. : 16 - 27

Community Structure and Central Node Identification through Leadership Function and Its Application

Renuka Anil Raut, Department of Computer Engineering, RSSOER, JSPM, NTC, PUNE

Email: raut.renuka15@gmail.com

Prof.Mr.R.H.Kulkarni, Department of Computer Engineering, RSSOER, JSPM, NTC, PUNE

Email: rkpv2002@gmail.com

Abstract

Complex network is defined as the network having large no. of nodes with complex relation (huge connectivity) among themselves. Mining communities in complex network is very important for analyzing the complex network. Many existing systems are not able to restore relation of nodes in network and do not maintain mining efficiency and community quality. A dynamical system to provide efficient mining and its membership in the community has been designed in the previous research paper, "Fast and Accurate Mining the Community Structure: Integrating Center Locating and Membership Optimization". In it, community detection algorithm to identify the central node and related network communities has been implemented. The designed algorithm is very effective and efficient to detect network communities in identifying most accurate central nodes of simple as well as complex network. Identifying centers of the communities is very important in analyzing the properties of the complex networks. In determining the membership in structure, the previous classical optimization and heuristic methods iteratively update the membership but cannot provide the optimized result. In the previous research paper, an effective mechanism to reveal the community structure by identifying the central nodes in the network has been implemented. It also optimizes the mining efficiency and the community quality and provides more efficient way for complex network.

In the enhancement part, a secure routing algorithm jointly optimizing underlay and overlay paths using key predistribution schemes has been implemented where data transmission between to nodes can take place.

Keywords: Community mining, complex networks, hierarchical structure, dynamical systems.

Introduction

A great number of natural and manmade systems are made up of a greatly sized number of connected dynamical units which can be well represented as complex networks. Network points are made by network units present in network and edges are made by network connections. Examples are everywhere, such as the net, the World Wide net, smart networks, having transport infrastructures, metabolic footways, protein effects on one another, human relationship and scientific collaborations. Community mining is an important work in complex network observations which aims at grouping the network points into so called parts of a greater unit or clusters such that they are thickly connected as to the rest of the network. This work often has relation to a group discovery. Though it is complex, mining or sensing network communities has wide applications, because nodes in the same group generally have as owner common (to 2 or more) properties or relationships in comparison with the network points interconnecting different groups of persons. For example, members in a grouping circle may have common interests, a group of closely connected proteins often work

together for a given functional realization, and tweets under the same topic always have the similar opinion. There is a great amount of facts supporting that community structures are very helpful for developing compared intelligent services, such as developing tools for precision marketing [2], [6], [5], identifying target points for drug discovery [3], [7], [8], searching and mining online social networks for trend predictions [4], [6], [9], [10]. A common characteristic of a real network is a hierarchical structure following the parent-child relationship among the nodes. Important nodes are always seen in the center. Therefore, Center node with the highest influence (connections) can form a loc. Just like k-means, a prototype based simple clustering technique which attempts to find the user-specified k clusters, a community can be also seen as a sub-graph, in which each node is closer to the given center (leader) than to others. However, similar to the well known distribution systems, such as random work dynamics [11] and consensus systems [12], k-means [13], [14] cannot provide the strict optimized result (by maximizing a given quality function which is used in many community detection algorithms). Therefore, given the central nodes of a network, how to identify the optimal communities they lead is a new and important topic

Review of Literature

As of now a great amount of study has been done on community structure. It consists of mainly two components: The quality function definition used to measures the importance of detected community partition, and The corresponding algorithm with which more efficient quality function can be produced.

In identification of central nodes, very first the concept of Modularity [16] was developed. It is one of the most wellknown quality functions. The Modularity can be explained as the comparison of actual volume of edges in a sub graph to the expected volume under some null hypothesis. More the volume of edges, better the community partition. Tanmoy et al. [17] proposed a novel community detection approach. It uses the distribution of inter-cluster connections from a community in its neighboring communities. This method is more efficient to produce the community structure than the existing modularity maximization algorithms. The classical Potts model can be also be used to design community structure, the idea behind which is to consider the communities as the spin states in the Potts model, and use the system energy to evaluate the quality of a candidate partition.

There are many extension methods and quality functions based on Potts model, such as the Reichardt Bornholdt model(RB) [32], the Erdos-Renyi null model,RBER [21], the configuration null model(RBCM) [21], Hofman and Wiggins(HW) model [22] and the Ronhovde and Nussinov (RN) model [23].

Most recently work from Aldecoa et al. [24] showed that no algorithms perform perfectly when applied on standard artificial benchmark networks, the results evaluated by their self-imposed quality functions often draw the contrast conclusion. Then, they proposed a new quality measure called Surprise, and proved that maximizing Surprise function is able to precisely reveal the community structure of real-world networks.

For algorithms, the Newman Fast (NF) method [16] is well-known for Modularity optimization, using a greedy approach to maximize the Modularity value. To unfold the community structure in large networks rapidly, the Louvain algorithm is proposed by Blondel et al. [15]. The algorithm starts by assigning a sole community label to each node Then, the algorithm performs two steps iteratively until the value of modularity cannot increase anymore.

In Ref. to [19], a new method for community detection called OCR has been proposed benefiting from partial synchronization of densely connected vertices in dynamical networks.

Label propagation is another famous algorithm for community detection [18]. Briefly, the algorithm starts by randomly assigning a community label to each node. Then, each node updates the label with the one which is most used by its neighbors. It has been also shown that the label propagation method is equivalent to finding the local energy minima of a simple zero temperature kinetic Potts model [20].

Sr. No.	Algorithms	Time Complexity
1.	CNM[33]	O(n, log2n)
2.	External Optimization, DA [06]	$O(n2, \log n)$
3.	Louvain[21]	O(n, log n)
4.	OCR-HK[32]	O(n2)
5.	Bayesian inference[31]	O(n, log an)
6.	Variational Bayian[22]	O(n1.44)
7.	RN Potts model[23]	O(n1.3)
8.	Label Propagation model[18]	O(n+m)
9	Community Detection Algorithm	O(n)
	(Existing Algorithm)	

The time complexities of various algorithms developed in this area of research are given below.

From the above table, it is clear that with every newly developed algorithm, along with accuracy, the speed of producing result has increased.

Problem Statement

Identifying the centers of the communities is very important to analyze the properties of the complex networks. In determining the membership in structure, the previous classical optimization and heuristic methods iteratively update the membership but cannot provide the optimized result. Hence accurately identifying center nodes and relevant community structure is very important.

Community structure is also very essential in understanding of functional properties of complex networks. It is in turn used for new tools, ideas, products, services, etc. Hence study of center node and relevant community structure identification has become an important task.

Existing Algorithms

1. Community detection algorithm via automatically determine the parameters

Input: A network G with n nodes and m edges; the maximum number of iterations R_{max}

Output: The community membership matrix *X*;

1: Initialize X(0) (described under section "Initiate the community configuration") and set X_{best} .

.....

2: repeat

3: Update n_{μ} , l_{μ}^{in} , l_{μ}^{out} , p_{μ}^{in} , and p_{μ}^{out} , using equations (2), (3), (4), (5), and (6), respectively.

- 4: Calculate $\delta H(t) / \delta x_{i\mu}(t)$ using Eq.(7),
- 5: Use Eq.(8) to update the membership matrix X(t)
- 6: Calculate the log-likelihood function LP(t) using
- Eq.(9)
- 7: **If** *R*_{max} is reached;
- go to step 8;

otherwise,

- go to step 1
- 8: Consider the X(t) with maximum LP(t) as the X_t
- 9: If $LP(X_t) > LP_{best}$, set $LP_{best} = LP(X_t)$ and $X_{best} = X_t$
- 10: If Rmax is reached
 - report X_{best} ;

otherwise,

go to step 1.

Initiate the community configuration

In the step 1 of Algorithm 1, the membership matrix $x_{i\mu}$ of regular nodes should be initialized (for centers, their membership vectors are first assigned to be the unit vector and do not vary anymore). In the following an efficient iterative way to initialize the matrix X is introduced. In the first time, it is initialized that

 $x_{i\mu}(0) = 1/c + y_{i\mu}; i = \{1, 2, \dots, n\}, \mu = \{1, 2, \dots, c\}, s.t. \sum y_{i\mu} = 0$ (1)

where $y_{i\mu}$ is a small Gaussian noise and this strategy will avoid the second type trivial solution . However, for the next times, an improved strategy is proposed which benefits from the best results of the previous iterations, i.e.,

 $\mathbf{x}_{iu}(0) = x_{iu}^{best} + y_{i\mu}$; $i = \{1, 2, \dots, n\},$

 $\mu = \{1, 2, \dots, c\}, \text{ s.t. } \sum y_{i\mu} = 0$

where x_{iu}^{best} is the (i, μ) entry of Xbest.

Using X(t), we can rewrite n_{μ} , l_{μ}^{in} , l_{μ}^{out} , p_{μ}^{in} , p_{μ}^{out} as

 $n_{\mu}(t) = \sum x_{i\mu}(t)$ (2)

 $l_{\mu}^{\ in}(t) = \frac{1}{2} \sum x_{i\mu}(t) \ x_{j\mu}(t) A_{ij} \qquad(3)$

Definition of above parameters is given below

C = the no. of communities

n =the no. of nodes

m = the no. of edges

 μ = the label of a community

 l_{μ}^{in} = the no. of intra -connection edges in μ

 $l_{\mu}^{~out~}$ = the no. of inter-connection edges in μ

 p_{μ}^{in} = the ratio of intra-cluster edges of community μ

 p_{μ}^{out} = the ratio of inter-cluster edges of community μ

 $x_{i\mu}$ = the probability of node i belonging to μ

 $k_{i\mu}$ = Node-Community degree of node i and μ

 $Q_{i\mu}$ = the fitness function of node i upon community μ

k(i) = the leadership function of node i

 d_{ij} = the shortest path distance from node i to j

 δ = the influence factor of leadership function

 R_{max} = maximum number of iterations

 $\not=$ for All

 \exists = there exists

 $y_{i\mu}$ = small Gaussian noise

LP(t) =log-likelihood function

Proposed System

With the existing algorithm, we obtain community structures with center nodes. Taking this as an input, we can transfer the data between the two nodes securely.

For this purpose, a secure routing algorithm jointly optimizing underlay and overlay paths using key pre-distribution schemes but not requiring explicit trust of other network nodes is developed. Modeling a network using key pre-distribution schemes with directed and weighted graphs, proposing a boolean LP problem (Linear programming) for optimal overlay routing in the resulting network graph, Analytically reducing the boolean LP problem to a relaxed LP problem and thereby solving the boolean LP in polynomial time, and Evaluating network performance, security, and energy consumption characteristics of the proposed algorithm for both symmetric and asymmetric key pre-distribution methods operating on top of on-demand routing protocols.

An example of proposed approach is given below.

Fig. 1 illustrates an example of routing in overlay networks.

Light blue links show the underlay connections and directed red links represent the overlay connections. In order to send a message to its destination node, a source node has to first find an overlay path. There may be several overlay paths from the source node toward the destination. Two such paths are shown in the figure. While the first overlay path reaches the destination through nodes 2 and 3, the second path reaches it through node 5. In the first path, any cryptographic (overlay) hop is formed by just one underlay hop. In the second path, the underlay path for the first cryptographic hop from the source node to node 5, is 2->3->4 where nodes 2, 3, and 4 are underlay neighboring nodes. The red overlay links are directed, i.e., node 1 is able to send a secure message to node 5 directly but not vice versa.

Hence after comparing these two paths, the proposed algorithm would select the second path to perform the data transmission.



Fig: Proposed data transmission approach

Algorithm

- 1. Identify the source and destination nodes in the network.
- 2. Identify all possible paths which can be used to reach from source node to destination.
- 3. Calculate the no. of hops required to reach from source to destination.
- 4. After comparing no. of hops required for all various paths, choose the path which requires least no. of hops.
- 5. Send the data from source node to destination node from the selected path.

Advantages

1. It illustrates one of the uses of community structure for the data transmission.

- 2. Data transmission can take place faster as the network is already rearranged through existing algorithm.
- 3. It helps to provide data to relevant and correct destination with minimum hops and thus in lesser time.

Modules

1) Source and destination nodes module

This module contains the information regarding the source node and the destination node.

2) No. of hops calculation module

This module calculates the no. of hops required by all possible paths and path which require minimum hops. It would be the output of this module.

Mathematical Model of proposed system

INPUT:

S-source Node

D-Destination Node

G-network

V-vertex (node or the point)

E-edge (Connection Between vertex or node)

OUTPUT:

A network path with minimum hops through which data from source node to destination node can be sent.

PROCESSING:

START

- 1. Call PATHFINDER(S,D,G(V,E)) # searches all possible paths from source, S to destination, D
- 2. If more than one path is found,

Call OPTIMUMPATHFINDER(S,D,G1(V,E))

Calculate no. of hops of all the possible paths.

Select the path with minimum hops required.

Else select the path .

3. Send data from source to destination.

END

Comparison with similar systems

With this project, the use of identifying central node and community structure has been enhanced. We have found the one application where the concept of central node and community structure is applied and is also useful.

Hence this project is an extended step of an existing algorithm.

Software Requirement Specification

Hardware Requirements

1) System: Pentium IV 2.4 GHz.

2) Hard Disk: 40 GB.

- 3) Monitor: 14 Colour Monitor.
- 4) Ram: 512 Mb

Software Requirements:

- 1) Operating system: Windows 7 Ultimate.
- 2) Java Development Kit : JDK 1.7 / JDK 1.8
- 3) Development Environment: Eclipse (Luna) IDE for Java
- 4) Programming Language: Java
- 5) Database: SQL Server 2012

Result

In existing system, network is constructed with more than 10000 users and 120 community. Forming of network used the graph structure G(E, V), So first file of dataset is edges between nodes and get all relation between them. Second file of dataset is community, in one community so many nodes are available on the basic of that edges and nodes (vertex) matrix are built.

The graphical structure shown in following graph,



Fig. : Total vertex and edges

On the basis of system calculate the center of the system. After finding center in network, members of community can share file to each other including some attribute like energy, received key and distance of node. The speed of sending of file measure by milliseconds, the following graph measures the performance of our file sending speed.



Conclusion

Thus, we have implemented and tested successfully the existing system explained in research paper, "**Fast and Accurate Mining the Community Structure: Integrating Center Locating and Membership Optimization**" We have also designed the enhancement part which shows data transmission between two nodes with the help of routing algorithm. For the future work, the community detection algorithm can be optimized in order to make it work for very complex network as well.

References

[1] X. F. Wang and G. Chen, "Complex networks: Small-world, scale-free and beyond," IEEE Circuits and Systems Magazine, vol.3, no.1, pp.6-20, Feb. 2003.

[2] R. D. Doverspike and J. Yates, "Optical network management and control," Proceedings of the IEEE, vol.100, no.5, pp.1092-1104, May. 2012.

[3] R. Albert and A. L. Barab#asi, "Statistical mechanics of complex networks," Reviews of Modern Physics, vol.74, no.1, p.47, 2002.

[4] M. E. J. Newman, "Networks: an introduction," Oxford University Press, 2010.

[5] J. L. Payne and M. J. Eppstein, "Evolutionary dynamics on scale-free interaction networks", IEEE Transactions on Evolutionary Computation, vol.13, no.4, pp.895-912, Aug. 2009.

[6] L. Wang and X. Li, "Spatial epidemiology of networked metapopulation: An overview", Chinese Science Bulletin, vol.59, no.28, pp. 3511-3522, Oct. 2014.

[7] S. Fortunato, "Community detection in graphs," Physics Reports, vol.486, no.3-5, pp.75-174, Feb. 2010.

[8] M. Gong, Q. Cai, X. Chen X and L. Ma, "Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition," IEEE Transactions on Evolutionary Computation, vol.18, no.1, pp.82-97, Feb. 2014.

[9] C. Pizzuti, "A multiobjective genetic algorithm to find communities in complex networks," IEEE Transactions on Evolutionary Computation, vol.16, no.3, pp.418-430, Jun. 2012.

[10] A. Rubio-Largo, M. A. Vega-Rodr#iguez, J. A. G#omez-Pulido and Juan M. S#anchez-P#erez, "Multiobjective Metaheuristics for Traffic Grooming in Optical Networks," IEEE Transactions on Evolutionary Computation, vol.17, no.4, pp.457-473, Aug. 2013.

[11] B. Yang, J. M. Liu, D. Y. Liu, "Characterizing and Extracting Multiplex Patterns in Complex Networks," IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol.42, no.2, pp.469-481, Apr. 2012.

[12] T. Rolland, M. Tasan, B. Charloteaux, et al., "A Proteome-Scale Map of the Human Interactome Network, Cell, vol.159, no.5,pp.1212-1226, Nov. 2014.

[13] Y. Liu, J. Moser, and S. Aviyente, "Network Community Structure Detection for Directional Neural Networks Inferred From Multichannel Multisubject EEG Data," IEEE Transactions on Biomedical Engineering, vol.61, no.7, pp.1919-1930, Jul. 2014.

[14] H. Gharavi and B. Hu, "Multigate communication network for smart grid", Proceedings of the IEEE, vol.99, no.6, pp.1028-1045,Jun. 2011.

[15] N. Tremblay and P. Borgnat, "Graph Wavelets for Multiscale Community Mining," IEEE Transactions on Signal Processing, vol.62, no.20, pp.5227-5239, Oct. 2014.

[16] H. J. Li and J. Daniels, "Social significance of community structure: Statistical view," Physical Review E, vol.91, no.1, p.012801,2015.

[17] A. Stanoev, D. Smilkov, L. Kocarev, "Identifying communities by influence dynamics in social networks," Physical Review E,vol.84, no.4, p.046102, 2011.

[18] J. Liu and T. Liu, "Detecting community structure in complex networks using simulated annealing with k-means algorithms," Physica A, vol.389, no.11, pp.2300-2309, 2010.

[19] S. Zhang, R. S. Wang and X. S. Zhang, "Identification of overlapping community structure in complex networks using fuzzy c-means clustering," Physica A, vol.374, no.1, pp.483-490, 2007.

[20] B. Yang, J. M. Liu, J. F. Feng, "On the Spectral Characterization and Scalable Mining of Network Communities," IEEE Transactions on Knowledge and Data Engineering, vol.24, no.2, pp.326-337, Feb. 2012.

[21] V. D. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment vol.2008, no.10, p.10008, Oct. 2008.

[22] Y. Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," Nature, vol.466, no.7307, pp.761-764, Aug. 2010.

[23] Y. V. Pehlivanoglu, "A new particle swarm optimization method enhanced with a periodic mutation strategy and neural networks," IEEE Transactions on Evolutionary Computation, vol.17, no.3, pp.436-452, Jun. 2013.

[24] S. Verel, G. Ochoa and M. Tomassini, "Local optima networks of NK landscapes with neutrality", IEEE Transactions on Evolutionary Computation, vol.15, no.6, pp.783-797, Dec. 2011.

[25] S. P. Mendes, G. Molina, M. A. Vega-Rodr#iguez and et al, "Benchmarking a wide spectrum of metaheuristic techniques for the radio network design problem," IEEE Transactions on Evolutionary Computation, vol.13, no.5, pp.1133-1150, Oct. 2009.

[26] S. Fortunato and M. Barth#elemy, "Resolution limit in community detection," Proceedings of the National Academy of Sciences of the United States of America, vol.104, no.1, pp.36-41, Jan. 2007.

[27] A. Khadivi, A. A. Rad and M. Hasler, "Network Community Detection Enhancement by Proper Weighting," Physical Review E, vol.83, no.4, p.046104, Apr. 2011.

[28] X. S. Zhang, R. S. Wang, Y. Wang, J. Wang, Y. Qiu, L. Wang and L. Chen, "Modularity optimization in community detection of complex networks," Europhysics Letters, vol.87, no.3, p.38002, Aug. 2009.

[29] G. Palla, I. Derenyi, I. Farkas and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," Nature, vol.435, no.7043, pp.814-818, Jun. 2005.

[30] L. Tang, H. Liu, And J. Zhang, "Identifying Evolving Groups In Dynamic Multimode Networks," Ieee Transactions On Knowledge And Data Engineering, Vol.24, No.1, Pp.72-85, Jan. 2012. [31] M. B. Hastings, "Community Detection As An Inference Problem," Phys. Rev. E, Vol. 74, No. 3, P. 035102, Sep. 2006. [32] S. Boccaletti, M. Ivanchenko, V. Latora, And A. Pluchino, "Detecting Complex Network Modularity By Dynamical Clustering," Phys. Rev. E, Vol. 75, No. 4, P. 045102, Apr. 2007. [33] A. Clauset, M. E. J. Newman, And C. Moore, "Finding Community Structure In Very Large Networks," Phys. Rev. E, Vol. 70, No. 6, P. 066111, Dec. 2004.