Scientific Journal of Impact Factor (SJIF): 4.72
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 10, October -2017

Community detection Using Modularity

HarekrishnaMili

Department of Mathematics,
Dibrugarh University, Dibrugarh-786004

Abstract -*In this paper, consider an algorithm of a community detection taking the diameter of the community as a key parameter. To measure the quality of the community structure in a network, a new metric called 'Modulariry' is used. To have a good quality community structure, the diameter of a community is make always small, upper bounded the diameter by a certain value k<<D , where D is the diameter of the network. In terms of Modularity of the community structure, the algorithm performs better than the existing algorithm by simulation on Zachary karate club, Dolphin network and football club. Also the definition of modularity given in this paper can compare the community structure of a given network better.*

Key Words:Community structure, Diameter, Modularity, Good quality Modularity.

1.INTRODUCTION

The term community structure in a network means natural division of the nodes into densely connected subgroups within and between them connection is sparse. By identifying and analysing the community structure of a network, we can understand and characterised the related network and its properties. It is evident that in graph or network, communities are loosely defined and it is not possible to compare the quality of two communities getting by different algorithm of the same graph [1]. By modularity metric, the quality of community structure is measured by Newman[1] , which is based on expected connectivity between nodes .A good high modularity nodes gives good quality community structure. The problem of community detection is that how to define the modularity metric that it may compare the quality of two communities and the partitioning of the nodes that miximizes modularity. Based on the topology of the given network itself not on the expected connectivity, this paper consider a new approach to compute modularity of a community structure. The modularity of a community structure is the average node modularity within it,where the node modularity is defined as the ratio of the number of neighbours of a node v within the same community to the total degree of the node v. So we can say that the modularity value varies between 0 and 1 . By our approach, by joining a node into a community ,it always attempts to maximize its modularity. A node can be addedto a community till that the diameter of the community structure is less than or equal to the predetermined constant k<<D, where D is the diameter of the original graph. Based on this criteria ,an algorithm is proposed for partitioning a given network into some communities with high modularity.

2. Preliminaries:

Girvan and Newman was first proposed the community detection in networks based on egdebetweenness centrality [1]. To reduce the computational complexity and to improve the modularity measure several algorithms are proposed[2-12]. These algorithms are measured the modularity by substractithe expected number of community edges from the actual number of intra- community edges[3], which mathematically given as:

$$Q = \frac{1}{2m}\sum_{uv}\left[A_{uv} - \frac{d_u d_v}{2m}\right]\delta(c_u c_v) \;\ldots\ldots.(1)$$

Where A_{uv} is the elements of the adjacency matrix of the network ,delta-function δ is equal to 1 if $c_u = c_v$, otherwise its value is 0, $m = \frac{1}{2}\sum_{uv} A_{uv}$is the total number of edges in the graph and $d_u$is the degree of node u ,which is equal to

$$d_u = \sum_v A_{uv} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Here modularity of a community structure is defined based on the existing edges are not on the expected connectivity [3] . Let C be a community and d_v^{in} is the number of edges of a node v ∈C .Then the node modularity is

$$v_{mod} = \frac{d_v^{in}}{d_v} \quad \text{......................(3)}$$

Then the community modularity is

$$c_{mod} = \frac{\sum_{\forall v \epsilon C} v_{mod}}{n_c} \quad \text{................(4)}$$

Where n_c is the number of nodes in C.

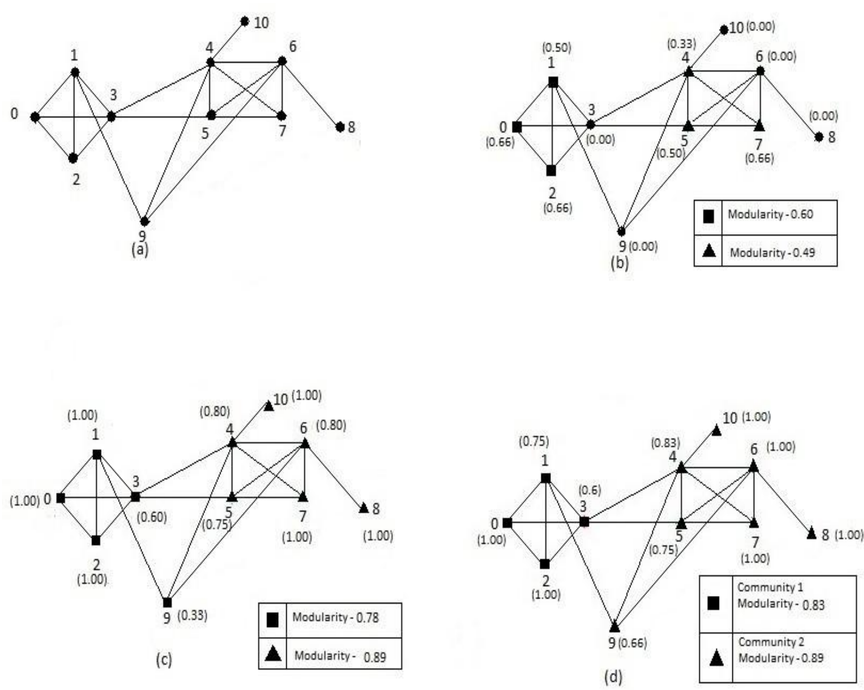

Figure 1: (a) Example Graph (b) All disjoint 3-cliques for the Example Graph (c) Intermediate steps (d) Communities of the Example Graph.

To improve the average modularity of the community keeping the diameter less than the predefined value k ,ie to produce a good quality community structure an iterative procedure by migrating a node from one community to other is proposed. A good community structure is that community which has not only high modularity value but also its diameter should be small so that it can maintain a good connectivity within intra community nodes.Bytaking the average of the modularities of the communities ,the modularity of the network is obtain, denoted as M.

4. Proposed Algorithm:

In this algorithm , only one node is allowed at a time to migrate from one community to another keeping the diameter less than k and making improvement in average modularity. Let Γ(v) denotes the neighbours of the node v, which is arranged in ascending order of degree. In Algorithm1, finding a set of disjoints 3-cliques, which is considered as initial set of components C. In case C=Φ, each node of the graph is treated as community.

Algorithm 1 Initial set of components algorithm

INPUT: Graph: $G(V, E), \Gamma(v)$

OUTPUT: A set of components : $C = \{c_1, c_2, ..., c_l\}$

sort the nodes $V = \{v_1, v_2, ..., v_n\}$ in increasing order of their degree

for each I **ϵV do**

if i is not marked **then foreach** $j \in \Gamma(i)$ **do**

if j isnotmarked**then if** $i > j$ **then**

k ← ~~fi~~ndacommonelementof$\Gamma(i)$and$\Gamma(j)$

if $k\neq$ **then**

if k is not marked **then**

make3-clique <i,j,k>andmark i, j, k.

set the diameter of the 3-clique to 1

For graph G(V,E), shown in fig1(a), first compute initial set of components,which is shown in fig1(b).At the initial level the graph has two initial components <0,1,2> denoted as component1 and <4,5,7> denoted as component 2. For each component ,first compute node modularity using (3) and than the modularity of the community by using (4), which is shown in fig1(b).In algorithm 2, each node of the graph move to the components such that for getting maximum modularity with keeping the diameter of the community less than or equal to pre-defined value k.In fig1(c), the node 3 and 9 joined with the component1, which gives the maximum increment in modularity and the nodes 6,8 & 10 are joined with component2 to get the same. Now migrating the node 9 from component1 to component2 , the overall modularity denoted by M, is improves ,so this migration is allowed.The fig1(d) shows the final results,where each components is term as community. For the example graph the pre-defined value is k=2.

Algorithm 2 Community detection algorithm

INPUT: Graph: $tt(V, E)$; Initial Component set: $C =$

$\{c_1, c_2 ...., c_l\}$; Threshold Diameter: k

OUTPUT: Final set of components called communities

$iter$1

whileiter≤$MAX\_ITERATION$**do**

foreach$v \epsilon V$**do**

find the component $c_x \epsilon$ C for maximum increment in community modularity with diameter is less than equal to k. In case of tie, choose one with minimumdiameter

if$flag(v)$=1and$v \epsilon c_y$**then**

ifthe transition of v from$c_y$to $c_x$improves M then

ifdiameter of$c_x \leq k$

$v$joins c_x

iftransition of v disconnects some nodes in $c_y$then remove those nodes from$c_y$and set the flagvalues to 0

update the diameter and community modularity of $c_y$and c_x

else

v **joins**c_x

$flag(v)$ 1

update the diameter and community modularity of c_x

ifno node joins any of the component **then**

return

Simulation and Testing

To compare th final modularity of communities ,here take Serial Louvain algorithm[5] and Girvan Newman ie GN[1] algorithm. Considered som small real worl networks like Zachary karate club[15], US College football club[16], and Dolphin's network[17] ,to test and simulate the proposed algorithm. This data sets artkaen from UCI network data resository. As the considered intian component have the diameter 1 and also the diameter of te small world network is less than equal to 6, so studied the solution for $2 \leq k \leq 5$.Here N is the number of communities.The value of Max-iteration is taken a large value.

Table 1: Comparison of number of communities N and the modularity M according to our definition

Algorithm		Karate		Dolphin		Football
	N	Modularity	N	Modularity	N	Modularity
GN	2	0:91	2	0:50	12	0:58
Louvain	4	0:76	10	0:58	10	0:70
Our Algo.	2	0:91	4	0:78	9	0:69

Observing the Table 1, for karate club the algorithm gives the same numbers of community with GN which equal to N=2,where the Louvains algorithm gives 4 communities with less modularity acieved by the proposed algorithm and GN algorithm. In case of Dolphin's network the proposed algorithm gives more communities than GN but less number of communities than Louvains with higher modularity than obtained by GN and Louvain. For football network the algorithm gives less number of communities than GN and Louvain but the modularity value is higher than GN and approximately equal to Louvain.

Table 2: Comparison of number of communities and the mod-ularity according to Newman's modularity definition

Algorithm		Karate		Dolphin		Football
	N	Modularity	N	Modularity	N	Modularity
GN	2	0:41	2	0:49	12	0:60
Louvain	4	0:41	10	0:51	10	0:60
Our Algo.	2	0:41	4	0:55	9	0:63

In Table 2, for karate club the algorithm gives the equal number of communities ie N=2, with GN but gives less number of communities than Louvain ie N=4. Here the modularity value is equafor all those three algorithm. For Dolphin's network, the algorithm gives the higher modularity value than GN and Louvain but the number of communities is less than the GN and Louvain. For football network ,the algorithm gives less number of communities than GN and Louvain but it is gives higher modularity value than GN and Louvain. Observing the results on table 2 ,for both the karate and football network, the modularity value achieved by GN and Louvain algorithm are same, though the number of communities are different. So in this case it is not possible to compare karate and football network by modularity. But from the table 2, using the uniquemodularity definition it can be easily compare the two networks,which shows that for karate club GN performs better whereas for football network Louvain algorithm gives better communities. So the given unique modularity definition is more successful to compare the quality of two communities.

Conclusion and Future Work

This paper has been proposed a new modularity definition for a community structure and an algorithm is designed to detect good quality community structure using modularity metric and keeping the diameter of the community to small. The proposed algorithm is tested yon small real world network data sets and compared it with the existing algorithm GN[1] and Louvain[5]. The simulation shows that the propsed algorithm gives equal or better community structure than the community obtained by GN and Louvain.

On large real world network and on large scale random network, the extensive simulation yet to be done to show the performance of the proposed algorithm and the new modularity definition. For large network the computational time will be high so like to parallalized the proposed algorithm to reduce its computational time.

REFERENCES

[1]M. E. Newman and M. Girvan, “Finding and evaluating community structure in networks,” Physical review E, vol. 69, no. 2, p. 026113, 2004.
[2]M. E. Newman, “Fast algorithm for detecting community structure in networks,” Physical review E, vol. 69, no. 6, p. 066133, 2004.
[3]A. Clauset, M. E. Newman, and C. Moore, “Finding community structure in very large networks,” Physical review E, vol. 70, no. 6, p. 066111, 2004.
[4]U. N. Raghavan, R. Albert, and S. Kumara, “Near linear time algorithm to detect community structures in large-scale networks,” Physical Review E, vol. 76, no. 3, p. 036106, 2007.
[5]V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and Lefebvre, “Fast unfolding of communities in large networks,” Journal of statistical mechanics: theory and experiment, vol. 2008, no. 10, p. P10008, 2008.
[6]E. J. Riedy, H. Meyerhenke, D. Ediger, and D. A. Bader, “Parallel community detection for massive graphs,” in International Conference on Parallel Processing and Applied Mathematics. Springer, 2011, pp. 286–296.
[7]J. Soman and A. Narang, “Fast community detection algorithm with gpus and multicore architectures,” in Parallel & Distributed Processing Symposium (IPDPS), 2011 IEEE International. IEEE, 2011, pp. 568–579.
[8]M. E. Newman, “Spectral methods for community detection and graph partitioning,” Physical Review E, vol. 88, no. 4, p. 042822, 2013.
[9]C. Y. Cheong, H. P. Huynh, D. Lo, and R. S. M. Goh, “Hierarchical parallel algorithm for modularity-based community detection using gpus,” in European Conference on Parallel Processing. Springer, 2013, pp. 775–787.
[10]H. Lu, M. Halappanavar, and A. Kalyanaraman, “Parallel heuristics for scalable community detection,” Parallel Computing, vol. 47, pp. 19–37, 2015.
[11]J. Cheng, L. Li, M. Leng, W. Lu, Y. Yao, and X. Chen, “A divisive spectral method for network community detection,” Journal of Statistical Mechanics: Theory and Experiment, vol. 2016, no. 3, p. 033403, 2016.
[12]P. Held and R. Kruse, “Online community detection by using nearest hubs,” arXiv preprint arXiv:1601.06527, 2016.
[13]U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Nikoloski, and D. Wagner, “On modularity clustering,” Knowledge and Data Engineering, IEEE Transactions on, vol. 20, no. 2, pp. 172–188, 2008.
[14]Y. Fujiwara, M. Onizuka, and M. Kitsuregawa, “Real-time diameter monitoring for time-evolving graphs,” in Database Systems for Advanced Applications. Springer, 2011, pp. 311–325.
[15]W. W. Zachary, “An information flow model for conflict and fission in small groups,” Journal of anthropological research, pp. 452–473, 1977.
[16]M. Girvan and M. E. Newman, “Community structure in social and biological networks,” Proceedings of the national academy of sciences, vol. 99, no. 12, pp. 7821–7826, 2002.
[17]D. Lusseau, “The emergent properties of a dolphin social network,” Proceedings of the Royal Society of London B: Biological Sciences, vol. 270, no. Suppl 2, pp. S186–S188, 2003.
[18]G. Li, D. Zhang, K. Xie, T. Huang, and Y. Li, “A gpu based fast community detection implementation for social

