

Modularity-Based Community Detection in Fuzzy Granular Social Networks

Nicole Belinda Dillen and Aruna Chakraborty

Abstract Social network analysis is an important task in the modern, globalised world and has several applications in crime, economy, and human psychology. An important aspect of social network analysis is community detection in which groups of closely connected individuals are identified separately from other groups. In this paper, we proposed a new method for detecting communities in a social network. Our method is inspired by fuzzy granular social networks (FGSN) and uses a popular heuristic modularity-based community clustering algorithm. The results obtained from our algorithm correlate well with those obtained by other popular modularity-based detection methods, making it a promising algorithm for community detection in non-overlapping networks.

Keywords Social network analysis · Community detection · Modularity · Fuzzy granular social networks

1 Introduction

A social network is a collection of individuals and their relationships with each other. An example, extremely relevant to our growing internet society would be the online social networks found on websites like Facebook and Twitter. All real-world social networks have one thing in common: proper community structure. Put simply, a community is a collection of individuals that share a common interest. In an online social network like Facebook, community structure could be identified by groups of “friends” with the same “Liked Pages” and, perhaps, even the same “mutual friends”.

N.B. Dillen (✉) · A. Chakraborty
Computer Science and Engineering Department,
St. Thomas' College of Engineering and Technology, Kolkata, India
e-mail: nicolebdillen@gmail.com

A. Chakraborty
e-mail: aruna.stcet@gmail.com

Based on the facts mentioned above, one could see how detecting community structure in networks could help in studying or predicting the overall behaviour of the networks. It could help business organisations target sections of society that would most likely respond positively to their products. It could help law enforcement agencies in unfolding pockets of criminal organisations which may not have been initially prominent. In large networks, it could even aid in the detection of important “key” personalities such as influential politicians or eminent scientists.

Since the mid-90s, several community detection algorithms have been developed. One of the most popular and earliest of these was the Girvan-Newman algorithm [1] which partitions the graph into a number of communities by removing connections that are most likely to occur “between” communities. Another algorithm by Pons and Latapy [2] used random walks to detect communities. An approach by Newman [3] used a fast modularity optimisation algorithm which was later improved by Clauset, Newman and Moore (the CNM algorithm) [4]. Wakita and Tsurumi [5] developed a new metric known as *consolidation ratio* which attempts to balance the communities detected by the CNM algorithm. Both Newman’s original algorithm as well as the CNM method inspired the well-known Louvain algorithm [6] which uses a greedy modular optimisation technique to accomplish the community detection task.

Many of the newer community detection algorithms include an additional scope of detecting overlapping communities, i.e., communities whose nodes may belong to other communities as well. Work in this area includes a modified modularity optimisation algorithm [7] for detecting overlapping communities and an algorithm based on the concept of Fuzzy Rough set theory [8].

In our paper, we have proposed a novel algorithm that is applicable to real-world social networks without any overlapping communities. Our concept is inspired by Fuzzy Granular Social Networks [9] as well as the Louvain algorithm [6]. While the Louvain algorithm deals with distinct, individual nodes, our algorithm extends this principle to the domain of fuzzy granular social networks (FGSN), the main aim of which is to reduce the set of nodes to a smaller set of granules.

Our paper is organised as follows: Sect. 2 deals with the preliminary information related to our algorithm while Sect. 3 describes the said algorithm. The results obtained are discussed in Sect. 4. Finally, Sect. 5 deals with our conclusions and possible future work.

2 Preliminary Concepts

In this section, we formally define and explain some of the necessary concepts related to social network analysis which have been used by our proposed algorithm.

Definition 1 Formally, a social network is a graph $G(V, E)$ where:

1. V is the set of all vertices or nodes or individuals
2. E is the set of all edges or links that interconnect the vertices in V

Definition 2 A community is a subset of nodes that are densely interconnected while being sparsely connected to the other nodes of other communities.

Our objective is, therefore, to identify such groups of densely interconnected nodes and classify them as communities. To do this, we model our social network system in the fuzzy domain using the concept of granules which we describe below.

Most algorithms take into account each and every node of a network. However, we could drastically reduce the number of computations if we group “similar” nodes together to form what are known as “granules” and perform the remaining computations solely on these granules. In fact, this is the very same approach used in creating FGSN [9]. The next four definitions serve to clarify this concept.

Definition 3 A granule [10] is a collection of similar, indistinguishable objects that can be treated as an independent unit. In the context of social networks, a granule, denoted by A_c , is represented by a centre vertex c , and a node’s relationship (usually a distance function) with c defining its membership in the granule.

Realistically speaking, some nodes may have equal or different memberships in multiple granules instead of just one. To account for this, the fuzzy domain [11] has been incorporated in the membership assignment of nodes to various granules. We denote this fuzzy membership as $\mu_c(v)$ which denotes the membership of node v (a monotonically non-increasing function) in the granule represented by centre c .

Definition 4 According to the FGSN [9] theory, the membership of a node v in a granule represented by centre c is denoted by $\mu_c(v)$, and defined in Eq. (1) below:

$$\mu_c(v) = \begin{cases} 0 & \text{for } d(c, v) > r \\ \frac{1}{1+d(c,v)} & \text{otherwise} \end{cases} \tag{1}$$

where $d(c, v)$ is the distance between node v and granule centre c , and r is the desired granular radius which may be varied. When the distance is 0, i.e., the vertex is c itself, $\mu_c(v) = 1$ while $\mu_c(v) = 0$ for infinite distance. Also, these membership values must be normalised as a node may belong to a number of granules with varying membership. The normalised membership value is then,

$$\tilde{\mu}_c(v) = \frac{\mu_c(v)}{\sum_{i \in C} \mu_i(v)}. \tag{2}$$

Thus, we have:

1. C , the set of vertices each representing a particular granule, and
2. $Gr = \{A_c | \forall c \in C, \sum_{v \in V} \tilde{\mu}_c(v)/v\}$, the set of all granules.

Another important aspect of social networks is embeddedness.

Definition 5 The embeddedness for a pair of granules, centred at a and b respectively, is the extent to which one is embedded in the other. It is nothing but the cardinality of the intersection of both granules and is denoted by $\varepsilon(a, b)$:

$$\varepsilon(a, b) = |A_a \cap A_b| = \sum_{v \in V} \min(\tilde{\mu}_a(v), \tilde{\mu}_b(v)). \quad (3)$$

We now provide a brief background of the Louvain algorithm [6] that is used to detect communities in social networks: the Louvain algorithm is a greedy optimisation algorithm which works on the principle of modularity. Like other optimisation algorithms [3, 4], its objective is to maximise the modularity by placing nodes in communities that result in a local maxima.

Definition 6 Modularity Q is a measure used to provide a qualitative assessment of the community partitions that have been detected in the network. It conveys the difference between the actual density of interconnections between nodes in a detected community and the corresponding connections in a random network possessing the same degree distribution as that of the actual network [3]:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j). \quad (4)$$

where A_{ij} is the weight of the edge between vertices i and j , k_i is the total weight of all edges linked to i , c_i is the community to which i is assigned and $\delta(c_i, c_j)$ is 1 when both i and j belong to the same community and is 0 otherwise. The total weight of all edges in the network is m , where $m = \frac{1}{2} \sum_{i,j} A_{ij}$.

The Louvain algorithm consists of two stages. The first, also called the “iterative stage” is the greedy stage which iteratively looks for the local maxima of the modularity. Each node is initially considered a single community. For each node i , we compute the change in modularity obtained by removing i from its community and placing it in the community of one of its neighbours. The node i is placed in the community for which this modularity change is both positive and maximum. This process is repeated for all other nodes in the network. The entire stage is then repeated iteratively until no further increase in modularity can be obtained.

The second stage, or “coarse-graining” stage, groups all the nodes belonging to a community into a single unit. A new network is formed whose nodes correspond to the communities detected during the iterative stage. Here, a link between two nodes is simply the sum of weights of the connections between the nodes of the corresponding communities. Similarly, self-loops may also be generated which have weights equal to the sum of intra-connections between nodes of the same community. The first stage is then reapplied to the new adjacency matrix. The two stages are repeated until the modularity cannot be optimised any further. A hierarchy is thus obtained consisting of the communities detected after each phase of iteration and coarse-graining.

3 Proposed Algorithm

In brief, our algorithm is simple. We first choose certain nodes as granule centres. This is done by computing the average degree of all nodes in the network and choosing only those nodes with a degree greater than the average as granule centres. We set the network diameter as the granule radius and construct the set of all granules according to the methods described in the previous section. Next, we seek to construct a new “network” whose nodes correspond to the granules. We assign a link between two nodes in our new network whose weight is equal to the embeddedness between the corresponding two granules. Note that this is applicable to self-loops as well. In the case of a self-loop, the loop weight will simply be the cardinality of the granule itself.

After we construct the adjacency matrix for our new network, we detect “granular communities” in it by means of the Louvain algorithm. We use these granular communities to construct the actual set of corresponding communities for the vertices in the social network. We first construct a Fuzzy-Rough community matrix [9] in the following manner: for every granular community g_i , we construct a corresponding community C_i in which a vertex v 's membership to C_i is set to 1 if all its positive granular memberships involve only those granules that have been assigned to g_i . If v is assigned positive memberships in granules belonging to multiple granular communities, its membership to C_i is equal to sum of its memberships of all granules in g_i . Obviously, if v possesses 0 membership in all granules of g_i , it will be assigned a membership of 0 to C_i . Finally, for every vertex v , we look for the community in which v has the highest membership value and set this value to 1. All other membership values are set to 0. Our algorithm, which we now call “GranLouv” is formally stated below.

Algorithm: GranLouv

1. *Start*
2. Set granule representative set: $C \leftarrow \{v: degree(v) > average\ degree\}$.
3. Form granule set $Gr = \{A_c \mid \forall c \in C, \sum_{v \in V} \tilde{\mu}_c(v) / v\}$ using distance function (1) and membership function (2).
4. Consider each granule as a ‘vertex’ and construct an adjacency matrix as below:
 - a. initialise an $N \times N$ matrix M where N is the number of granules.
 - b. $M(i, j) \leftarrow \epsilon(a, b)$, from equation (3).
5. Granular-communities $g \leftarrow Louvain(M)$.
6. Obtain Fuzzy Rough Community matrix from g .
7. $\forall v \in V$, find C_i for which $C_i(v)$ is maximum. Set $C_i(v)=1$ and $C_j(v)=0, \forall j \neq i$
8. *End*

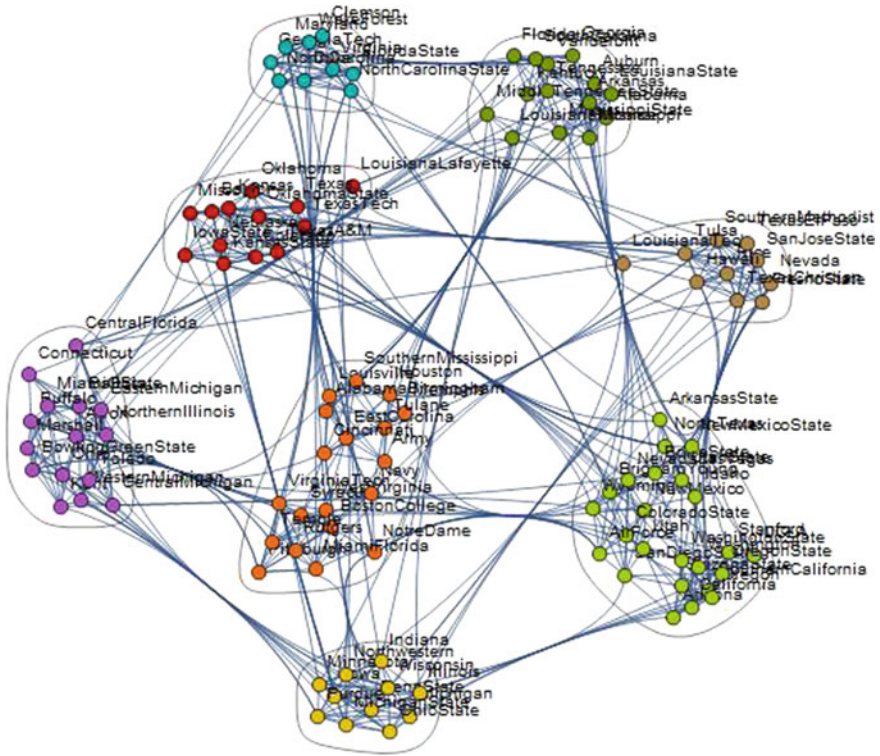


Fig. 3 American college football social network: communities detected using GranLouv

4 Application and Results

Our algorithm was tested using a 2.40 GHz dual core CPU with 2.00 GB RAM and was implemented with Mathematica 10.0. We considered three different real-world datasets, namely, the Dolphin [12], Les Miserables [13] and the American College Football [1] social networks. The results obtained for each of these networks compared with those obtained by various reference algorithms are provided below (Figs. 1, 2, 3 and Tables 1, 2 and 3). In all cases, we considered the highest modularity obtained in a series of iterations of the algorithm.

Table 1 Dolphin social network: modularity

Serial no.	Algorithm	Modularity
1.	Louvain	0.518
2.	GranLouv	0.509

Table 2 Les miserables social network: modularity

Serial no.	Algorithm	Modularity
1.	Newman and Girvan [14]	0.540
2.	Louvain	0.555
3.	GranLouv	0.534

Table 3 American college football social network: modularity

Serial no.	Algorithm	Modularity
1.	Girvan-Newman [1]	0.601
2.	Louvain	0.604
3.	GranLouv	0.599

5 Conclusion and Future Work

We have seen from our results that our algorithm produces results comparable to other popular algorithms. While other modularity-based algorithms have considered each and every node of the network during the detection process, we have accomplished the optimisation task by considering only a few granules instead.

Scope for improvement lies in the selection of granule centres as, in our implementation, we have considered only those nodes with degree greater than the average degree of the network. However, this may not be the best way to select the most significant or important nodes. A better selection algorithm could yield even better granular communities and, thus, better final communities. Our future work will include modifying the algorithm to address the granule centre selection problem mentioned above as well as extending our algorithm to accommodate overlapping communities.

Acknowledgments The authors of this paper wish to convey their thanks to Prof. Sankar Kumar Pal as well as Mr. S. Kundu from the Center for Soft Computing Research, Indian Statistical Institute, for their continued support and assistance provided during the conception of this paper.

References

1. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99(12):7821–7826, June 2002.
2. Pons, P., Latapy, M., 2006 *Journal of Graph Algorithms and Applications* 10 191–218.
3. Newman M.E.J.: Fast algorithm for detecting community structure in networks. *Phys. Rev. E* 69, 066133 (2004).
4. Clauset, A., Newman, M., Moore, C.: Finding community structure in very large networks. *Physical Review E*, 70(6):1–6, 2004.
5. Wakita, K., Tsurumi T.: 2007 *Proceedings of IADIS international conference on WWW/Internet 2007* 153.
6. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10), P10008 (12 pp) doi:10.1088/1742-5468/2008/10/P10008. arXiv:<http://arxiv.org/abs/0803.0476>.
7. Ganganath, N., Chen, G., Cheng, C.-T.: Detecting Hierarchical and Overlapping Community Structures in Networks. 2014 *International Symposium on Nonlinear Theory and its Applications, NOLTA2014*, Luzern, Switzerland, September 14–18, 2014.
8. Kundu, S., Pal, S.K.: FGSN: Fuzzy Granular Social Networks – Model and applications. *Information Sciences*, vol. 314, 1 September 2015, pp. 100–117.
9. Kundu, S., Pal, S.K.: Fuzzy-rough community in social networks. *Pattern Recognition Letters* (2015), <http://dx.doi.org/10.1016/j.patrec.2015.02.005>.

10. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* 90 (1997) 111–127.
11. Zadeh, L.: Fuzzy sets, *Inform. Control* 8 (1965) 338–353.
12. Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M.: The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396–405, 2003.
13. Knuth, D.E.: *The Stanford GraphBase: A Platform for Combinatorial Computing*. Addison-Wesley, 1993.
14. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Phys. Rev. E*, vol. 69, p. 026113, 2004.