

Quantifying Influential Communities in Granular Social Networks Using Fuzzy Theory

Anisha Kumari[1], Ranjan Kumar Behera[1(✉)], Abhishek Sai Shukla[2], Satya Prakash Sahoo[1], Sanjay Misra[3], and Sanatanu Kumar Rath[4]

[1] Department of CSE, Veer Surendra Sai University of Technology, Burla, Odisha, India
anishamishracs@gmail.com, jranjanb.19@gmail.com, sahoo.satyaprakash@gmail.com

[2] PwC Service Delivery Center, Bengaluru, India
deva.abhi96@gmail.com

[3] Department of Electrical and Information Engineering, Covenant University, 1023, Ota, Nigeria
ssopam@gmail.com

[4] Department of CSE, National Institute of Technology, Rourkela, Rourkela, India
skrath@nitrkl.ac.in

Abstract. Community detection and centrality analysis in social networks are identified as pertinent research topics in the field of social network analysis. Community detection focuses on identifying the subgraphs (communities) which have dense connections within it as compared to outside of it, whereas centrality analysis focuses on identifying significant nodes in a social network based on different aspects of importance. A number of research works have focused on identifying community structure in large-scale network. However, very less effort has been emphasized on quantifying the influence of the communities. In this paper, group of nodes that are likely to form communities are first uncovered and then they are quantified based on the influencing ability in the network. Identifying exact boundaries of communities are quite challenging in large scale network. The major contribution in this paper is to develop a model termed as FRC-FGSN (Fuzzy Rough Communities in Fuzzy Granular Social Network), to identify the communities with the help of fuzzy and rough set theory. The proposed model is based on a idea that, the degree of belongingness a node in a community may not be binary but can be models through fuzzy membership. The second contribution is to quantifying the influence of the community using eigenvector centrality. In order to improve the scalability, several steps in the proposed model have been implemented using map-reduce programming paradigm in a cluster-computing framework like Hadoop. Comparative analysis of FRC-FGSN with other parallel algorithms as available in the literature has been presented to demonstrate the scalability and effectiveness of the algorithm.

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12252, pp. 906–917, 2020.
https://doi.org/10.1007/978-3-030-58811-3_64

Keywords: Influential community · Eigenvector centrality · Granular social network · Cluster-computing

1 Introduction

A social network can be defined as a graph structure where nodes represent the individuals and the edges within it represent the relationships among the individuals. With recent growth in the online social networking sites (OSNs) such as Facebook, Twitter, Instagram etc., the analysis of the static and dynamic features plays an important role in every aspect of human life. Real-world social networks have some distinctive characteristics like, small-world effect [1], power law degree distribution [2] and community structure [3], which make the analysis of these networks even more interesting. Small world effect and power law degree distribution basically deals with degree of association among the nodes in the network. Communities are defined as the functional units of the network which may be formed due to the mentioned features. It can also be defined as the dense subgraph which have large number of edges within themselves and less number of edges going out of them. Since the real-world social network follows power law degree distribution, some of the nodes are more significant as compared to others [4,5]. The communities having more number of influential nodes are said to have high influencing factor as compared to the communities having less significant nodes. The identification of influential communities passes through two phases:

1. **Phase I- Community detection process**: With the increase in complexity and size, exploring communities in large scale network is found to be a challenging task. Pal et al. have developed an efficient algorithm to explore the fuzzy-rough communities [6]. This algorithm has been designed to run on a granular social network model which is based on fuzzy and rough-set theory known as fuzzy-granular social network (FGSN) [7]. Unlike other community detection algorithms, this approach gives better results when the network consist of bigger number of overlapping communities [6]. The extension of the work has been implemented in the proposed work.
2. **Phase II- Influence maximization for the Communities**: Bonacich has studied the unique properties of eigenvector centrality [8]. Unlike graph-theoretic centralities like degree, betweenness, and closeness centrality, eigenvector centrality considers the centrality of its neighbors [8,9]. Thus, eigenvector centrality measure helps in estimating the influence of a certain node in a real-world scenario. In this work, eigenvector centrality has been considered to calculate the influence of the communities.

Social networks have similar characteristics of big data namely volume, variety and velocity (3 V's) [10]. Real-world social networks with number of nodes of the order $\geq 10^6$ are quite difficult to process with conventional tools. In this paper, map-reduce based community detection algorithms have been designed which is implemented in Apache Spark distributed framework. *Apache Spark* is a cluster-computing framework, widely used for processing distributed applications.

The subsequent sections of the papers are described as follows: Sect. 2 presents the motivation about fuzzy implementation of social network. Section 3 presents the literature survey of the work. The background details is discussed in Sect. 4. Section 5 presents the problem statement and the proposed algorithm of the work. Section 6 presents the implementation work of the proposed algorithms. The experimental results and the observation is presented in Sect. 7. The conclusion and future work is presented in Sect. 8.

2 Motivation

Although a good number of research works have been carried out in the literature for community detection in social networks, quantifying them in term of influential factor is still a challenging task. Each individual entity and relationship may not have significant impact in community formation. For example, if there is an edge, that exists between two entities, it is wrong to conclude that they have a strong relationship and if no edge exists between them they are totally unfamiliar with each other. Exact degree of bonding for a relationship or a node in a community is quite difficult to measure. In this paper, social network has been modeled with fuzzy relationships in order to capture the degree of fuzziness in bonding.

Analyzing social networks can be a tedious task mostly because of its larger data size [10]. Conventional tools and methods are either not scalable or consume high computational power. Apache Spark, Hadoop, Sqoop, Hive, NOSQL etc. are the few available tools in order to handle the big data applications in scalable and time efficient manner. These are the cluster-computing frameworks where the computation is distributed among several nodes in a cluster. The results from all the nodes are combined together (by the master node) in order to get the final output. In this work, we have considered Apache Spark, which is one of the popular tools for cluster computing and is nearly 100 times faster than Apache Hadoop [11].

3 Literature Survey

Girvan *et al.,* in their seminal work on community structure in biological and social networks have presented one of the most popular community detection algorithms [12]. This algorithm is based on divisive approach where, edges are removed in the order of their edge betweenness value (number of shortest paths going through them). The intuition behind the algorithm is that communities are connected through the edges with larger betweenness value. This concept is found to be widely adopted in several others methods for community detection [3, 13,14]. These algorithms may be classified into two major categories; first, those methods where a node is considered to be member of any single community at a particular instance of time and second, those where a node may belong to more than one community (overlapping communities).

Pal *et al.* came up with the idea of fuzzy-rough based communities detection algorithm [6]. In their work, social network was modeled using fuzzy-granular concepts, which is necessary to detect fuzzy-rough communities [7]. Unlike conventional social network models, in fuzzy-granular social networks (FGSN's) nodes are clustered into number of granules and the unit of processing is granules rather than nodes. This kind of modeling approach helps in calculating the influence of communities by calculating the overlap of communities (granular embeddedness) which helps us to obtain a weighted undirected graph whose eigenvector centrality can be easily calculated.

Li *et al.,* have presented a method to detect influential communities [15]. In their model of social network, each node is associated with a weight which measures the importance of node. The communities have been detected using k-cores (group of connected nodes whose degree is at least k). In their work the influence of community is defined as the minimum weight of the nodes in the community. Though a scalable version of the algorithm has been presented, but it has certain limitations. Firstly, it doesn't consider the overlapping nature of real-world communities. Secondly, the least weight of a node may not be the perfect measure for quantifying the influence of community.

Wang *et al.,* have presented a unique approach for detecting influential community in large scale networks [16]. In this approach community is detected using kr-cliques and influence is then measured using influence maximization technique. This approach evaluates the influence efficiently, but the dataset necessary for evaluating influence using this measure needs to have activation probability of every edge in the graph. Activation probability of an edge is the probability with which one node can influence its adjacent node along that edge. Obtaining these probabilities is a quite cumbersome process [17]. Thus, calculating influence using this approach is highly computational expensive which is not feasible for real world network.

4 Background Details

4.1 Fuzzy-Granular Social Network (FGSN)

Social network consists of set of entities (individual, group or organization) and relationships among the entities. As the size of the social network is obviously large in nature, real world networks have huge contribution to the era of big data world. It is quite a difficult task to investigate each individual node and relationships for large scale network. Social network can be modeled as a Fuzzy-granular social network (FGSN) where the network consists of set of macro units known as granules [7]. A granule around a node can be constructed with the help of fuzzy set [18] consisting of neighboring nodes with the degree of membership. Each granule is identified by a granule representative. A fuzzy-granular social network can be represented by four parameters as follows:

$S = (C, V, G, A_c)$ such that,

- V is the set of vertices in the network.
- C is the set of granule representatives, where $C \subseteq V$.
- G is the set of all granules around each granule representative $c \in C$
- A_c is the granule having center at c.

The membership value of a node in a granule can be obtained by various parameters associated in the network. Usually the membership value of a node should be decreased as the distance from granule representative increases. In this paper, the fuzzy membership function is defined as follows: where $\mu_c(v, r)$ is the membership value of node v with respect to granule representative c. r is the radius of the granule (Granule Radius). For the method presented in this paper, $d(c, v)$ has been considered as the minimum hop distance between the granule representative c and node v.

Few terms related to FGSN used throughout this paper are discussed below:

- **Granule Radius**: It is defined as the maximum allowable hop distance of a node from the granule center in a granule. The value of granule radius depends on the user preference and problem in hand.
- **Granule Representative**: The node around which the granule is being formed is designated as granule representative. The distance term $d(c, v)$ in Eq. 1 is calculated with respect to granule representative.
- **Granular Degree**: Granular degree of a granule centered at node c is defined as the cardinality of the fuzzy set A_c that represents the granule. Mathematically it can be expressed as follows:

$$D(A_c) = |A_c| = \sum_{v \in V} \mu_{A_c}(v, r) \tag{1}$$

  where $\mu_A(v, r)$ is the membership value of node v in granule A_c and r is the granular radius for the granule A_c
- **θ-Core**: A granule A_c can be represented as θ-core, if granular degree is greater than or equals to a threshold value θ, *i.e.,* $D(A_p) \geq \theta$
- **Neighbourhood of a granule**: Set of all granules are said to be neighborhood for A_c, if their centers lies in the support set of A_c. It can be defined as below:

$$Nb(A_c) = \{A_i | i \in Support(A_c) \text{ and } i \neq c\} \tag{2}$$

  where $N_b(A_c)$ is the set of neighboring granules for A_c. The support set of A_c can be defined as follows:

$$Support(A_c) = \{v | \overline{\mu_A}(v, r) > 0\} \tag{3}$$

- **Normalized granular embeddedness**: The normalized granular embeddedness of two granules, A_p and A_q is defined as the ratio of union and intersection of the fuzzy sets that represents the granules. It can be expressed as follows:

$$\varepsilon(A_p, A_q) = \frac{|A_p \bigcap A_q|}{|A_p \bigcup A_q|} \tag{4}$$

  The value of ϵ lies in the range from 0 and 1.

5 Fuzzy-Rough Community (FRC) Detection

5.1 Problem Statement

A social network may be considered, consisting of three parameters (C, V, G), and two constant θ, and, ϵ, where C is the set of communities, V is the set of users or entities and G the set of granules. The objective is to identify a community C such that it is a non-empty subset of granules G that satisfies the following constraints:

- $\forall A_p, A_q \in C, A_p$ and A_q are the community connected θ-cores.
- $\forall A_p \in C, f_{nge}(A_p, \bigcup_{A_q \in C \setminus A_p} A_q) > \epsilon$

where f_{nge} is the function that computes the normalized granular embeddedness of two granules. After obtaining the set of communities C, influential score of each community has been quantified.

5.2 Proposed Algorithm

Unlike conventional community detection methods Fuzzy-Rough Community (FRC) has the ability to capture the behavior of real world network. In this approach, the social network is modeled into a fuzzy-granular social network [6,7]. Similar to the normal perception of associating nodes with high connectivity to the same community, FRC detection considers granules with similar features belonging to the same community. Few terms as noted below help to understand the algorithm in a better way [6]:

Fuzzy-Rough Community. Let n communities to be discovered in a social network be $C_1, C_2, C_3, ..., C_n$, and the upper and lower approximation of the ith community be $\overline{C_i\theta}$ *and* $\underline{C_i\theta}$ respectively. Then,

$$\underline{C_i\theta} = \{x | x \in Support(A_p) \wedge x \notin Support(A_q C_i)\} \tag{5}$$

where $\forall A_p \in C_i$ *and* $A_q \in C_j; i \neq j$

$$\overline{C_i\theta} = \{x | x \in Support(A_p); A_p \in C_i\} \tag{6}$$

Fuzzy-Rough Membership. The membership function, that characterizes the Fuzzy-rough community C_i can be expressed as follows:

$$\delta^{\theta}_{C_i}(x, r) = \begin{cases} 1 & \text{if } x \in \underline{C_i\theta} \\ \sum_{c \in C_i\theta} \overline{\mu_c}(x, r) & \text{if } x \in \overline{C_i\theta} \setminus \underline{C_i\theta} \\ 0 & otherwise \end{cases} \tag{7}$$

For a given social network, communities are identified based on the granular computing approach. In the proposed algorithm, various communities are detected with the help of fuzzy and rough-set representation defined over a granular model of social network. The isolated node that are not included in any of

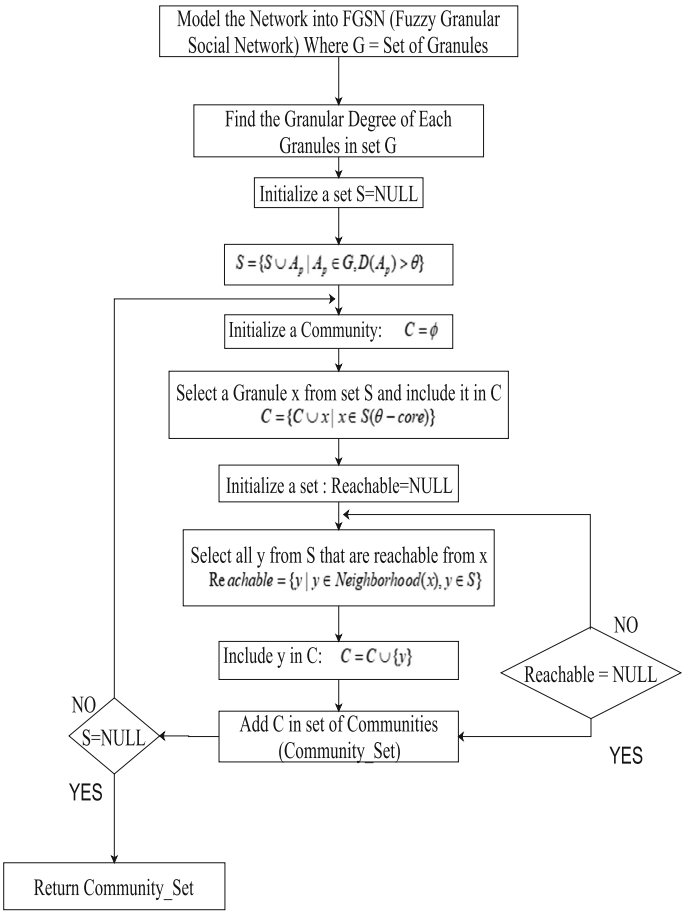

Fig. 1. Execution flow for fuzzy-rough community detection

the detected communities are being treated as outliers. The proposed algorithm i.e., FRC-FGSN is presented in Algorithm 1. For the sake of simplicity and low space, the detail map-reduce implementation of all-pair shortest path and eigen vector centrality as mentioned in step-2 and step-10 respectively in algorithm 1 is omitted. The overall execution flow for enumerating all the communities is presented in Fig. 1.

6 Implementation

6.1 Dataset Used

The proposed algorithm has been experimented on a small-scale network (<1000 nodes and <10000 edges) and three real world large-scale networks (>1000 nodes

Algorithm 1. Map-Reduce based FRC-FGSN Algorithm

1: Transform the input dataset in edge list format into adjacency list format for ease of processing in map-reduce model.
2: All pairs shortest path of the network using a modified ring-search method [22] are obtained using three pair of map-reduce functions.
3: Granular embeddedness of every pair of granule are calculated.
4: Granules with granular degree less than θ i.e. $D(A_p) < \theta$ are filtered.
5: The θ-cores which constitute communities are obtained.
6: $A_c \leftarrow \theta$-cores are initialized
7: The membership value of a node in a community, is calculated
8: The nodes which belong to the lower approximation of a community are filtered out.
9: Normalised community overlap factor is obtained.
10: Eigenvector centrality is being calculated using map-reduce based distributed algorithm.
11: Output: return the set of communities.

and >10000 edges). Small dataset is used to optimize the value of θ and ϵ, which are the major parameters of the FRC-FGSN algorithm. The value of θ and ϵ usually vary from dataset to dataset. However for the sake of simplicity the parameter value has been optimized for the small dataset only and the same value has been fixed for other datasets. The description of the datasets are listed in Table 1.

Table 1. Datasets used for the experiment

Dataset	No. of vertices	No. of edges	Size	Type
Dolphin	62	152	Small scale	Undirected
ego-Facebook	4039	88234	Large scale	Directed
ego-Gplus	107, 614	13, 673, 453	Large scale	Undirected
ego-Twitter	81, 306	1, 768, 149	Large scale	Directed

6.2 Experimental Setup

The proposed algorithm is implemented in Hadoop distributed platform. The cluster used in distributed platform consists of five nodes with symmetric configuration. Each node is of i7 processor with 3.4 GHz clock speed. The secondary memory space and main memory space in each system is 1 TB and 20 GB respectively. In the cluster, one system is dedicated for master node and rest of the node act as slave or datanode. The master node is configured as namenode as well as datanode so that it can also contribute to the data processing whenever it is required.

6.3 Normalized Community Overlap Factor

In order to quantify the influential communities we have introduced a measure known as the normalized community overlap factor, $N_{p,q}$. It is defined as the ratio between the cardinality of the fuzzy intersection and the fuzzy union of the two communities p and q.

$$N_{p,q} = \frac{|A_p \bigcap A_q|}{|A_p \bigcup A_q|} \tag{8}$$

After community detection process is completed, the network is transformed into an un-directed weighted network. In this new network the nodes are represented by the community representatives (granule representatives with highest granular degree in the community) and the edge weights are represented by the normalized community overlap factor ($N_{p,q}$) as given in Eq. 8. Thus the resulting network is the network on which the eigenvector centrality algorithm is run to get the centrality values of the nodes.

The proposed influential community detection algorithm (starting from FRC detection to influence calculation) has been implemented in map-reduce fashion in order to make it scalable. The program comprises of 14 mapper-reduce pairs. The sequence of steps followed are presented in Algorithm 1.

Table 2. Observations for varying θ and ϵ

Serial no.	θ		ϵ		No. of communities	Convergence
	Quartile	Value	Quartile	Value		
1	Lower	0.7563	Lower	0.1151	2	Yes
2	Lower	0.7563	Middle	0.3647	3	No
3	Lower	0.7563	Upper	0.5503	7	Yes
4	Middle	1.0415	Lower	0.1151	2	Yes
5	Middle	1.0415	Middle	0.3647	3	No
6	Middle	1.0415	Upper	0.5503	5	Yes
7	Upper	1.1405	Lower	0.1151	2	Yes
8	Upper	1.1405	Middle	0.3647	3	No
9	Upper	1.1405	Upper	0.5503	2	Yes

7 Results and Observations

Identification of fuzzy-rough communities depends on two parameters namely, granular embeddedness threshold (ϵ) and granular degree (θ). The proposed algorithm has been executed on a small-scale network and three large scale networks. Since smaller networks are easier to visualize and analyze, a small

scale network was used to optimize the value of ϵ and θ. The results obtained have been presented in Table 2.

The experiments have been performed on one small scale network and three large scale networks. The small-scale network used in this work is the Dolphin social network [20,21]. Since the optimal values of ϵ and θ are unknown, three different combinations of θ and ϵ are chosen for running the experiments. The value of maximum granular embeddedness (ϵ_{max}) and the maximum possible value of granular degree (θ_{max}) might vary from one network to another. In order to generalize the optimal values of ϵ and θ, the values of θ and ϵ have been considered as the lower quartile, median and upper quartile from the range of values of θ and ϵ. The radius of the granules were set to three. Since real-world social networks show small-world property the average diameter of such networks is fixed to six [22,23]. Searching for optimal values of ϵ and θ is important because if the values are not optimal two possible problems may be encountered:

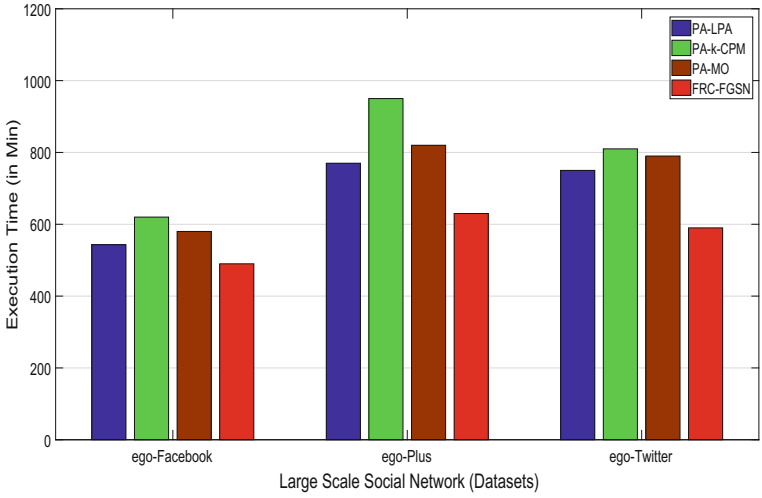

Fig. 2. Execution time (in Mins) for different algorithms

1. The network of community representatives obtained after the community detection is completed, may not have a dominant eigen value in its adjacency matrix form. Since the eigenvector centrality is calculated using power method it is not possible to obtain eigenvector for a matrix with more than one dominant eigen value [24].
2. The number of communities is either too small or too big.

Nine observations have been considered with varying values of θ and ϵ (lower, middle and upper quartile). The results are presented in Table 1.

From the Table 2, it can be observed that the number of communities obtained depends on the value of θ and ϵ. It can also be observed that the

power method used for obtaining eigenvector centrality is converged only when value of θ and ϵ lie in the range of lower-quartile and upper-quartile respectively.

After obtaining the optimal values for θ and ϵ the algorithm was being executed for large scale networks like Facebook, Google circle and Twitter [25]. The proposed algorithm has been compared with few other community detection algorithms based on execution time. As the FRC-FGSN algorithm is implemented in distributed platform, for comparison we have considered the parallel version of Label Propagation (PA-LPA), k- Clique percolation Method (PA-k-CPM) and Modularity Optimization (PA-MO) algorithms. The execution time has been compared on large scale network only as for small scale network, the FRC-FGSN algorithm might take longer time than expectation due to communication latency in clustering setup. The execution time is observed to be least for FRC-FGSN as shown in Fig. 2.

8 Conclusion

In this work, we have presented an efficient algorithm (FRC-FGSN) to identify the fuzzy rough communities in granular social network. Real world social networks are modeled into fuzzy granular social network (FGSN), where the basic entity for processing is fuzzy granule rather than a single node. The identified communities are characterized by crisp lower and fuzzy upper memberships which are called as fuzzy-rough communities. Each node is associated with fuzzy membership value. A node is said to be member of a community, if its membership value lies in the range of boundary region *i.e.,* Upper-Lower. A node may belong to multiple communities. The proposed approach is suitable for exploring overlapping communities. We have also detected the outliers in the network which may not belongs to any communities. Communities are then quantified based on the eigen-centrality measures. We have implemented the model using distributed programming paradigm like map-reduce. The extensive comparison of FR-FGSN has been made with other parallel algorithms which conclude that, it has less execution time as compared to other.

In future some other distributed framework like spark can be adopted to measure the performance of the model. Other parameters like NMI, AUC, Accuracy may be considered in future to extend the experimental analysis.

References

1. Travers, J., Milgram, S.: The small world problem. Phychol. Today **1**, 61–67 (1967)
2. Stephen, A.T., Toubia, O.: Explaining the power-law degree distribution in a social commerce network. Soc. Netw. **31**(4), 262–270 (2009)
3. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3), 75–174 (2010)
4. Albert, R., Jeong, H., Barabási, A.-L.: Internet: diameter of the world-wide web. Nature **401**(6749), 130–131 (1999)
5. Broder, A., et al.: Graph structure in the web. Comput. Netw. **33**(1), 309–320 (2000)

6. Kundu, S., Pal, S.K.: Fuzzy-rough community in social networks. Pattern Recogn. Lett. **67**, 145–152 (2015)
7. Kundu, S., Pal, S.K.: FGSN: fuzzy granular social networks-model and applications. Inf. Sci. **314**, 100–117 (2015)
8. Bonacich, P.: Some unique properties of eigenvector centrality. Soc. Netw. **29**(4), 555–564 (2007)
9. Landherr, A., Friedl, B., Heidemann, J.: A critical review of centrality measures in social networks. Bus. Inf. Syst. Eng. **2**(6), 371–385 (2010). https://doi.org/10.1007/s12599-010-0127-3
10. McAfee, A., Brynjolfsson, E., Davenport, T.H., et al.: Big data: the management revolution. Harvard Business Rev. **90**(10), 60–68 (2012)
11. Zaharia, M., et al.: Fast and interactive analytics over Hadoop data with Spark. USENIX Login **37**(4), 45–51 (2012)
12. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)
13. Newman, M.E.: Fast algorithm for detecting community structure in networks. Phys. Rev. E **69**(6), 066133 (2004)
14. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Phys. Rev. E **76**(3), 036106 (2007)
15. Li, R.-H., Qin, L., Yu, J.X., Mao, R.: Influential community search in large networks. Proc. VLDB Endow. **8**(5), 509–520 (2015)
16. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146. ACM (2003)
17. Anagnostopoulos, A., Kumar, R., Mahdian, M.: Influence and correlation in social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 7–15. ACM (2008)
18. Zadeh, L.A.: Fuzzy sets. In: Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh, pp. 394–432. World Scientific (1996)
19. Behera, R.K., Shukla, A., Mahapatra, S., Rath, S.K., Sahoo, B., Bhattacharya, S.: Map-reduce based link prediction for large scale social network. In: Proceedings of the 29th International Conference on Software Engineering and Knowledge Engineering, pp. 1–4. IEEE (2017)
20. Dolphins network dataset - KONECT, April 2017
21. 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. Behav. Ecol. Sociobiol. **54**, 396–405 (2003). https://doi.org/10.1007/s00265-003-0651-y
22. de Sola Pool, I., Kochen, M.: Contacts and influence. Soc. Netw. **1**(1), 5–51 (1978)
23. Zhang, L., Tu, W.: Six degrees of separation in online society. In: Proceedings of the WebSci 2009, Society On-Line, 18–20 March 2009, Athens, Greece, pp. 1–5 (2009)
24. Davidson, E.R.: The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. J. Comput. Phys. **17**(1), 87–94 (1975)
25. Leskovec, J., Mcauley, J.J.: Learning to discover social circles in ego networks. In: Advances in Neural Information Processing Systems, pp. 539–547 (2012)