Overlapping community detection in social networks using a quantum-based genetic algorithm

Alireza Saleh Sedghpour
School of Computer Engineering
Iran University of Science and Technology
Tehran, Iran
alireza_ssp@yahoo.com

Amin Nikanjam
Faculty of Computer Engineering
K.N. Toosi University of Technology
Tehran, Iran
nikanjam@kntu.ac.ir

ABSTRACT
One of the main research areas in social networks is community detection among them. In this research, the structure of communities of a social network considering their overlap has been detected with reliance on the quantum-based genetic algorithm (QIGA). This is a new dual-phase algorithm for community detection as an optimization problem. This algorithm has taken advantage of the modularity function as the target function. In order to detect communities considering their overlap, Jaccard's coefficient and the Jaccard median have been used in pre-processing stage. Results indicate that different amounts of overlap and modularity are achievable by incorporating a coefficient into the existing correlation between Jaccard's coefficient and the Jaccard median. This coefficient has been regulated using the three networks of karate, dolphin, and soccer. The huge air traffic control network has also benefitted from the mentioned regulated number. Implementation outcomes show that the suggested algorithm is able to detect community structure in overlapping networks with pinpoint accuracy.

CCS Concepts
• Theory of computation → Evolutionary algorithms

Keywords
Social networks, community structure, modularity, quantum-based genetic algorithm, overlapping communities.

1. INTRODUCTION
For determining communities in social network, one of the innovative methods is the GN algorithm proposed by Girvan and Newman [1].The CNM algorithm has been proposed by Clauset and his colleagues [2] to improve the performance of the GN-Fast algorithm that renders a widespread analysis of community structure in networks.

Since one of the most significant research areas nowadays is the expansion of searching methods based on natural evolution principles, and considering the fact that abstract evolutionary calculations have been inspired by those principles to find optimum solutions for various problems, researchers have jumped on evolutionary algorithms. There are evolutionary algorithms in this scope namely: DA or EO algorithm [3], GATHB genetic algorithm [4], MOGA-Net genetic algorithm [5], QIGA genetic algorithm [6].

The QIGA is a new 2-stage algorithm introduced by Gupta and his colleagues for community detection as an optimization problem. The proposed algorithm deploys the community modularity as a target function for the purpose of measuring community relativity assuming that communities do not overlap in social networks. At the first stage, each node, *i*, is marked to be located in the community it belongs to. At the second stage, a hierarchical segmentation strategy is deployed which iteratively divides the network into halves and maximizes the modularity (Q) of the network. This stage has benefited from Quantum Inspired Genetic Algorithm (QIGA). The proposed method does not need any background knowledge about the number of communities and works properly both with directed and undirected networks. Experiments on virtual and real life networks indicate that this method is able to manifest the community structure successfully at a high modularity.

In this research, we aim to detect the structure of communities in a social network with reliance on quantum genetic algorithm introduced in [6]. In this regard, overlap of communities will also be taken into account in order to improve the accuracy of results.

2. THE PROPOSED METHOD
Community detection method with deployment of quantum genetic algorithm including two major stages:

Phase one: Pre-processing

Phase two: QIGA algorithm

2.1 Phase one: Pre-processing
In this phase, the nodes with a higher chance of being located in the same community are classified. The first stage of this phase starts with each node i in its own community and detects the nodes farthest from the two edges crossing the node i. P-neighborhood of the node i has been deployed for this purpose. Neighborhood-1 (N1) and neighborhood-2 (N2) could be obtained through the following operations:

$$N^1(i) = \{i\} \cup \{j | A_{ij} \neq 0\} \cup \{j | A_{ji} \neq 0\}$$

$$N^2(i) = N^1(i) \cup \{j | \exists k \in N^1(i), (A_{jk} \neq 0 \vee A_{kj} \neq 0)\} \quad (1)$$

The Jaccard's coefficient (Jij) has been used to calculate neighborhood overlap. Nodes j will be added to the node i's community if Jij is greater than a specific coefficient of overlap average for node i (Ji):

$$J_{ij} > \beta\alpha J_i \quad (2)$$

GECCO '17 Companion, July 15-19, 2017, Berlin, Germany

ACM ISBN 978-1-4503-4939-0/17/07.
http://dx.doi.org/10.1145/3067695.3076000

In which the amount of the parameter $\beta\alpha$ is obtained with regard to experimental results of various data. It is essential to point out that the reference [6] has stated the above in equation as $J_{ij} > \alpha J_i$ and determined the amount of the parameter α within the range of 2 and 3 (the accurate amount used in [6] equals $\frac{ln(n)}{ln(4.9*ceil(\frac{n}{490}))}$, n is the node number and ceil means rounding up). With respect to the numerous implementations in this research, the amount of α turned out to be different from that used in this reference so overlapping state is resulted. Thus, for the totality of the above relations not to be trivialized as well as the parameter α used in this reference to remain intact, this inequation was multiplied by a coefficient namely β whose amount regarding numerous implementations on various networks must be somewhere between 0 and 1.

2.2 Phase two: QIGA algorithms

This phase proceed as the same as [6].

3. Results

The proposed algorithm has been tested on karate, soccer, dolphin, and air traffic control social networks. Parameter β must be regulated with regard to the experimental outcomes of different networks.

The value of parameter β has been modified between 0 and 1 to obtain its effects on the following parameters in pre-processing stage: Modularity value, Number of communities, Percentage of modularity changes, Number of overlapping nodes and Percentage of overlap.

The obtained number of communities, modularity value, and modularity changes in percentage for different values of β have been illustrated in Figure 1. According to this diagram, there is obviously no change in modularity for β=1 because our basis is the variance of modularity for β=1. For β>0.8, the amount of modularity changes is lower than 5%. Also, for β<0.3 more than 30% of modularity value is changed (Minus sign indicates a decrease in modularity).

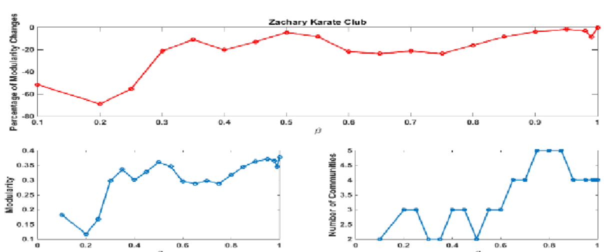

Figure 1. The influence of parameter β on modularity values, number of communities and percentage of modularity changes.

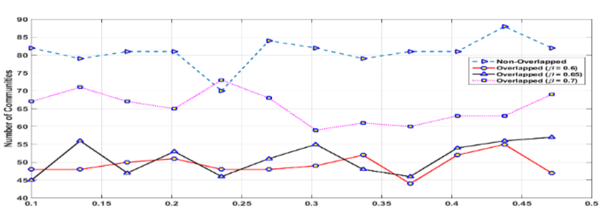

Figure 2. Number of communities of air traffic control network in overlapping state for β=0.6, 0.65, 0.7 and non-overlapping state according to parameter μ.

According to the above analyses, the value β>0.3 could be proposed for karate network, and the range 0.3<β<0.8 could also be a proposal if at least 5% of overlapping is also desired.

The following outcomes were extracted from implementations.

Given the chosen range in Table 1, implementation of air traffic control network has been done according to values β=0.6, 0.65, 0.7. Figure 2 illustrates the number of communities of air traffic control network in overlapping state for β=0.6, 0.65, 0.7 and non-overlapping state according to parameter μ. Almost for all values of μ, the number of communities is the highest in non-overlapping state and the lowest when β=0.6, 0.65.

Table 1. The final range of parameter β

Type of Network	Condition	Range
All networks	With stipulation of minimum 5% overlapping	0.6<β<0.75

4. CONCLUSION

According to the outcomes of implementation, overlapping and non-overlapping states are obviously different. In this research, an overlapping-based model was presented for community detection in social networks which is able to predict the existing overlap in communities and distinguish them from each other properly. Furthermore, given the high optimization ability of smart algorithms such as quantum genetic algorithm, the outcomes of implementation could be claimed to be of pinpoint accuracy and be taken into account as a powerful tool for community detection and pursuing further researches.

5. REFERENCES

[1] M. Girvan and M. EJ. Newman, Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, Vol. 99, No.12, pp. 7821-7826, 2002.

[2] A. Clauset, M. EJ. Newman and C. Moore, Finding community structure in very large networks, *Physical review E* 70, No. 6, 2004.

[3] J. Duch and A. Arenas, Community detection in complex networks using extremal optimization. *Physical review E* 72, No. 2, 2005.

[4] M. Tasgin, A. Herdagdelen and H. Bingol, Community detection in complex networks using genetic algorithms. *arXiv preprint* arXiv:0711.0491, 2007.

[5] C. Pizzuti, A multiobjective genetic algorithm to find communities in complex networks. *IEEE Transactions on Evolutionary Computation*, Vol. 16, No. 3, pp. 418-430, 2012.

[6] S. Gupta, S. Taneja and N. Kumar, Quantum inspired genetic algorithm for community structure detection in social networks. *Proceedings of the 2014 conference on Genetic and evolutionary computation*, ACM, pp. 1119-1126, July 2014.