INTELLIGENT COMMUNITY DETECTION: REVIEW

Jenan Moosa jhasan@ahlia.edu.bh
Dr. Wasan Awad wawad@ahlia.edu.bh
Dr. Tatiana Kalganova tatiana.kalganova@brunel.ac.uk

Abstract
An emerging multidisciplinary field nowadays is graph mining; which is basically a form of data mining that deals with graphs instead of normal data, (Cook & Holder, 2006) it intends to discover repetitive sub-graphs and interesting patterns that occurs in the input graph. This research is focusing on community detection using graph mining techniques; the objectives are to study the existing methods used to solve this problem by comparing the evaluation parameters, and then develop a method for community detection in social networks, which meets the criteria identified for assessing a good algorithm. Various techniques of graph mining will be tested to determine wither it can meet the identified criteria.
Keywords: Graph Mining, Community Detection.

1. Introduction

Living in an era of technology, a dramatic amount of data is being produced; according to the world internet usage and population statistics 2019, there are over 4.5 billion users on the internet (World Internet Usage and Population Statistics, 2019). The expansion use of the internet by people of different ages, backgrounds and interests, and their engagement in the social media accounts, rich social media content is being generated. As a result, even a bigger massive amount of data will be created in the coming years. Having all this amount of data will not only help in performing the task it was created for, as it can also facilitate in executing new functions and outcomes; when first created, the data is raw and senseless. However, instead of neglecting it, raw data and its' relationships can be transformed into meaningful information and structures when applying the appropriate mining tools. Here where Graph mining plays a significant scope, as it helps analysing the networks and extracting useful information.

2. Problem Formulation

The ideal way to analyse a network is to represent it as a graph; it is the closest formulation to the real-world situation. Using graphs to represent a network can simplify the problem as it provides different points of view, in addition to the appropriate tools that can be used to solve the problem. A social network can be represented as a graph G(V,E); where V is the set of vertices (users) and E is the set of edges (which is any sort of relationships) (Mishra et al. 2014). Which means that the users will be denoted by nodes; and since there are relationship between the users, the nodes will be connected together, and the relationships will be denoted by the edge and this graph representation will help analysing the social network.

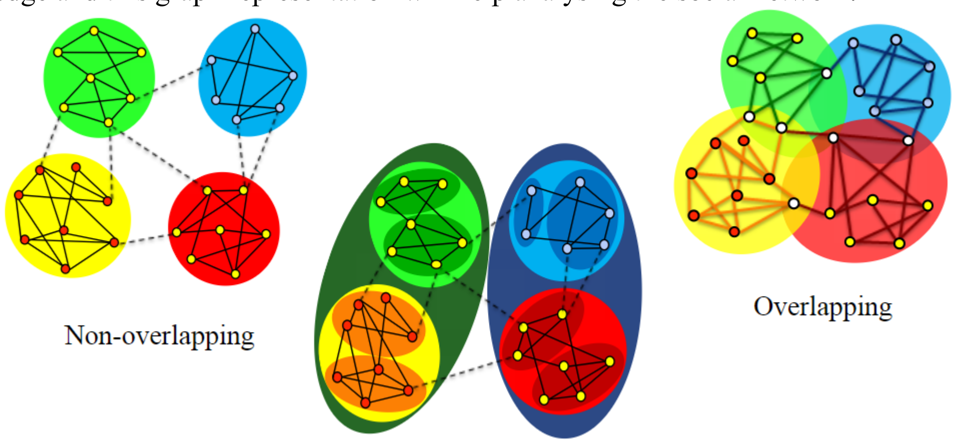

Figure1: Problem Detection of Community Detection (Moradi, Olovsson, & Tsigas)

3. Literature Review & Related Work

Community detection is being widely studied, and different algorithms and techniques have been proposed so far, which includes traditional algorithms, evolutionary algorithms etc. Most of these algorithms designed to discover disjoint communities, assuming that nodes in each network belong to exactly one community. Bedi & Sharma (2016) stated that any network can be outlined in a graph; (Bedi & Sharma, 2016) the graph is composed of a set of nodes which can be individuals or entities, and edges that represent the connections and interactions between the nodes. Structuring a social network by a graph can result in performing a number of tasks (Tang & Liu, 2010), such as centrality analysis, position/role analysis, network modelling, information diffusion, network classification and outlier detection, viral marketing and link prediction, and community detection.

In this paper we will focus on Community detection; as it recognizes the communities formed by the actors in a social network, by studying the network topology. (Papadopoulos, Kompatsiari, Vakali, & Spyridonos, 2012) Community detection is an important tool for analysing complex networks by studying the structures and functional characteristics of the networks (Bedi & Sharma, 2016).

A network is consisted of connected communities that are created by the communicated individuals, and a community is a densely connected group of nodes that is also connected to the rest of the network (Steve Harenberg, et al., 2014). In different research fields, a community is also referred to as group, cluster, cohesive subgroup or module. Communities can be defined based on the context. The main goal of community detection is to detect and discover the strongly connected members as compared to the rest of the members, and thereby form the people of similar taste, choices and preferences in a virtual cluster or a community. (Verma & Bharadwaj, 2017) The extracted communities in the community detection procedure can considerably analyse building ontology for semantic web, detecting topics in tagging systems, analysing the behaviour of a community, and personalized search and recommendations.

An example of community detection application is social network. Where Social Network Analysis (SNA) refers to the study of relations and connections between nodes, (Wassermanx & Faust 1994) and it includes the analysis of social structures, positions, roles and many others. SNA has emerged out of the need to better understand social networks, by exploring how people are connecting and determining the network's strength and weakness by calculating the weights of edges; which could be frequency of interaction between nodes, number of items exchanged, individual perceptions of strength of relationship , cost in communication/ exchange, or a combination of these. (Cheliotis 2010).

In real world, a node (or a person) in a certain network, does not necessarily belong to exactly one community; as a person can be in more than one community at a time, and here is where overlapping occurs- which is illustrated in Figure1. Originally, a community consists of densely connected nodes, so intuitively, when a node connects multiple communities with similar strength, it is more likely to be an overlapping node (Cai, et al., 2014). For instance, if node i has both l links with community a and b, then we can regard i as an overlapping node.

Table1 illustrates some of the algorithms used to solve the community detection problem. It shows that researchers have classified community detection algorithms in several ways depending on the aim of their research, however, the main aim is to detect communities or structures in social networks. While some focused on large scale, complexed or weighted networks, other researchers focused on detecting overlapped communities without taking into consideration the type or size of the network. We can also observe most of them used the evolutionary algorithms such as genetic algorithms, particle swarm and neural networks. Other approaches were also used such as quantum algorithms and neural networks.

Table 1 previous related work summary

Reference	Aim	Algorithm/Technique
(Rahimi, Abdollahpouri, & Moradi, 2018)	Solve the graph clustering problem by detecting the structures which are closer to real ones	Particle swarm optimization
(Moayedikia, 2018)	Detect communities, incorporate nodes' attribute	Node Importance Analysis

	information and estimating nodes' importance	
(Ebrahimi, Shahmoradi, Heshmati, & Salehi, 2018)	Use all Pareto fronts to detect overlapping communities	Pareto fronts
(Yuanyuan & Xiyu, 2018)	Discover communities in social networks by optimizing modularity	Quantum inspired evolutionary algorithm
(Cheng, Cui, Su, Niu, & Zhang, 2018)	Improve the quality (accuracy) of community detection by utilizing local communities	A local information based multi-objective evolutionary algorithm
(Zhang, et al., 2018)	Detect communities in Large-Scale Complex Networks	A Network Reduction-Based Multiobjective Evolutionary Algorithm
(Verma & Bharadwaj, 2017)	Identify community structure in a multi-relational network through relational learning	GA k-means clustering algorithm
(Gupta, et al., 2017)	Detect high-quality communities for varied datasets and work well for both weighted and un-weighted networks.	Parallel quantum-inspired
(Wen, et al., 2017)	Detected overlapping communities	A Maximal Clique
(Reihanian, Feizi-Derakhshi, & Aghdasi, 2017)	- Detect communities of social networks with node attributes - Measure the similarity of node attributes in a community of a network	multi-objective biogeography based optimization
(Said, Ayaz Abbasi, Maqbool, Daud, & Aljohani, 2018)	Detect communities in social and complex network by generating the initial population and the mutation method.	Genetic Algorithm
(Guerrero, Montoya, Baños, Alcayde, & Gil, 2017)	Enable a flexible and adaptive analysis of the characteristics of a network from different levels of detail according to an analyst's needs.	
(Liu & Li, 2017)	Optimize the modularity for community detection with local search Strategy	
(Ji, Zhang, & Zhou, 2019)	Solve complex network community detection problems	Ant colony
(Bruna & Li, 2017)	Observe instances of graphs together with their true community structure and attempt to learn a mapping between graphs and their predicted communities.	Neural networks

There is a wide number of techniques used to solve the community detection problems, and those were classified in a multidisciplinary review (Javed, Younis, Qadir, & Baig, 2018) and illustrated in Figure2.

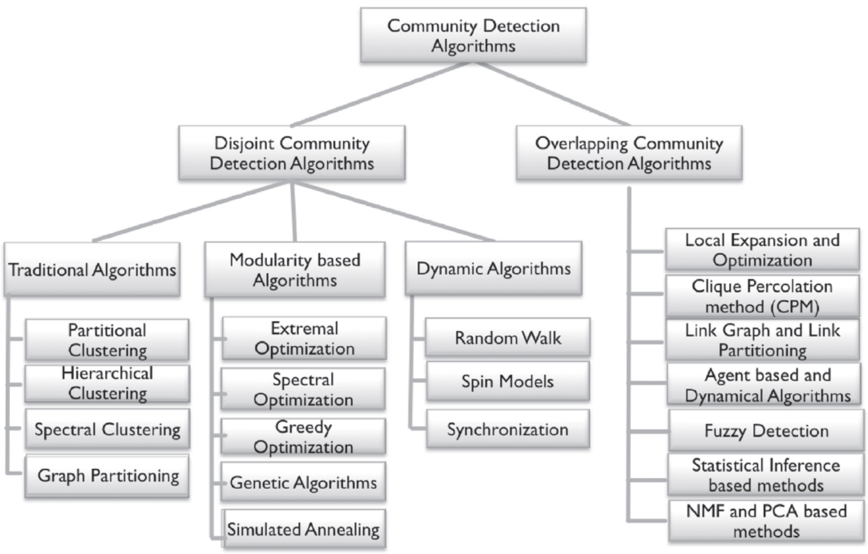

Figure2: Classification breakdown of algorithms for community detection, (Javed, Younis, Qadir, & Baig, 2018)

The main difference between disjoint and overlapping community detection algorithms is that the latter can discover overlapped communities, where nodes can belong more than one community at the same time; as a node can share the same taste or preferences of more than one group of people. Although overlapping is an important aspect in social networks if we tend to reflect the real-wold social networks were users interact with one another in a humanly way, this classification shows that not many techniques were used to solve the overlapping problem; and by reviewing the non-overlapping techniques, most of them stated that their future work will focus on finding the overlapping in the networks, as an example: (Said, et al., 2018), (Bilal & Abdelouahab, 2017), (Reihanian, et al., 2017), (Gupta, et al., 2017), (Verma & Bharadwaj, 2017). In our research we will concentrate on detecting the overlapped communities within the network, by applying hybrid intelligent techniques, which could be combining disjoint and overlapping methods or more than one overlapping method.

4. Research question

Can the accuracy of community detection in social networks be improved by using hybrid intelligent techniques? Does combining more than one method make good use of different efficient techniques for better performance?

5. Research methodology

1. Review the related work and available techniques previously to solve the community detection problem.
2. Design the proposed algorithm to solve the problem.
3. Implement the proposed algorithm.
4. Analyse the results by evaluating the algorithm using the evaluation methods used by other researchers.

6. Conclusion

In this paper we have reviewed the definition of community detection problem. And we focused on social networks as an application. This paper investigates the different approaches and algorithms used to detect communities in social network; and that include disjoint and overlapping community detection algorithms. As a future work, we intend to design a hybrid intelligent technique to improve the accuracy of community detection in social networks.

REFERENCES

Bilal, S., & Abdelouahab, M. (2017). Evolutionary algorithm and modularity for detecting communities in networks. *Physica A, 473*, 89–96.

Wen, X., Chen, W.-N., Lin, Y., Gu, T., Zhang, H., Li, Y., . . . Zhang, J. (2017). A Maximal Clique Based Multiobjective Evolutionary Algorithm for Overlapping Community Detection. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, 21*(3), 363-377.

Zhang, X., Zhou, K., Pan, H., Zhang , L., Zeng, X., & Jin, Y. (2018). A Network Reduction-Based Multiobjective Evolutionary Algorithm for Community Detection in Large-Scale Complex Networks. *A Network Reduction-Based Multiobjective Evolutionary Algorithm for Community Detection in Large-Scale Complex Networks*.

Bedi, P., & Sharma, C. (2016, Feb 19). Community detection in social networks. *WIREs Data Mining Knowledge Discovery*, pp. 115–135.

Bruna, J., & Li, X. (2017). Community Detection with Graph Neural Networks. *stat*, 27.

Cai, Q., Ma, L., & Gong, M. (2014). A survey on network community detection based on evolutionary computation. *International Journal of Bio-Inspired Computation*.

Cai, Y., Shi, C., Dong, Y., Ke, Q., & Wu, B. (2011). A Novel Genetic Algorithm for Overlapping Community Detection. *International Conference on Advanced Data Mining and Applications. 7120*, pp. 97-108. Berlin, Heidelberg: Springer.

Chen, D., Zou, F., Lu, R., Yu, L., Li, Z., & Wang, J. (2016). Multi-objective optimization of community detection using discrete teaching–learning-based optimization with decomposition. *Information Sciences, 369*, 402-418.

Cheng, F., Cui, T., Su, Y., Niu, Y., & Zhang, X. (2018). A local information based multi-objective evolutionary algorithm for community detection in complex networks. *Applied Soft Computing, 69*, 357-367.

Cook, D. J., & Holder, L. B. (2006). *Mining Graph Data.* John Wiley & Sons.

Du, N., Wu, B., Pei, X., Wang, B., & Xu, L. (2007). Community Detection in Large-Scale Social Networks. *the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 16-25). San Jose, California: ACM.

Ebrahimi, M., Shahmoradi, M. R., Heshmati, Z., & Salehi, M. (2018). A novel method for overlapping community detection using Multi-objective optimization. *Physica A, 505*, 825–835.

Ghaffaripour, Z., Abdollahpouri , A., & Moradi, P. (2016). A multi-objective genetic algorithm for community detection in weighted networks. *Eighth International Conference on Information and Knowledge Technology (IKT)* (pp. 194-199). Hamedan, Iran: IEEE.

Gong, M., Ma, L., Zhang, Q., & Jiao, L. (2012). Community detection in networks by using multiobjective evolutionary algorithm with decomposition. *Physica A: Statistical Mechanics and its Applications, 931*(15), 4050–4060.

Guerrero, M., Montoya, F. G., Baños, R., Alcayde, A., & Gil, C. (2017). Adaptive community detection in complex networks using genetic algorithms. *Neurocomputing, 266*, 101-113.

Gupta, S., Mittal, S., Gupta, T., Singhal, I., Khatri, B., Gupta, A. K., & Kumar, N. (2017). Parallel quantum-inspired evolutionary algorithms for community detection in social networks. *Applied Soft Computing, 61*, 331-353.

Hafez, A. I., Zawbaa, H. M., Hassanien, A., & Fahmy, A. (2014). Networks community detection using artificial bee colony swarm optimization. *the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA* (pp. 229-239). Cham: Springer.

Hassan, E. A., Hafez, A. I., Hassanien, A., & Fahmy, A. A. (2015). Community Detection Algorithm Based on Artificial Fish Swarm Optimization. *Intelligent Systems'2014*, 509-521.

Javed, M. A., Younis, M. S., Qadir, J., & Baig, A. (2018). Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 87–111.

Ji, P., Zhang, S., & Zhou, Z. (2019). A decomposition-based ant colony optimization algorithm for the multi-objective community detection. *Journal of Ambient Intelligence and Humanized Computing*.

Li, J., & Song, Y. (2013). Community detection in complex networks using extended compact genetic algorithm. *Soft Computing, 17*(6), 925–937.

Li, Y., Wang, Y., Chen, J., Jiao, L., & Shang, R. (2015). Overlapping community detection through an improved multi-objective quantum-behaved particle swarm optimization. *Journal of Heuristics, 21*(4), 549-575.

Liu , H., Yang, F., & Liu, D. (2016). Genetic algorithm optimizing modularity for community detection in complex networks. *The 35th Chinese Control Conference* (pp. 1252-1256). Chengdu, China: IEEE.

Liu , S., & Li, Z. (2017). A modified genetic algorithm for community detection in complex networks. *International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET).* Chennai, India: IEEE.

Meng, X., Dong, L., Li, Y., & Guo, W. W. (2017). A genetic algorithm using K-path initialization for community detection in complex networks. *Cluster Computing, 20*(1), 311–320.

Moayedikia, A. (2018). Multi-objective community detection algorithm with node importance analysis in attributed networks. *Applied Soft Computing, 67*, 434–451.

Papadopoulos, S., Kompatsiari, Y., Vakali, A., & Spyridonos, P. (2012). Community detection in Social Media. *Data Mining and Knowledge Discovery, 24*(3), 515–554.

Pizzuti, C. (2008). GA-Net: A Genetic Algorithm for Community Detection in Social Networks. *In: Parallel Problem Solving from Nature–PPSN X.* Springer.

Pizzuti, C. (2009). A Multi-objective Genetic Algorithm for Community Detection in Networks. *International Conference on Tools with Artificial Intelligence, ICTAI*, (pp. 379-386). New Jersey, USA.

Pizzuti, C. (2012). A Multiobjective Genetic Algorithm to Find Communities in Complex Networks. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, 16*(3), 418-430.

Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *76*(3).

Rahimi, S., Abdollahpouri, A., & Moradi, P. (2018). A multi-objective particle swarm optimization algorithm for community detection in complex networks. *Swarm and Evolutionary Computation, 39*, 297–309.

Reihanian, A., Feizi-Derakhshi, M. R., & Aghdasi, H. S. (2017). Community detection in social networks with node attributes based on multi-objective biogeography based optimization. *Engineering Applications of Artificial Intelligence, 62*, 51-67.

Said, A., Ayaz Abbasi, R., Maqbool, O., Daud, A., & Aljohani, N. R. (2018). CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks. *Applied Soft Computing, 63*, 59–70.

Shi, C., Yan, Z., Cai, Y., & Wu, B. (2011). Multi-objective community detection in complex networks. *Applied Soft Computing, 12*(2), 850-859.

Shi, C., Yu, P. S., Cai, Y., Yan, Z., & Wu, B. (2011). On Selection of Objective Functions in Multi-Objective Community Detection. *The 20th ACM international conference on Information and knowledge management* (pp. 2301-2304). ACM.

Steve Harenberg, Bello, G., Gjeltema, L., Ranshous, S., Harlalka, J., Seay, R., . . . Samatova, N. (2014). Community detection in large-scale networks: a survey and empirical evaluation. *Wiley Interdisciplinary Reviews: Computational Statistics, 6*(6), 426-439.

Tang, L., & Liu, H. (2010). Graph Mining Applications to Social Network Analysis. In *Managing and Mining Graph Data.* US: Springer.

Verma, A., & Bharadwaj, K. (2017). Identifying community structurein a multi-relational network employing non-negative tensorfactorization and GA k-means clustering. *WIREs Data Mining Knowl Discoery*, p. e1196.

WORLD INTERNET USAGE AND POPULATION STATISTICS. (2019, June). Retrieved from Internet World Stats: https://www.internetworldstats.com/stats.htm

Xie, J., & Szymanski, B. K. (2012). Towards Linear Time Overlapping Community Detection in Social Networks. *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 25-36). Berlin, Heidelberg: Springer.

Yuanyuan, M., & Xiyu, L. (2018). Quantum inspired evolutionary algorithm for community detection in complex networks. *Physics Letters A, 382*(34), 2305–2312.

Yun, L., Gang, L., & Song-yang, L. (2013). A genetic algorithm for community detection in complex networks. *Journal of Central South University, 20*(5), 1269–1276.

Zadeh, P. M., & Kobti, Z. (2015). A Multi-Population Cultural Algorithm for Community Detection in Social Networks. *The 6th International Conference on Ambient Systems, Networks and Technologies* (pp. 342 – 349). Procedia Computer Science.

Zhu, X., Ma, Y., & Liu, Z. (2018). A novel evolutionary algorithm on communities detection in signed networks. *Physica A, 503* , 938–946.