Are *Evolutionary Computation*-Based Methods Comparable to State-of-the-art non-Evolutionary Methods for Community Detection?

Ami Hauptman
Computer Science Department
Sapir Academic College
M. P. Hof Ashkelon, Israel
amih@mail.sapir.ac.il

ABSTRACT

One important aspect of graphs representing complex systems is community (or group) structure—assigning vertices to groups, which have dense intra-group connections and relatively sparse inter-group connections. Community detection is of great importance in various domains, where real-world complex systems are represented as graphs, since communities facilitate our understanding of the graph and thus of the underlying system. However, this is known to be a hard optimization problem.

In this study we pursue the following question: Have *Evolutionary Computation-Based Methods* proven their worth for this complex domain, or is it currently better to rely on other state-of-the-art methods? While several works compare state-of-the-art methods for community detection (see [8] and [11] for recent surveys), we are unaware of other attempts to compare methods based on evolutionary computation to other methods.

After describing some recent algorithms for this problem, and comparing them in various ways, we conclude that evolutionary computation-based method for community detection have indeed developed to hold their own against other methods for several variants of this problem. However, they still need to be applied to more difficult problems and improve further to make them in par with other methods.

Keywords

Genetic Algorithms; Cluster Analysis; Community Detection

1. INTRODUCTION

Community Detection is the problem of finding groups of strongly connected nodes in a graph, which are relatively weakly connected to nodes outside the group. This problem is of considerable importance for various types of networks, including large social networks (e.g. Facebook), web graphs, biological networks and climate networks, and thus have been studied extensively (see [8] and [11] for recent surveys).

Community detection may reveal important information about the studied system. The most distinct example is that entities which belong to the same community tend to be more similar than entities which do not. This assists in making predictions regarding these entities. Other examples include detecting central and outlier nodes, graph hierarchy and shortest paths.

However, as community detection is closely related to the *clique* problem, it is very hard to solve optimally. Moreover, in the Big Data era very large networks are typically analyzed (for example, social network graphs with millions of nodes), and thus powerful methods are required for detecting communities in reasonable time.

Communities come in different flavors. They can be overlapping (nodes may belong to more than one community)– crisp or fuzzy (different extent of belonging to each), weighted (non-binary degree of membership), hierarchical (large communities can contain smaller ones), directed (based on directed graphs) and even signed (e.g. friend or foe relations). These aspects add yet more degrees of complication.

We now describe and compare two categories of methods employed for this problem: methods based on evolutionary computation, and other state-of-the-art methods.

2. EVOLUTIONARY COMPUTATION-BASED METHODS

We now present some recently published methods based on evolutionary computation.

A Multi-objective Genetic Algorithm for Community Detection in Networks [7] is a method which generated hierarchical network divisions, and automatically detects the number of communities.

Multi-objective Evolutionary Algorithm Based on Decomposition ($MOEA/D$)(e.g. [5]) is a generic algorithm framework which decomposes a multi-objective optimization problem into a number of different single objective optimization subproblems and then uses a population-based method to optimize these subproblems simultaneously.

Genetic Algorithm for Overlapping Community Detection($GaoCD$) [9] addresses the problem of finding overlapping communities by clustering links instead of nodes.

Multi-objective EA for signed social networks [4] adds signs to weights in social networks, and combines direct

GECCO'16 Companion July 20-24, 2016, Denver, CO, USA

ACM ISBN 978-1-4503-4323-7/16/07.
DOI: http://dx.doi.org/10.1145/2908961.2931643

and indirect representations for large-scale networks. This method also detects overlapping communities.

A Genetic Algorithm for Fuzzy Community Detection(*GAFCD*) [10] was recently purposed to deal with fuzzy communities, and outperformed several other algorithms, in terms of finding max-modularity communities.

3. OTHER STATE-OF-THE-ART METHODS

We now present an overview of 5 *recent* state-of-the-art (all published since 2010) methods which are not based on evolutionary computation.

Speaker-listener Label Propagation Algorithm (*SLPA*) [12] identifies both overlapping and non-overlapping communities by propagating labels representing community membership between nodes in a graph. This label propagation process is repeated for a user-defined maximum number of iterations T.

Top Graph Clusters (*TopGC*) [6] is a probabilistic clustering algorithm that finds the top well-connected clusters in a graph. The main idea is to find sets of nodes whose neighborhoods are highly overlapping, as these nodes should be clustered together.

SVINET [2] is based on a Bayesian model for graphs with overlapping communities that uses a mixed-membership stochastic blockmodel. This method can also detect overlapping communities.

DSE Matrix Blocking via Cosine Similarity [1] is inspired by matrix blocking–the process of reordering the rows and columns such that a critical mass of the nonzero elements are along or near the diagonal.

Speed and Performance In Clustering (*SPICi*) [3] is a greedy heuristic algorithm that produces an incomplete clustering and is designed to work on large biological networks.

4. COMPARISON AND CONCLUSIONS

As it is beyond the scope of this work to analyze the performance of each of the above methods individually, we assess the *overall* performance of these two approaches using several measures, which are summarized in Table 1. For brevity, we write *EC* and *non-EC* to denote both method categories, respectively.

(1) Maximal network size tackled–nonEC was applied to far larger networks. However, it is unclear if EC could tackle such networks or not; (2) Community type variants–all types we found were covered by both methods; (3) Goodness of communities found–we could not use this measure since the its values vary too greatly within each method; (4) Running Times - fastest algorithms are *GAFCD* (EC) and *SPICi* (non-EC), the latter being faster; (5) Computing Requirements (HW) are the same; (6) Standard measures - both were testes with standard benchmarks.

To conclude, Evolutionary Computation-based methods have been successfully applied to a wide-range of problems. However, standard methods have not only been successfully applied to all such problems, but also to larger networks than those reported for evolutionary computation. The reason could be that time complexity requirements are better for standard methods. However, we hypothesize that standard methods are practiced by a much larger corpus of researchers, which have found better ways to apply them. Therefore, albeit the wide-range of evolutionary-computation based methods applied to this domain, more research is needed to push the boundary even further.

Table 1: Comparison Summary–Evolutionary-Computation (EC) Based Methods compared to non-Evolutionary Methods (non-EC) applied to The Community Detection Problem. Note: $n = |V|$

Aspect	**EC**	**non-EC**
`Maximal Network Nodes`	$93,000$	$297,000$
`Maximal Network Edges`	$344,000$	7.7 million
`Standard Measures Applied`	Yes	Yes
`Overlapping`	Yes	Yes
`Directed`	Yes	Yes
`Weighted`	Yes	Yes
`Signed`	Yes	Yes
`Community Complexity`	Complex	Complex
`Best Running Times`	$O(n^2)$	$O(n \cdot log(n) + n)$
`Hardware Requirements`	Standard	Standard
Applicable to Real World?	**Yes**	**Yes**
Applied to Real World?	**Less**	**Yes**

5. REFERENCES

[1] J. Chen and Y. Saad. Dense subgraph extraction with application to community detection. *IEEE Trans. Knowl. Data Eng.*, 24(7):1216–1230, 2012.

[2] P. Gopalan and D. M. Blei. Modeling overlapping communities with node popularities. In *Advances in Neural Information Processing Systems 26, 2013, Nevada, United States.*, pages 2850–2858, 2013.

[3] P. Jiang and M. Singh. Spici: a fast clustering algorithm for large biological networks. *Bioinformatics*, 26(8):1105–1111, 2010.

[4] C. Liu, J. Liu, and Z. Jiang. A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks. *IEEE Trans. Cybernetics*, 44(12):2274–2287, 2014.

[5] J. Liu, W. Zhong, H. A. Abbass, and D. G. Green. Separated and overlapping community detection in complex networks using multiobjective evolutionary algorithms. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2010, Barcelona, Spain*, pages 1–7, 2010.

[6] K. Macropol and A. K. Singh. Scalable discovery of best clusters on large graphs. *PVLDB*, 3(1):693–702, 2010.

[7] C. Pizzuti. A multi-objective genetic algorithm for community detection in networks. In *ICTAI 2009, 21st IEEE International Conference on Tools with A.I., New Jersey, USA*, pages 379–386, 2009.

[8] B. Saha, A. Mandal, S. B. Tripathy, and D. Mukherjee. Complex networks, communities and clustering: A survey. *CoRR*, abs/1503.06277, 2015.

[9] C. Shi, Y. Cai, D. Fu, Y. Dong, and B. Wu. A link clustering based overlapping community detection algorithm. *Data Knowl. Eng.*, 87:394–404, 2013.

[10] J. Su and T. C. Havens. Fuzzy community detection in social networks using a genetic algortihm. In *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, pages 2039–2046, July 2014.

[11] P. Wadhwa and M. Bhatia. Community detection approaches in real world networks: A survey and classification. *IJVCSN*, 6(1):35–51, 2014.

[12] J. Xie and B. K. Szymanski. Towards linear time overlapping community detection in social networks. In *Advances in Knowledge Discovery and Data Mining 2012, Kuala Lumpur, Malaysia*, pages 25–36, 2012.