A Multimodal Optimization and Surprise Based Consensus Community Detection Algorithm

Guanbo Jia
School of Computer Science
University of Birmingham
Birmingham, B15 2TT, UK
gxj288@cs.bham.ac.uk

Shan He*
School of Computer Science
University of Birmingham
Birmingham, B15 2TT, UK
s.he@cs.bham.ac.uk

Zexuan Zhu
College of Computer Science
and Software Engineering
Shenzhen University
Shenzhen, 518060, China
zhuzx@szu.edu.cn

Jing Liu
Institute of Intelligent
Information Processing
Xidian University
Xi'an, 710071, China
neouma@mail.xidian.edu.cn

Ke Tang
School of Computer Science
and Technology
University of Science and
Technology of China
Hefei, 230027, China
ketang@ustc.edu.cn

ABSTRACT

A new community structure measure called Surprise has been proposed to address the resolution limit problem of modularity. However, our analysis shows that, similar to modularity, Surprise also suffers from the so-called extreme degeneracy problem, which leads to unstable module identification results. To solve this problem, we propose a novel **M**ultimodal **O**ptimization and **S**urprise based **C**onsensus **Co**mmunity **D**etection (**MOSCCoD**) algorithm. Experimental results show that MOSCCoD has overcome the extreme degeneracy problem of Surprise and shown a very competitive performance in terms of stability and accuracy.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

Keywords

Complex network; community detection; extreme degeneracy; consensus clustering; multimodal optimization

1. INTRODUCTION

Complex systems in real world can be abstracted and represented mathematically by complex networks consisting of vertices and edges which represent individuals and their relationships in systems. One important property of complex networks is called community structure [8]. The identification of community structures is called community detection. To quantify community structures, Newman and Girvan proposed the modularity [8] and many modularity maximization algorithms have been proposed [4]. However, the modularity suffers at least two serious problems, i.e., the resolution limit problem [3] and the extreme degeneracy problem [4], which deteriorates the performance of modularity maximization algorithms on real-world complex networks.

Recently, another measure called Surprise [2] has been proposed, which overcomes the resolution limit problem as shown in [2], and several Surprise maximization algorithms have been developed [2] . However, using the karate [2] and dolphin [8] networks in Figures 1, we show that, similar to modularity, Surprise also suffers from the extreme degeneracy problem, which makes Surprise maximization algorithms finding the optimal Surprise partition very difficult among an exponential number of structurally dissimilar suboptimal partitions whose Surprise values are very close to that of the optimal partition.

2. THE PROPOSED ALGORITHM

To overcome the extreme degeneracy problem of Surprise, we propose a novel algorithm called MOSCCoD. It combines Surprise and consensus clustering [5], but is reformulated as a multimodal optimisation problem [6] to detect representative community structures and employs the differential evolution (DE) [9] with deterministic crowding technique [6] as the multimodal optimisation algorithm. Its framework is:

Step 1) Set generation number $g = 0$ and maximum g_m.

Step 2) Generate the initial population P_g randomly.

Step 3) Evaluate individuals in P_g based on the Surprise.

Step 4) Evolve P_g by DE and deterministic crowding.

Step 5) If $g < g_m$, $g = g + 1$ and go back to **Step 3**).

Step 6) Perform the consensus clustering scheme on P_g to construct the consensus matrix D.

Step 7) If partitions in P_g are not all equal, then reconstruct the network based on D and go back to **Step 1**); otherwise, stop and output the optimal partition.

*Dr. Shan He is the corresponding author.

GECCO'15 Companion, July 11–15, 2015, Madrid, Spain.

ACM ISBN 978-1-4503-3488-4/15/07.

DOI: http://dx.doi.org/10.1145/2739482.2764656.

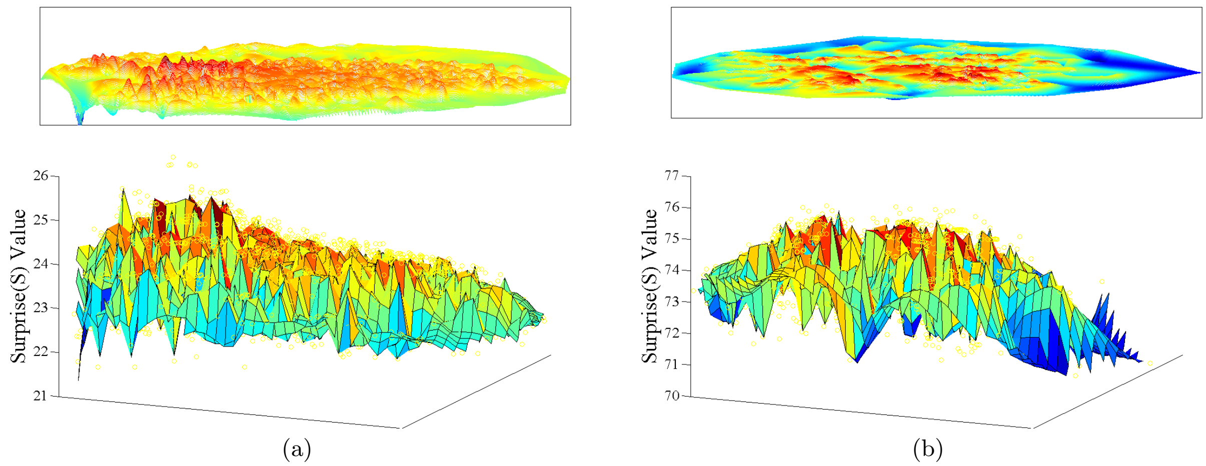

Figure 1: The Surprise functions of the (a) karate and (b) dolphin networks. The x- and y-axes are the embedding dimensions which are complicated functions of the original partition space of networks, and the vertical axis indicates the Surprise (S) values.

Table 1: Results of algorithms on tested networks.

Network	Algorithm	NMI_{bst}	$NMI_{avg} \pm NMI_{std}$
Karate	MOSCCoD	**0.618**	$\mathbf{0.587 \pm 0.007}$
	UVCluster	0.471	$0.427 \pm 0.023^*$
	RCluster	0.517	$0.457 \pm 0.029^*$
	SCluster	0.525	$0.485 \pm 0.041^*$
Dolphin	MOSCCoD	**0.608**	$\mathbf{0.585 \pm 0.011}$
	UVCluster	0.577	$0.546 \pm 0.012^*$
	RCluster	0.586	$0.560 \pm 0.020^*$
	SCluster	0.585	$0.562 \pm 0.022^*$
Football	MOSCCoD	**0.920**	$\mathbf{0.897 \pm 0.012}$
	UVCluster	0.917	$0.886 \pm 0.019^*$
	RCluster	0.911	$0.884 \pm 0.018^*$
	SCluster	0.911	$0.885 \pm 0.019^*$
Books	MOSCCoD	**0.531**	$\mathbf{0.523 \pm 0.016}$
	UVCluster	0.515	$0.417 \pm 0.047^*$
	RCluster	0.522	$0.439 \pm 0.057^*$
	SCluster	0.521	$0.423 \pm 0.047^*$

3. EXPERIMENTAL RESULTS

We ran MOSCCoD 20 times with parameters: population size of 100; maximum generation of 200; DE employs "rand/1" mutation and binomial crossover strategies with scaling factor of 0.9 and control crossover parameter of 0.3. We selected four widely-used real-world benchmark networks (i.e., Karate [2], Dolphin [8], Football [2] and Books [7] networks) all of which have known community structures. We employed UVCluster, RCluster and SCluster algorithms from [1] for comparison. Results are collected in Table 1 using the best, average and standard deviation of Normalized Mutual Information (NMI) [5] with the t-test at a 0.05 significance level. Results with asterisks are significantly worse than that of MOSCCoD. From Table 1, we can see that MOSCCoD always obtained the best NMI_{bst} and NMI_{avg} on these networks. Thus, it can be concluded that MOSCCoD has overcome the extreme degeneracy problem of Surprise and can effectively identify community structures in complex networks.

4. CONCLUSION

To overcome this extreme degeneracy problem of Surprise, we have proposed a novel community detection algorithm called MOSCCoD. To the best of our knowledge, this is the first time the consensus clustering is reformulated as a multimodal optimization problem. The experimental results have shown that MOSCCoD has overcome the extreme degeneracy problem of Surprise and demonstrated a very excellent performance compared with other state-of-the-art community detection algorithms.

5. ACKNOWLEDGMENTS

This work was supported by the Shenzhen Basic Research Program (Grant No. JCYJ20130401170306880).

6. REFERENCES

[1] R. Aldecoa and I. Marín. Jerarca: efficient analysis of complex networks using hierarchical clustering. *PLoS ONE*, 5(7):e1158, 2010.

[2] R. Aldecoa and I. Marín. Deciphering network community structure by surprise. *PLoS ONE*, 6, 2011.

[3] S. Fortunato and M. Barthélemy. Resolution limit in community detection. *Proc. Natl. Acad. Sci.*, 2007.

[4] B. H. Good, Y.-A. Montjoye, and A. Clauset. Performance of modularity maximization in practical contexts. *Phys. Rev. E*, 81(4):046106, 2010.

[5] A. Lancichinetti and S. Fortunato. Consensus clustering in complex networks. *Sci. Rep.*, 2, 2012.

[6] S. W. Mahfoud. *Niching methods for genetic algorithms*. PhD thesis, University of Illinois at Urbana-Champaign, 1995.

[7] M. E. J. Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci.*, 2006.

[8] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(4):026113, 2004.

[9] R. Storn and K. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *J. Glo. Opt.*, 11(4):341–359, 1997.