Identifying Relationship Patterns Inside Communities

Rafael de Santiago[1,2] and Luís C. Lamb[2]

[1] Universidade do Vale do Itajaí, Itajaí SC, Brazil
[2] Institute of Informatics
Universidade Federal do Rio Grande do Sul,
Porto Alegre RS, Brazil

Some studies on networks require to isolate groups of elements, known as Communities. Some examples include (*i*) detection of protein complexes [10]; (*ii*) mapping metabolic networks [9]; (*iii*) identification of criminal organizations [5].

There are many algorithms to find communities and a commonly used evaluation approach is the maximization of the Modularity Clustering. This technique was proposed by Newman and Girvan [7][11] and is based on the number of edges that are into a community minus the expected number of edges from *null model*. The *null model* is a graph $G' = (V, E')$, obtained from original network $G = (V, E)$, with $|E'| = |E|$ and each edge is distributed over uniform probability, so each node in G' has the same degree on equivalent node in G. The objective function is to maximize the Equation $\frac{1}{2}\sum_{u,v \in E}(\frac{A_{u,v}-(d_u d_v)}{2m})\sigma(u, v)$, where the d_w is degree of $w \in V$, $|E| = m$ and function σ returns 1, if u and v are in the same community and 0 otherwise [4] [3]. Some criticism and reviem over Newman and Girvan model can be seen in [6], [2] and [8].

Community detection is an important problem for Computer and other sciences. Following Agarwal and Kempe [1] one of the most important reasons to make clustering over a network is to identify the function/role of each element in a community.

If the communities have hundreds or thousands of elements, it is important to understand the functions of internal elements, but that will require an automatic process. In this context, we propose to develop a model, capable to identify elements with common features in different communities, based on the connection between elements and communities, agreeing with Newman and Girvan model features.

To understand the patterns within communities and carry out the research, the steps below are necessary. The schedule is as follows:

1. Developing a survey about network clustering and algorithms to solve the problem (aug. 2013 – jun. 2014);
2. Understanding the possible dynamics to elements function classification (jan. 2014 – dec. 2014);
3. Developing a model for dynamics function classification (jan. 2015 – dec. 2015);

4. To propose algorithms and to identify the model computational complexity (mar. 2015 – apr. 2016);
5. Verifying the model with networks instances and understand the adherence (mar. 2015 – jun. 2016).

References

1. Agarwal, G., Kempe, D.: Modularity-maximizing graph communities via mathematical programming. The European Physical Journal B 66(3), 409–418 (Nov 2008)
2. Cafieri, S., Hansen, P., Liberti, L.: Loops and multiple edges in modularity maximization of networks. Physical review. E, Statistical, nonlinear, and soft matter physics 81(4 Pt 2), 046102 (Apr 2010)
3. De Meo, P., Ferrara, E., Fiumara, G., Provetti, A.: Mixing local and global information for community detection in large networks. Journal of Computer and System Sciences 80(1), 72–87 (Feb 2014)
4. Dinh, T., Thai, M.: Community Detection in Scale-Free Networks: Approximation Algorithms for Maximizing Modularity. . . . Areas in Communications, IEEE Journal on 31(6), 997–1006 (2013)
5. (in press)Ferrara, E., De Meo, P., Catanese, S., Fiumara, G.: Detecting criminal organizations in mobile phone networks. Expert Systems with Applications (April) (Mar 2014)
6. Fortunato, S., Barthélemy, M.: Resolution limit in community detection. Proceedings of the National Academy of Sciences of the United States of America 104(1), 36–41 (Jan 2007)
7. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99(12), 7821–6 (Jun 2002)
8. Good, B.H., de Montjoye, Y.A., Clauset, A.: Performance of modularity maximization in practical contexts. Physical Review E 81(4), 046106 (Apr 2010)
9. Guimera, R., Amaral, L.: Functional cartography of complex metabolic networks. Nature 433(February), 895–900 (2005)
10. Nepusz, T., Yu, H., Paccanaro, A.: Detecting overlapping protein complexes in protein-protein interaction networks. Nature Methods 9 pp. 471–472 (2012)
11. Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69(2), 026113 (Feb 2004)