CT 1.6.4

Fast unfolding of community hierarchies in large networks

Vincent D. Blondel[1], Jean-Loup Guillaume[2], Renaud Lambiotte[1,3] and Etienne Lefebvre[1]

[1]Department of Mathematical Engineering, Université catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium.
[2]LIP6, Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris, France
[3]Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2AZ, U.K.

Introduction

The typical size of large networks such as social network services, mobile phone networks or the web now counts in millions when not billions of nodes and these scales demand new methods to retrieve comprehensive information from their structure. A promising approach consists in decomposing the networks into communities of strongly connected nodes, with the nodes belonging to different communities only sparsely connected. Finding exact optimal partitions in networks is known to be computationally intractable, mainly due to the explosion of the number of possible partitions as the number of nodes increases. It is therefore of high interest to propose algorithms to find reasonably "good" solutions of the problem in a reasonably "fast" way. One of the fastest algorithms consists in optimizing the modularity of the partition in a greedy way (Clauset *et al*, 2004), a method that, even improved, does not allow to analyze more than a few millions nodes (Wakita *et al*, 2007).

Results

In this paper, we introduce an algorithm (Blondel *et al*, 2008) that finds high modularity partitions of large networks in short time and that unfolds a complete hierarchical community structure for the network, thereby giving access to different resolutions of community detection. Contrary to most of the other community detection algorithms, the network size limits that we are facing with our algorithm are due to limited storage capacity rather than limited computation time: identifying communities in a 118 million nodes network took only 152 minutes on a standard computer. The algorithm is successfully applied to several networks and, in particular, to a large Belgian mobile phone network. This dataset is especially interesting because of the special situation of Belgium that is composed of two main linguistic communities and has had a very unstable political status in the last eight months due to the impossibility to form a government and to the lack of consensus between its communities. Our algorithm allows to unravel the community structure of this social network and to reveal interesting topological features, such as the strong linguistic segregation in the country.

Discussion

The algorithm that we propose has several advantages. First, its steps are intuitive and easy to implement, and the outcome is unsupervised. Moreover, the algorithm is extremely fast, which allows to analyse systems of unprecedented sizes. Our algorithm is also unaffected by the so-called resolution limit problem of modularity (Fortunato *et al*, 2007), i.e. the fact that modularity optimization may fail to identify

communities smaller than a certain size. Our approach provides instead a hierarchical tree of communities, i.e. a decomposition of the network into communities for different levels of organization. This flexibility allows the end-user to zoom in the network and to observe its structure with the desired resolution.

Acknowledgement
This research was supported by the Communauté Française de Belgique through a grant ARC and by the Belgian Network DYSCO, funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. J.-L. G. is also supported by the project MAPE (ANR France) and MAPAP (Safer Internet Plus Programme, European Union).

References

V.D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre (2008): Fast unfolding of community hierarchies in large networks, arXiv:0803.0476

A. Clauset, M.E.J. Newman and C. Moore (2004): Finding community structure in very large networks. Phys. Rev. E 70, 066111

S. Fortunato and M. Barthélemy (2007): Resolution limit in community detection. Proc. Natl. Acad. Sci. USA 104(1), 36-41

K. Wakita and T. Tsurumi (2007): Finding community structure in a mega-scale social networking service. in proceedings of IADIS international conference on WWW/Internet 2007, 153-162