Unfolding of nodes with the same behavior in large networks

Maguy Trefois Jean-Charles Delvenne

Department of Applied Mathematics
Université catholique de Louvain (UCL)
Bâtiment Euler-4, Avenue G. Lemaître B-1348 Louvain-La-Neuve
Belgium
Email: maguy.trefois@uclouvain.be

1 Introduction

A graph, also called network, is a mathematical structure which allows to model complex systems with interacting agents like the Internet or the human societies. Such structures can be partitioned into subgraphs, called communities, with few links between them. This problem of graph partitioning is called community detection and is fundamental for the understanding of the underlying system. Several methods have been developed to detect communities in large graphs. We present an adaptation of these methods in order to detect groups of nodes with the same behavior, that is we partition the graph so that all the nodes of a group communicate in the same way with the nodes of the other groups.

2 Community detection

Several objective functions have been defined in order to partition the graph into communities. We briefly present two of them : the cut and the modularity.

2.1 The cut

The cut function is defined in order to divide the graph into two communities C_1 and C_2. This function counts the number of edges between the two groups of nodes, that is :

$$R = \sum_{i \in C_1, j \in C_2} A_{ij},$$

where A_{ij} denotes the $(i,j)-$entry of the adjacency matrix of the graph to partition.

A relevant partition is a partition minimizing R (see [1] for efficient algorithms).

2.2 The modularity

The modularity function compares the fraction of edges between two nodes in a same community with the expected fraction of edges between these two nodes if the edges were placed randomly in the graph (respecting the degrees of the nodes), that is :

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j),$$

where m is the number of edges in the graph, k_i is the degree of node i, δ is the Kronecker symbol and c_i is the community index to which node i belongs.

A relevant partition is a partition maximizing Q (see [1] and [2] for efficient algorithms).

3 Detecting groups of nodes with the same behavior

We will say that some nodes have the same behavior (or are of the same type) if they communicate in the same way with the nodes of the other groups. Illustrate this with the following examples.

- A movie-actor network : in this graph, a node is either a movie or an actor. If an actor plays in a movie, then their corresponding nodes are connected. So, the movies are only connected to actors and vice versa. In this network, there are two kinds of nodes : the movies and the actors.
- A food web network : in this directed graph, a node represents an individual. If there is an edge from node i to node j, that means that individual i eats individual j. In this graph, all individuals of species A only eat individuals of species B, and so on. So, a group of nodes with the same behavior matches with the individuals belonging to a same species.

The objective functions described in previous section have been generalized to detect such groups of nodes. We will study the flow of information in the system whose communication topology is described by the graph in order to detect nodes with the same behavior.

References

[1] M.E.J. Newman. *Networks : An introduction*. Oxford University Press, Oxford UK, 2010.

[2] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefevre. *Fast unfolding of communities in large networks.* Journal of Statistical Mechanics: Theory and Experiment, page P10008, 2008.