Networks in Economics and Social Science
Homework 2019/2020

Roberto Casarin

Abstract

Prepare a report containing the definition and properties of network modularity and a modularity analysis of the interlocking network data provided. Your report should provide a reply to the questions given in this document. You can use the LaTeXsource file of this document to generate your report. Your Homework is due by 30th of March by e-mail (`r.casarin@unive.it`).

1 Reading and comprehension

Let $C = (C_1, \ldots, C_R)$ be a partition of the vertex set V such that $C_i \cap C_j = \emptyset$ for $i \neq j$ and $C_1 \cup \ldots \cup C_R = V$. The modularity of the partition C is a scalar value between -1 and 1. The modularity measures the density of links inside communities as compared to links between communities.

Following Blondel et al. (2008) the modularity of a partition $C = (C_1, \ldots, C_R)$ of a graph with n nodes and adjacency matrix A is defined as follows

$$Q(C) = \frac{1}{2d} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(A_{ij} - \frac{d_i d_j}{2d} \right) \delta(c_i, c_j) \tag{1}$$

where

$$d = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}, \quad d_i = \sum_{j=1}^{n} A_{ij}$$

are the i-th node and the graph total degree, respectively, c_i is the partition element to which the node i belongs, and $\delta(x, y)$ is a Dirac function which takes value 1 if $x = y$ and value 0 otherwise.

1. Provide a summary of the heuristic rule used by Blondel et al. (2008) to find the optimal value of $Q(C)$.

2. Show that the modularity

$$Q(C) \;=\; \frac{1}{2d}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(A_{ij}-\frac{d_i d_j}{2d}\right)\delta(c_i,c_j)$$

is upper bounded by 1 and lower bounded by -1.

3. Let $\tilde{C}$ be a partition obtained from C by setting $\tilde{C}_r = C_r$ for $r = 3,\ldots,R$ and by moving the node h from the first to the second community, that is: $\tilde{C}_1 = C_1 - \{h\}$ and $\tilde{C}_2 = C_2 \cup \{h\}$. Show that

$$Q(\tilde{C}) - Q(C) \;=\; = \frac{\kappa_{2h}}{d} - \frac{\kappa_{1h}}{d} + \frac{d_h\kappa_1}{2d^2} - \frac{d_h\kappa_2}{2d^2}$$

where

$$\kappa_{rh} = \sum_{i\neq h} A_{ih}\delta(c_i,c_h) = \sum_{i\neq h} A_{ih}\delta(c_i,r) = \sum_{i\in C_r, i\neq h} A_{ih}$$

is the number of nodes of the community C_r which are connected to h and

$$\kappa_r = \sum_{i\neq h} d_i\delta(c_i,c_h) = \sum_{i\in C_r, i\neq h} d_i$$

is the sum of the degree of the nodes in the community C_r excluding node h, for a given partition C such that the node h is in the r-the element, i.e. $c_h = C_r$.

4. Let C be a partition such that the first community in the sequence contains and isolated node, that is $C_1 = \{h\}$. Let $\tilde{C}$ be a partition obtained from C by setting $\tilde{C}_r = C_r$ for $r = 3,\ldots,R$ and by moving the node h from the first to the second community, that is: $\tilde{C}_1 = \emptyset$ and $\tilde{C}_2 = C_2 \cup \{h\}$. Show that

$$Q(C) - Q(\tilde{C}) = \frac{d_h\kappa_2}{2d^2} - \frac{\kappa_{2h}}{d}$$

2 Interlocking directorate

2.1 Dataset description

The nodes of the network are companies. If the same individual is in the boards of two companies then there is an edge between those two companies. This set of nodes and edges defines an *interlocking directorate network*. The two files `VIPDedgelist.csv` and `VIPDnodelist.csv` contain the edge and node list of the interlocking directorate network for two Italian provinces (Vicenza and Padova).

The vertex set, which comprises 33340 nodes, decomposes in two subsets following the province (see column partition in the node list file).

Multiple edge between nodes are allowed. There are three types of edges: Female, Male and Firm, which correspond to three different type of board members (directors) in the interlocking directorate.

The edge set, which includes 16896 edges, decomposes in three subsets following the type of edge (see column partition in the edge list file, where 1 means Male, 2 Female and 3 Firm)

In this homework you will work on the subgraph described in the following section.

2.2 Subgraph extraction

Import the *interlocking directorate network*, delete vertices with degree less or equal to 8 and export the results to a new workspace. In the new workspace apply Force Atlas 2 layout with parameters: scaling 1.5, gravity 10.0, dissuade hubs true, linlog mode true and prevent overlaps true. The size of the node has to be proportional to their degree (size between 1 and 16).

Exhibit the result (export the graph from the preview window to a `.pdf` format). Your network should be similar to the one in Figure 1. In your figure use **coloured nodes** following the node-partition

2.3 Connected components

Find:

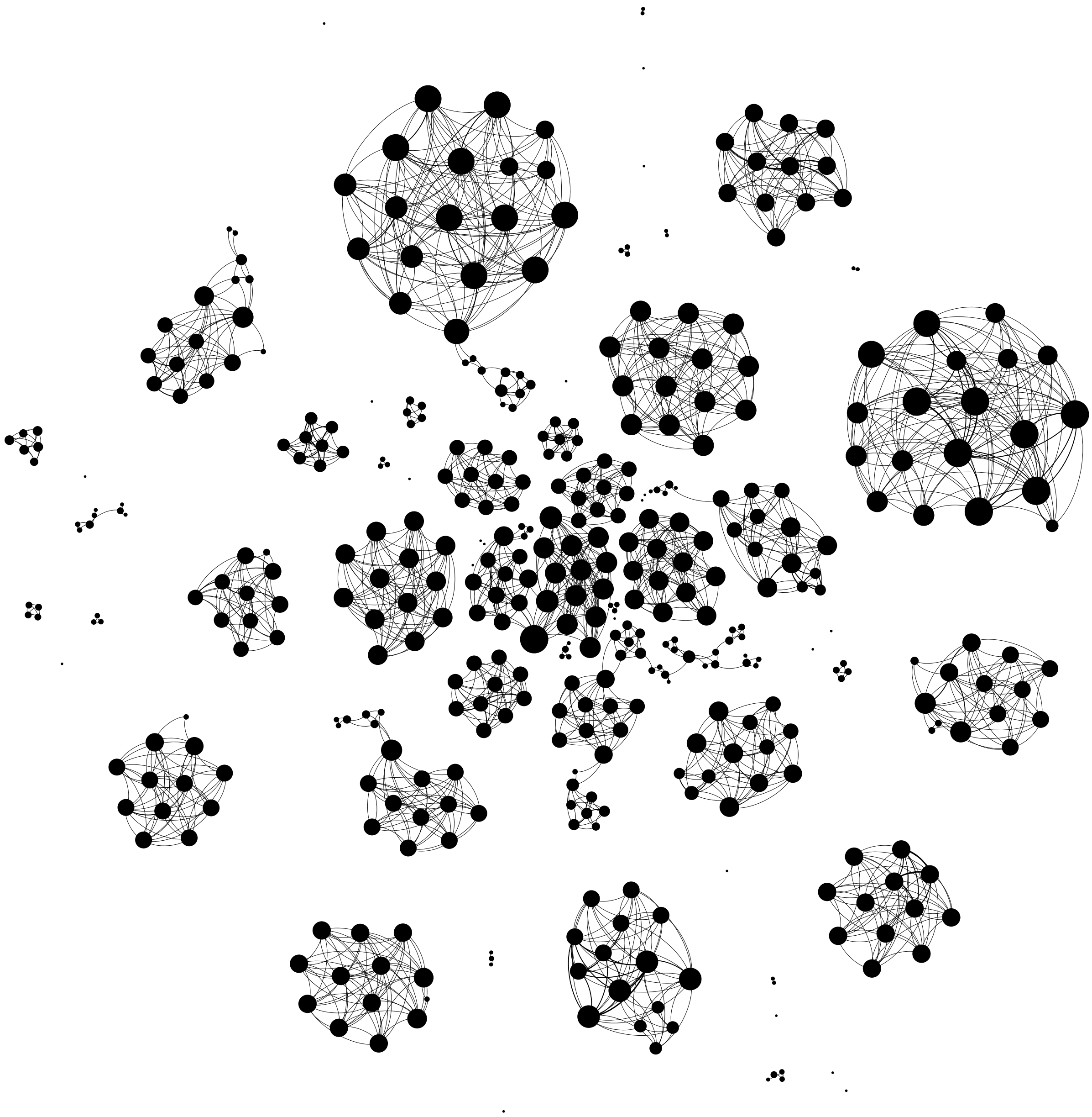

Figure 1: Network

- the number of nodes;
- the number of edges;
- the number of connected components;

and color in red the nodes and edges of the largest connected component (i.e., giant component) and the second largest connected component.

2.4 Communities

Apply Blondel et al. (2008) to find the communities in the network, with three alternative settings for the resolution parameter: 0.5, 1 and 2 (do not use edge weights).

For each resolution setting:

- report the number of communities;
- report the number of communities belonging to the giant component;
- use as node size the betweenness centrality and color the nodes and edges of the communities using a separate color for each community. Exhibit the resulting graph.

Why the number of communities and the number of connected components can be different?

3 References

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, (10), P1000.