Role Detection: Network Partitioning and Optimal Model of the Lumped Markov Chain

Maguy Trefois and Jean-Charles Delvenne

Univ. catholique de Louvain (UCL) Dept. of Applied Mathematics, Louvain-la-Neuve 1348, Belgium

IAP VI/4 - DYSCO

Motivation

Partitioning the network in such a way that the dynamics on the lumped network is a good model of the lumped Markov chain.

In particular, our method allows to detect:
- if the network is multipartite
- the communities in a well clustered network
- a non trivial partition with respect to which the Markov chain (represented by the network) is lumpable, if any.

In general, our method detects the most relevant partition which best respects the lumpability property on the Markov chain (represented by the network).

Definition of Roles

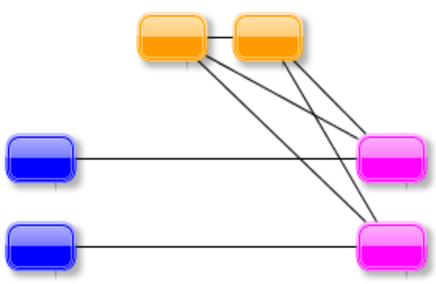
Idea: we look for a partition of the network and for a Markov chain on the blocks that is a good model of the lumped Markov chain.

This idea can be translated in the following way:

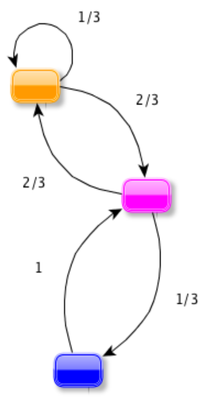
Definition 1: A node partition $S = \{S_1, ..., S_n\}$ of a network is a perfect role partition if for all roles S_k, S_l, any two nodes of role S_k have the same proportion of their neighbors in role S_l.

Example

Network



Lumped network



The nodes of the lumped network correspond to the different roles and the weight of a directed edge in the lumped network is the probability of jumping from the source node to the target node.

Objective Function (1)

In general, there does not exist a non trivial perfect role partition. That is why we are interested in the most relevant partition whose Markov chain defined on the blocks is a good model of the lumped Markov chain.

This partition is obtained by minimizing the following function defined on the partitions $S = \{S_1, ..., S_n\}$ of the network:

$$f(S) = \frac{1}{n^2} \sum_{k,l=1}^{n} e_{kl} - E(e_{kl}).$$

Objective Function (2)

For any two blocks S_k, S_l, we compute the « variance » e_{kl} of the proportion of neighbors of a node of S_k belonging to S_l. Then, we compare this « variance » with its expected value $E(e_{kl})$ in a null model (here, the Erdos-Rényi model) and finally, we compute the mean of these differences on all pairs of blocks.

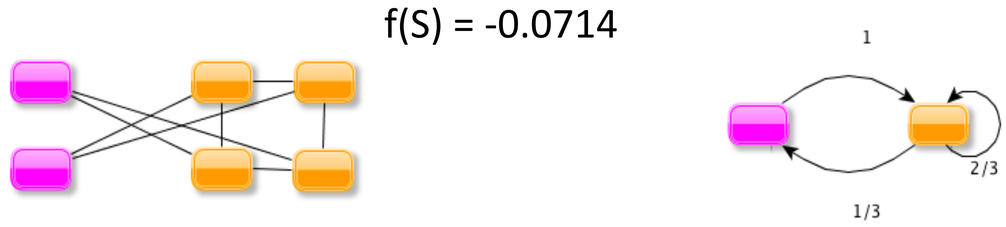
The weight of the edge from S_k to S_l in the lumped network is given by the arithmetic mean of the proportions of neighbors of the nodes of S_k belonging to S_l.

Results

We tested the objective function on toy networks. Here are some important results:

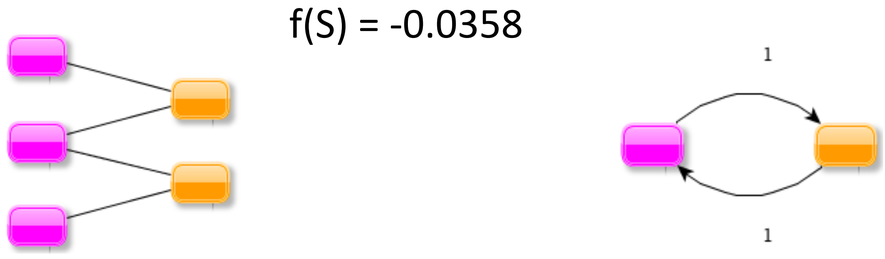
- A lumpable Markov chain

Network — f(S) = -0.0714 — Lumped network



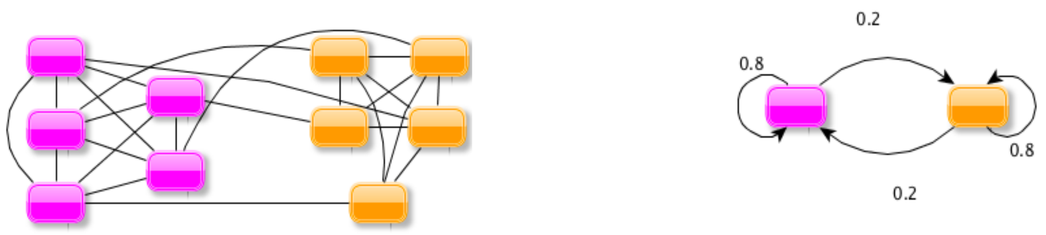
- A bipartite graph

Network — f(S) = -0.0358 — Lumped network



- A well clustered network

Network — f(S) = -0.0223 — Lumped network



Future Work

In the future, we will test the efficiency of the presented method through real networks. For examples,
- in food webs, we hope to detect the different species
- in Zachary's network, we hope to detect the two clans
- …

References

[1] T. Cason, **Node-to-node Similarity Measures and Role Extraction in Networks**, PhD thesis, Université catholique de Louvain, Belgium, 2012.
[2] K. Cooper, M. Barahona, **Role-based similarity in directed networks**, e-print arXiv: 1012.2726, 2010.
[3] W. E., T. Li, E. Vanden-Eijnden, **Optimal partition and effective dynamics of complex networks**, PNAS, vol. 105, p.7907-7912, Jun. 2008.
[4] J. Reichardt, D.R. White, **Role models for complex networks**, Eur. Phys. J. B 60, 217-224 (2007.)