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

Maguy Trefois[1] and Jean-Charles Delvenne[123]

[1] Department of Applied Mathematics, Université catholique de Louvain, Belgium
[2] Namur Complex Systems Center (NAXYS), Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium
[3] Center for Operations Research and Econometrics (CORE), Université catholique de Louvain, Belgium

## 1 Introduction

Nowadays, complex networks are present in many fields (social science, information theory, chemistry, biology, computer science, ...) as they allow to model systems with interacting agents. In many cases, the number of interacting agents is large (from hundreds to millions of nodes). In order to get information about the functionality of the underlying system, we are interested in studying the structure of the network. One way to do that is by partitioning the network into communities (many links within the clusters and few links between them). In the last decade, this community detection problem has attracted many interest in research [1–3, 6, 8–10, 12, 13].

In this paper, we present a method to detect a partition of the network such that the dynamics of a random walker on the lumped network is a good model of the dynamics of a random walker in the original network. In particular, our strategy allows to find the communities in a well clustered network, or to discover if the network is multipartite. Moreover, in the case of a lumpable Markov chain, this strategy provides the partition with respect to which the chain is lumpable [7].

## 2 The Partitioning Problem

Consider an undirected and unweighted network. The dynamics defined on the network is the following: being at node $i$, the probability of jumping to node $j$ is

$$p_{ij} = \begin{cases} \frac{A_{ij}}{k_i} & \text{if } k_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $k_i$ is the degree of node $i$ and $A$ is the adjacency matrix of the network. This dynamical process is a Markov chain on the network.

We are interested in partitioning the network so that the dynamics defined on the blocks is a good model of the dynamics in the original network. More precisely, we look for a partition $S = \{S_1, ..., S_n\}$ such that for any blocks $S_k$, $S_l$ and for any nodes $i, j \in S_k$,

$$\sum_{m \in S_l} p_{im} = \sum_{m \in S_l} p_{jm} \quad . \tag{1}$$

Notice this partitioning problem exactly corresponds to the lumpability of the Markov chain defined on the original network.

In general, the Markov chain defined on the network in not lumpable, which means that there does not exist a relevant partition $S$ having exactly property (1). That is why we are interested in the most relevant partition whose dynamics on the blocks is a good model of the dynamics in the original network. The blocks of this partition will be called "roles" (this role definition differs from those proposed in [4, 5, 11]). In next section, we present our strategy to find such a partition.

## 3 The Objective Function

In [7], E *et al* suggest a method in order to partition the network as defined in previous section. However, in their method they have to fix in advance the number of roles to detect. As this number is a priori unknown, this seems to be a big disadvantage of their strategy. That is why we present another strategy in which the relevant number of roles is provided by the method itself.

The role partition will be represented by a lumped network. The nodes of the lumped network correspond to the different roles and the weight of the directed edge from node $n_k$ to node $n_l$ in the lumped network represents the probability of jumping from node $n_k$ to node $n_l$.

Given a role partition $S = \{S_1, ..., S_n\}$ of the original network, the weight $m_{kl}$ of the edge from node $n_k$ to node $n_l$ in the corresponding lumped network is given by the arithmetic mean of the probabilities of jumping from a node of role $S_k$ to any node of role $S_l$, that is

$$m_{kl} = \frac{1}{|S_k|} \sum_{i \in S_k} p(i, S_l) \quad ,$$

where $p(i, S_l) = \sum_{j \in S_l} p_{ij}$ is the probability of jumping from node $i$ to any node of role $S_l$.

We would like to find a partition $S = \{S_1, ..., S_n\}$ such that for any nodes $i$ and $j$ belonging to a same block and for any block $S_l$, the probabilities $p(i, S_l)$ and $p(j, S_l)$ are very similar, that is we would like to find a partition $S$ which minimizes the expression:

$$\sum_{k,l=1}^{n} \sum_{i \in S_k} \frac{(p(i, S_l) - m_{kl})^2}{|S_k|} \quad .$$

However, the partition with only one block and the partition with the maximum number of blocks (that is, any node of the original network is a block) are trivial solutions. So, minimizing previous expression does not provide a relevant partition. To deal with this problem, we compare the observed "variance" $e_{kl} := \sum_{i\in S_k} \frac{(p(i,S_l)-m_{kl})^2}{|S_k|}$ with its expected value $E(e_{kl})$ in a null model (e.g., the Erdos-Rényi model). Then, we compute the mean of these differences on all pairs of blocks.

Consequently, we would like to find a partition minimizing the function:

$$f(S = \{S_1, ..., S_n\}) = \frac{1}{n^2} \sum_{k,l=1}^{n} e_{kl} - E(e_{kl}) \ .$$

Notice that the partition with only one block and the partition with the maximum number of blocks are not trivial minimizers of $f$.

We will show the efficiency of this objective function through several examples.

**Acknowledgments.** We acknowledge support from the Belgian Programme of Interuniversity Attraction Poles and an Action de Recherche Concertée (ARC) of the French Community of Belgium.

## References

1. Aynaud, T., Guillaume, J.L.: Multi-Step Community Detection and Hierarchical Time Segmentation in Evolving Networks. Proceedings of the 5th SNA-KDD Workshop '11
2. Blondel, V., Guillaume, J.-L., Lambiotte, R., Lefèvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, Oct. 2008
3. Browet, A., Absil, P.A., Van Dooren, P.: Community Detection for Hierarchical Image Segmentation. IWCIA'11 Proceedings of the 14th international conference on Combinatorial image analysis, p. 358-371, 2011
4. Cason, T.: Node-to-node Similarity Measures and Role Extraction in Networks. PhD thesis, Université catholique de Louvain, Belgium, 2012
5. Cooper, K., Barahona, M.: Role-based similarity in directed networks. E-print arXiv: 1012.2726, 2010
6. Delvenne, J.C., Yaliraki, S.N., Barahona, M.: Stability of graph communities across time scales. PNAS, vol. 107, p. 12755-12760, Jul. 2010
7. E, W., Li, T., Vanden-Eijnden, E.: Optimal partition and effective dynamics of complex networks. PNAS, vol.105, p.7907-7912, Jun. 2008
8. Fortunato, S.: Community detection in graphs. Physics Reports 486, 75-174 (2010)
9. Lambiotte, R., Delvenne, J.-C., Barahona, M.: Laplacian Dynamics and Multiscale Modular Structure in Networks. ArXiv:0812.1770, Oct. 2009
10. Mucha, P., *et al*: Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science 328, 876(2010)

11. Reichardt, J., White; D.R.: Role models for complex networks. Eur.Phys.J. B 60, 217-224(2007)
12. Schaub, M., Delvenne, J.-C., Yaliraki, S.N., Barahona, M.: Markov Dynamics as a Zooming Lens for Multiscale Community Detection: Non Clique-Like Communities and the Field-of-View Limit. PLoS ONE, vol. 7, e32210, Feb. 2012
13. Traag, V.A., Van Dooren, P., Nesterov, Y.: Narrow scope for resolution-limit-free community detection. Phys. Rev. E 84, 016114 (2011)