Multilevel Analysis of Co-authorship Networks: Evaluating the Impact of Exogenous Factors on the Conduct of Scientific Collaborations

Robin Lamarche-Perrin[1], Lionel Tabourier[1], and Fabien Tarissan[2]

{FirstName.LastName}@lip6.fr

[1]Sorbonne Universités, UPMC, Université Paris 6, CNRS, LIP6

[2]Université Paris-Saclay, ISP, ENS Cachan, CNRS

Session: Multilevel Network Perspectives in and around Organizations: Theory, Structure and Dynamics

Keywords: Co-authorship networks, multilevel graph analysis, from individuals to institutional levels, information-theoretic data aggregation.

Science studies have shown that a large number of exogenous factors influence the conduct of scientific activities, such as the social, institutional, economical, political, and geographical context of academic research. Hence, the *description* and *explanation* of scientific practices cannot go without a clear understanding of the impact of these macroscopic contexts on the individuals. In this communication, we propose a multilevel analysis method to describe co-authorship networks at various scales and thus address these challenging issues.

Our method exploits *information-theoretic data aggregation* to summarise the information that is contained at the individual level in a given data set and to highlight consistent patterns that arises at higher-levels [1, 2, 3]. Applied to relational data, this method consists in partitioning the adjacency matrix of co-authorship networks into "rectangular tiles". Each such tile represents an *aggregate edge* between two *aggregate nodes*, that is the aggregation of several collaboration relations between two groups of researchers. Similarly to *block-modeling methods* – introduced for the analysis of social networks – the quality of such an aggregate depends on the homogeneity of the relations it contains, that is the homogeneity of collaborations between the two groups of researchers. We propose to use information-based measures, such as Kullback-Leibler divergence, to quantify and optimise this homogeneity criterion.

Data aggregation thus provides an abstraction method to identify homogeneous macroscopic patterns within the network. However, one can require additional constraints to apply on the set of feasible abstractions depending on the exogenous factors one is interested in. For example, one could only consider as valid abstractions the groups of researchers that are defined at some institutional level: *e.g.*, individuals aggregated as research teams, as research departments, as institutes, and so on. By introducing such constraints within the aggregation process, the network is then macroscopically described on the basis on this exogenous vocabulary, thus leading to general empirical assertions regarding the collaborations within the different institutional levels, as well as the individual level.

To go further, we are currently interested in the combination of several sets of constraints, modelling several exogenous contexts. We hence propose to add some other exogenous classifications in order to allow more abstractions for the aggregation purposes: *e.g.*, geographical categories (cities, countries, regions), epistemic categories (subfields, fields, disciplines), and so on. Our aggregation method should hence allow to identify the categories, and eventually the combinations of categories, that are the most efficient to macroscopically describe co-authorship networks, thus providing different points of view on scientific activities by mixing the individual level to several multiscale exogenous contexts.

References

[1] R. Lamarche-Perrin. Optimal Partition: A Toolbox to Solve Structured Versions of the Set Partitioning Problem with Decomposable Objectives. `https://github.com/Lamarche-Perrin/optimal_partition/`, 2015.

[2] R. Lamarche-Perrin, Y. Demazeau, and J.-M. Vincent. A Generic Algorithmic Framework to Solve Special Versions of the Set Partitioning Problem. In *Proceedings of the 26th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'14)*, pages 891–897. IEEE Computer Society, 2014.

[3] R. Lamarche-Perrin, Y. Demazeau, and J.-M. Vincent. Building Optimal Macroscopic Representations of Complex Multi-agent Systems. In *Transactions on Computational Collective Intelligence*, volume XV of *LNCS 8670*, pages 1–27. Springer-Verlag, 2014.