

Estimation of static community memberships from multiplex and temporal network data

K. Avrachenkov¹, M. Dreveton¹ and L. Leskelä²

¹ *Inria Sophia-Antipolis, France, {k.avrachenkov,maximilien.dreveton}@inria.fr*

² *Aalto University, Finland, lasse.leskela@aalto.fi*

Data sets in many application domains consist of pairwise interactions observed over time. Pair interactions are often characterized by the type of interacting objects, and a set of objects with a common type is called a community. Community recovery or clustering is the unsupervised task of inferring the community memberships from the observed pair interactions.

As an information-theoretic benchmark, we study data sets generated by a homogeneous block model where the pairwise interactions take values in a measurable space \mathcal{S} . The interactions within a block are distributed according to f_{in} , and interactions between blocks according to f_{out} . We derive a lower bound on the expected number of misclassified nodes made by *any* clustering algorithm. This naturally extends the recent results of [3] to a non-asymptotic setting which makes no regularity assumptions on f_{in} , f_{out} nor on the underlying space \mathcal{S} of interaction types. In particular, we can consider a multiplex (dynamic) network where the number of layers (snapshots) grows with the number of nodes N . Then, we show that this bound is achieved by an *ad-hoc* algorithm.

If we denote by D the Rényi-divergence between f_{in} and f_{out} , then for same-size clusters almost exact recovery (the expected proportion of misclassified nodes going to zero) is possible if $ND \ll 1$, and is impossible otherwise. This provides a natural extension to known results in block models [4]. We later apply those results to dynamic networks where the interaction kernel has a Markov structure, generalizing the results of [2] on a multiplex SBM with independent layers.

In the second part, we propose several clustering algorithms, both offline and online, which fully utilize the temporal nature of the observed data.

For further details and proofs, we refer to our preprint [1].

References

- [1] Avrachenkov, K., Dreveton, M., and Leskelä, L. (2020) Estimation of Static Community Memberships from Temporal Network Data. *arXiv:2008.04790*.
- [2] Paul, S., and Chen, Y. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electronic Journal of Statistics*, **10(2)**, 3807-3870.
- [3] Xu, M., Jog, V., and Loh, P. L. (2020). Optimal rates for community estimation in the weighted stochastic block model. *The Annals of Statistics*, **48(1)**, 183-204.
- [4] Zhang, A. Y., and Zhou, H. H. (2016). Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, **44(5)**, 2252-2280.