Assortative-Constrained Stochastic Block Models

Daniel Gribel [1], Thibaut Vidal [1], Michel Gendreau [2]

2020/12/08

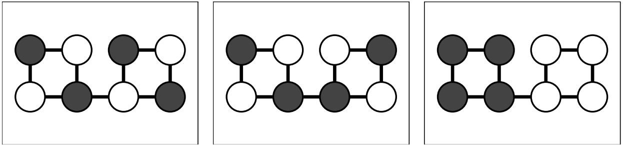
[1] *Pontifícia Universidade Católica do Rio de Janeiro, Brazil*
[2] *CIRRELT and Polytechnique Montréal, Canada*

Introduction

- **Stochastic Block Models (SBM)** [Holland et al., 1983], [Nowicki and Snijders, 2001] are random graph models, in which fitting the parameters to empirical graphs is a prominent way of discovering **communities**.
- Typical network structures:
  - *Assortative* networks: connections within communities are more frequent than in between communities.
  - *Disassortative* networks: connections within communities are less frequent than in between communities.

Introduction

- SBMs are agnostic to assortativity, and can indifferently model assortative and disassortative structures.
  - ⇒ Especially in sparse graphs (or with lightly assortative structures), non-assortative solutions with a better likelihood may substitute the assortative solutions which were originally sought.



(a) Opt. solution $\log \mathcal{L} = -2.7616$
(b) 2^{nd} best solution $\log \mathcal{L} = -5.3193$
(c) 3^{rd} best solution $\log \mathcal{L} = -5.9012$

Figure 1: The three best solutions in an example case.

Key contributions

1. We introduce a SBM variant which incorporates assortativity constraints to represent prior user knowledge;
2. We propose an efficient solution approach based on local optimization and interior-point algorithms for this model;
3. Through extensive computational experiments, we identify the regimes in which it contributes to improve community detection practice.

Degree-corrected SBM (DC-SBM)

- In its most fundamental form, the DC-SBM considers N nodes allocated to K groups.
- The number of edges between a pair of nodes (i, j) depends only on the groups to which the nodes belong and their degrees.

Key concepts and Notations

DC-SBM log-likelihood

$$\log P(A|\Omega, Z) = \frac{1}{2} \sum_{rs}^{K} \sum_{ij}^{N} \left(A_{ij} \ln(\omega_{rs}) - \frac{k_i k_j}{2m} \omega_{rs} \right) z_{ir} z_{js}, \quad (1)$$

where A is the observed adjacency matrix, k_i is the degree of node i, and m is the total number of edges. Variable Z represents the memberships, where $z_{ir} = 1$ indicates that node i is assigned to group r, and Ω is a symmetric $K \times K$ edge probability matrix.

Key concepts and Notations

Strong assortativity

- All diagonal terms of Ω are greater or equal than all off-diagonal terms:

$$\omega_{qq} \geq \omega_{rs} \quad \forall\, q, r, s \in \{1, \ldots, K\}, r \neq s. \tag{2}$$

Assortative-Constrained DC-SBM (AC-DC-SBM)

AC-DC-SBM log-likelihood

$$
\begin{aligned}
\max_{\Omega, Z, \lambda} \quad & \frac{1}{2} \sum_{rs}^{K} \sum_{ij}^{N} \left(A_{ij} \log(\omega_{rs}) - \frac{k_i k_j}{2m} \omega_{rs} \right) z_{ir} z_{js} && \text{(3a)} \\
\text{s.t.} \quad & \omega_{qq} \geq \lambda \quad \forall q \in \{1, \ldots, K\} && \text{(3b)} \\
& \omega_{rs} \leq \lambda \quad \forall r, s \in \{1, \ldots, K\}, r \neq s && \text{(3c)} \\
& \omega_{rs} \geq 0 \quad \forall r, s \in \{1, \ldots, K\}, && \text{(3d)}
\end{aligned}
$$

where λ represents a continuous variable acting as a threshold.

Assortative-Constrained DC-SBM (AC-DC-SBM)

Solution approach

We propose a maximum-likelihood iterative algorithm to solve (3a – 3d), in which we combine two techniques:

(i) an incremental move evaluation approach, using the log-likelihood of the unconstrained problem to filter relocation candidates, and keeping this solution if it satisfies the assortativity constraints;

Assortative-Constrained DC-SBM (AC-DC-SBM)

Solution approach

(ii) an interior point solver for the convex subproblem (4a – 4d), only used if the relocation candidate is not feasible:

$$\max_{\Omega,\lambda} \quad \frac{1}{2}\sum_{rs}^{K} \left(m_{rs}\log(\omega_{rs}) - T_{rs}\omega_{rs}\right) \tag{4a}$$

$$\text{s.t.} \quad \omega_{qq} \geq \lambda \quad \forall q \in \{1,\ldots,K\} \tag{4b}$$

$$\omega_{rs} \leq \lambda \quad \forall r,s \in \{1,\ldots,K\}, r \neq s \tag{4c}$$

$$\omega_{rs} \geq 0 \quad \forall r,s \in \{1,\ldots,K\}, \tag{4d}$$

where m_{rs} represents the number of edges between communities r and s according to the fixed partition and $T_{rs} = \left(\sum_t^K m_{rt} \sum_t^K m_{st}\right)/2m$.

Synthetic Networks

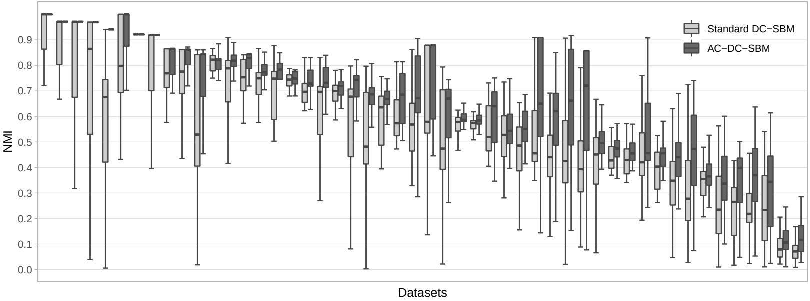


Figure 2: Performance of DC-SBM and AC-DC-SBM on networks generated from general SBMs. The results are ordered by median NMI.

Synthetic Networks

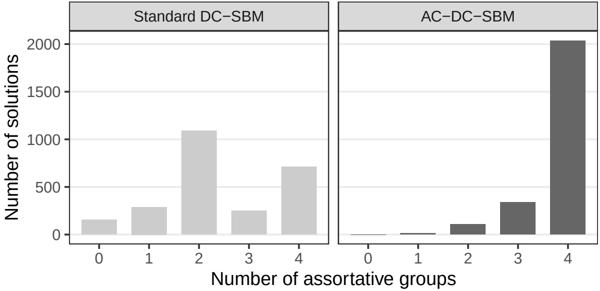
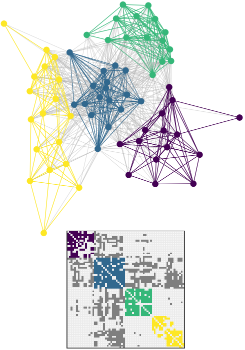
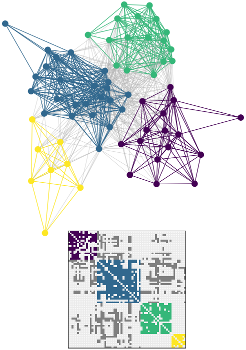


Figure 3: Distribution of the number of assortative communities found by AC-DC-SBM and DC-SBM on networks with $K = 4$ groups.

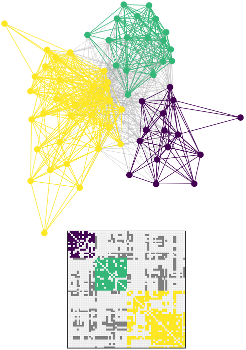
Brain Cortex Networks



(a) Standard DC-SBM



(b) AC-DC-SBM



(c) Modularity maximization model

Figure 4: The best among 100 network partitions found by different models in the cats cortex network.

Conclusions

- In this work, we show that the AC-DC-SBM significantly outperforms unconstrained community detection methods in lightly assortative graphs.
- In these circumstances, the classic SBM has a strong tendency to converge towards non-assortative solutions, while the modularity maximization model does not generalize well to graphs in which the number of edges between groups widely varies.
- For research perspectives, we recommend to investigate different algorithmic paradigms to improve the solution of constrained SBMs, and to pursue the study of the AC-DC-SBM in a wider range of application contexts.

Source code and datasets

- The algorithms of this paper were implemented in Julia (version 1.0.5) and the source code is available at: `http://github.com/danielgribel/AssortativeSBM`

References I

Amini, A. A., Levina, E., et al. (2018).
On semidefinite relaxations for the block model.
The Annals of Statistics, 46(1):149–179.

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011).
Inference and phase transitions in the detection of modules in sparse networks.
Physical Review Letters, 107(6):065701.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983).
Stochastic blockmodels: First steps.
Social Networks, 5(2):109–137.

References II

Nowicki, K. and Snijders, T. A. B. (2001).
Estimation and prediction for stochastic blockstructures.
Journal of the American Statistical Association, 96(455):1077–1087.