Community Detection on Block Models with Geometric Kernels

Konstantin Avrachenkov[1], **B. R. Vinay Kumar**[2], and Lasse Leskelä[3]

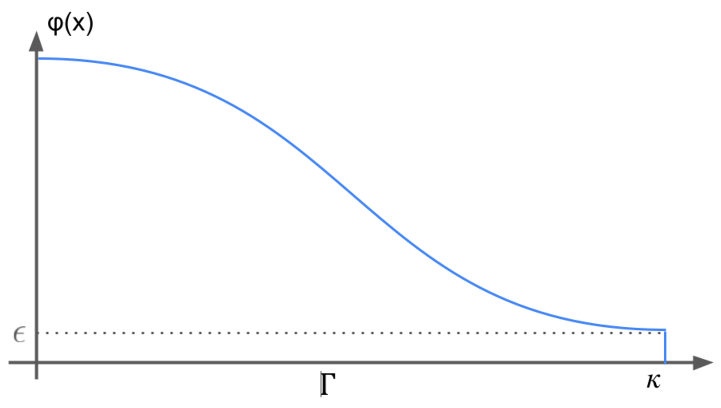
[1]INRIA Sophia Antipolis, France, [2]TU Eindhoven, The Netherlands, [3]Aalto University, Finland

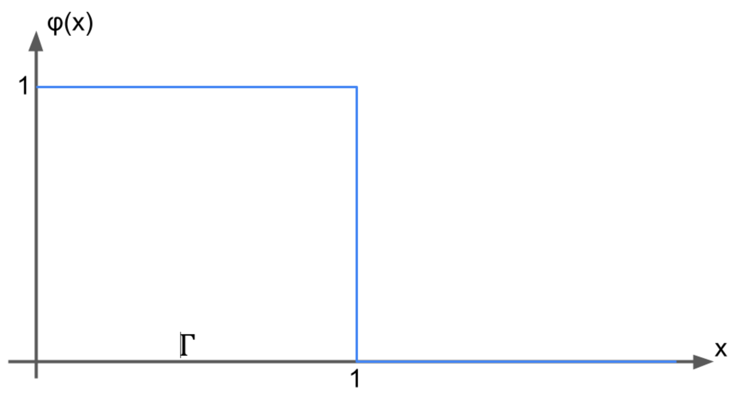
Model

- Let $\mathbf{S} = \left[\frac{-1}{2}, \frac{1}{2}\right]^d$. Given $n \geq 1$ and $\lambda > 0$.
- Sample $N \sim$ Poi(λn)
- Locations: $\mathbf{X} = (X_u)_{u=1}^N$, $X_u \sim$ Unif$(\mathbf{S})$
- Two communities: $\boldsymbol{\sigma} = (\sigma(1), \cdots, \sigma(N))$

$$\mathbb{P}(\sigma(u) = +1) = \mathbb{P}(\sigma(u) = -1) = \frac{1}{2}$$

- Geometric kernel: $\phi \colon \mathbb{R}_+ \to [0,1]$ measurable



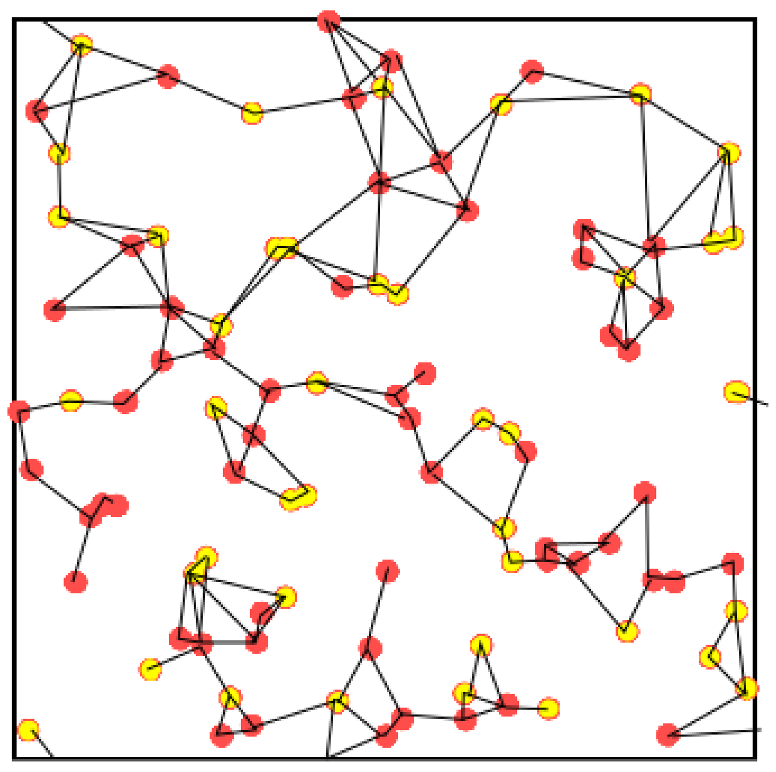


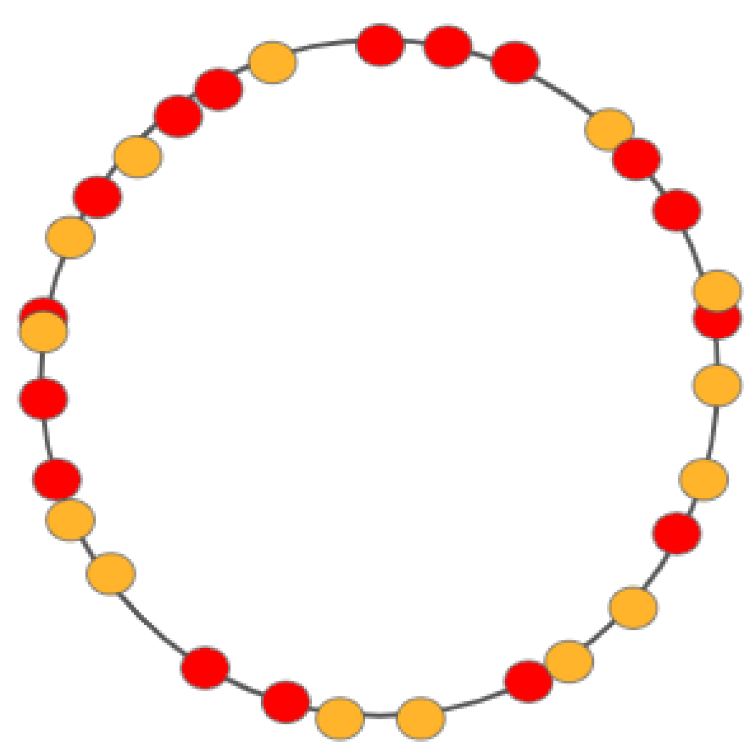
Given locations $\mathbf{X}$ and communities $\boldsymbol{\sigma}$

$$A_{uv} \sim \begin{cases} \text{Ber}(p\psi_n(X_u, X_v)) & \text{if } \sigma(u) = \sigma(v), \\ \text{Ber}(q\psi_n(X_u, X_v)) & \text{if } \sigma(u) \neq \sigma(v), \end{cases}$$

where $\psi_n(x,y) \;=\; \phi\left(\frac{n}{\log n} d(x,y)\right)$.

Illustration

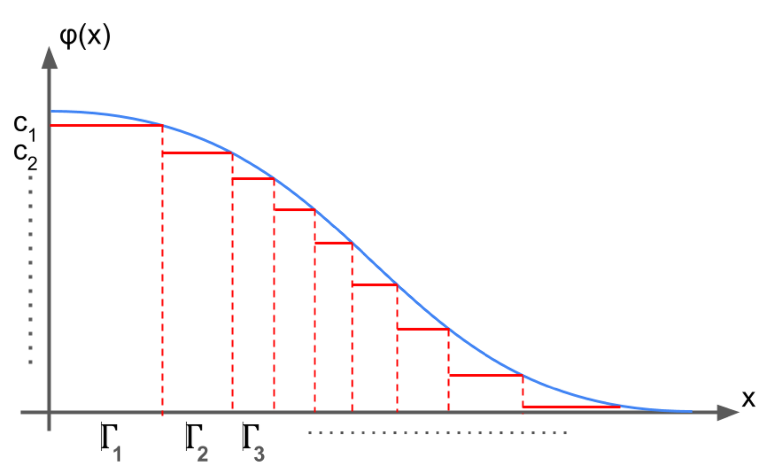


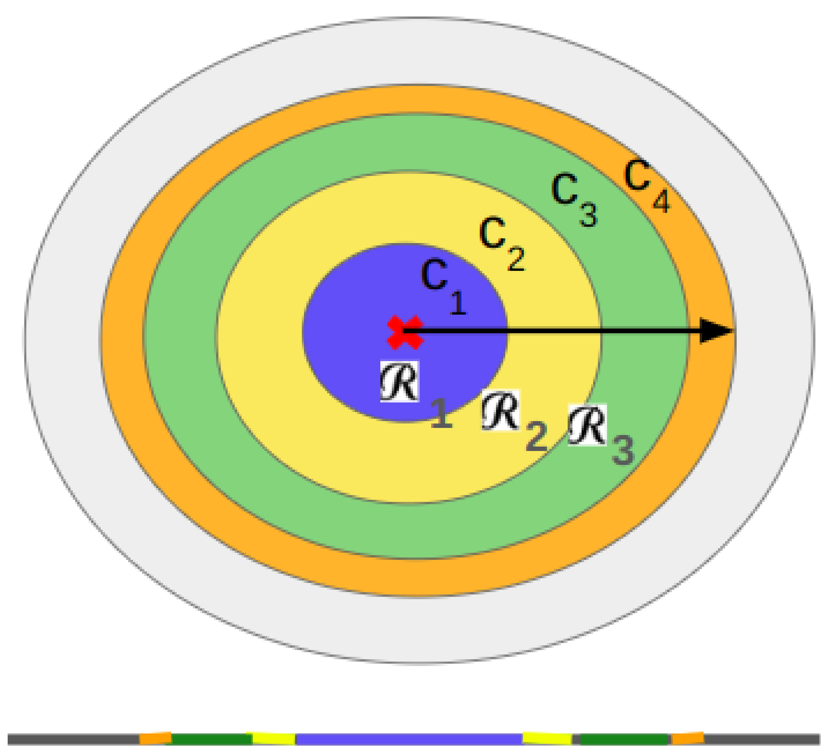


Impossibility

Log-likelihood function:

$$\sum_{\substack{v\sim 0\\ \sigma_v=\sigma_0}} \log\left(p\psi_{0v}\right) + \sum_{\substack{v\sim 0\\ \sigma_v\neq\sigma_0}} \log\left(q\psi_{0v}\right) + \sum_{\substack{v\nsim 0\\ \sigma_v=\sigma_0}} \log\left(1 - p\psi_{0v}\right) + \sum_{\substack{v\nsim 0\\ \sigma_v\neq\sigma_0}} \log\left(1 - q\psi_{0v}\right)$$





Problem Statement

Problem: Given the locations $\mathbf{X}$ and the adjacency matrix $\mathbf{A}$, recover $\boldsymbol{\sigma}$ exactly.

- Estimate $\hat{\boldsymbol{\sigma}}_n$ of $\boldsymbol{\sigma}_n$ recovers the communities exactly if

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{\boldsymbol{\sigma}}_n \in \{\pm\boldsymbol{\sigma}_n\}\right) = 1$$

- An estimate is said to recover the communities *almost-exactly* if for any $\eta > 0$, there exists an n_0 large enough such that for all $n \geq n_0$

$$\mathbb{P}\left(\max_{s\in\{\pm 1\}} |\{v \colon \tilde{\sigma}(v) = s\sigma(v)\}| \geq (1-\eta)n\right) \;=\; 1 - o(1).$$

Main Results

Define an information metric

$$I_\phi(p,q) \;:=\; 2\int_{\mathbb{R}_+}\left(1 - \sqrt{pq}\phi(x) - \sqrt{(1-p\phi(x))(1-q\phi(x))}\right)dx$$

and a normalised interaction range

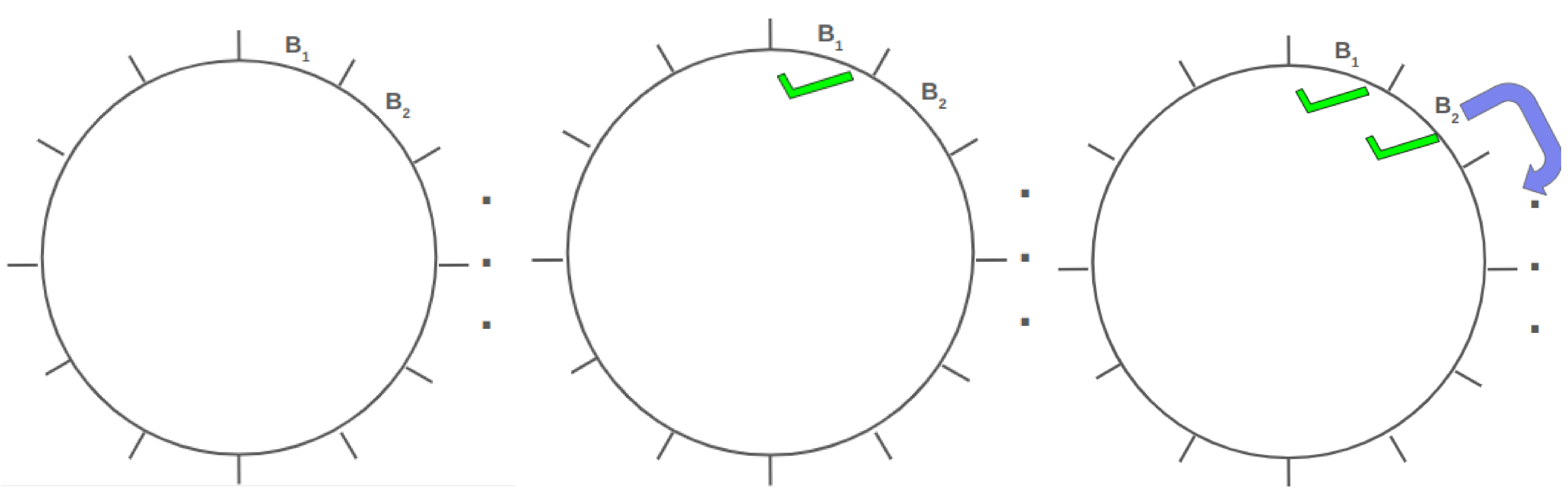
$$\kappa \;:=\; \sup\{x \colon \phi(x) \neq 0\}.$$

Impossibility. *Let $0 < \kappa < \infty$. If $\lambda\kappa < 1$ or $\lambda I_\phi(p,q) < 1$, then no estimator recovers the community structure exactly.*

Achievability. *Let $0 < \kappa < \infty$ and $\phi(x) > 0$ for all $x \in [0,\kappa]$. If $\lambda\kappa > 1$ and $\lambda I_\phi(p,q) > 1$, then there exists a linear-time algorithm (in the number of edges) that recovers the community structure exactly.*

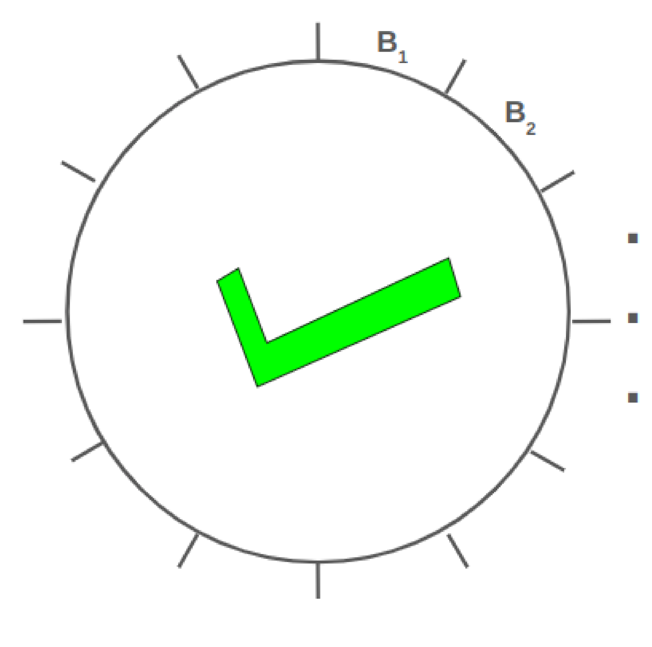
Algorithm

Phase I: Almost-exact recovery



Division into blocks — Initialization — Propagation

Phase II: Exact recovery



Refinement

References

- E. Abbe, F. Baccelli, and A. Sankararaman. *Community detection on Euclidean random graphs.* Information and Inference: A Journal of the IMA, 10(1):109–160, 2021.
- J. Gaudio, X. Niu, and E. Wei. *Exact community recovery in the geometric SBM.* In ACM-SIAM Symposium on Discrete Algorithms (SODA), 2024
- K. Avrachenkov, B. R. Vinay Kumar, L. Leskelä. *Community Detection on Block Models with Geometric Kernels.* Available at https://arxiv.org/abs/2403.02802