

Community detection via matrix factorization under the degree-corrected block model

Alexandra Dache*, Arnaud Vandaele*, and Nicolas Gillis*

* Department of Mathematics and Operational Research, University of Mons, Mons, Belgium,
✉ firstname.lastname@umons.ac.be

Community detection is a fundamental task in data analysis. Block models are a standard approach to partition nodes according to a graph model, facilitating the analysis and interpretation of the network structure. By grouping nodes with similar connection patterns, they enable the identification of a wide variety of underlying structures. The degree-corrected block model (DCBM) [3] is an established model that accounts for the heterogeneity of node degrees. However, standard inference methods for the DCBM are computationally costly and highly sensitive to initialization, while cheaper alternatives such as spectral or modularity-based approaches are limited to detecting only specific structures (typically assortative).

In this work, we reformulate the problem of DCBM inference as a constrained nonnegative matrix factorization (NMF) problem [2]. A DCBM with n nodes and r communities can be fully characterized by two parameter matrices. The first is an $n \times r$ matrix Z , where rows correspond to nodes and columns to communities. Each row of Z contains exactly one nonzero entry, indicating the node's community membership, with the value of that entry encoding its degree parameter. The second is an $r \times r$ matrix θ , capturing community interactions, where $\theta(k, \ell)$ quantifies the expected number of edges between nodes in communities k and ℓ . Under the DCBM, the expected adjacency matrix satisfies $\mathbb{E}[A] = Z\theta Z^\top$. Inference under the original Poisson-based DCBM [3] is equivalent to minimizing the Kullback-Leibler (KL) divergence between the observed adjacency matrix A and the low-rank factorization $Z\theta Z^\top$, subject to the orthogonality constraint $Z^\top Z = I_r$ and nonnegativity constraints $Z \geq 0$ and $\theta \geq 0$. Enforcing orthogonality among the nonnegative columns of Z ensures that each row contains at most one nonzero entry, corresponding to a single community assignment.

We propose replacing the KL divergence with the Frobenius norm, leading to the following alternative to DCBM inference for community detection:

$$\min_{Z \in \mathbb{R}_+^{n \times r}, \theta \in \mathbb{R}_+^{r \times r}} \|A - Z\theta Z^\top\|_F^2 \quad \text{s.t.} \quad Z^\top Z = I_r. \quad (1)$$

We solve this problem using an alternating optimization scheme. For a fixed Z , the matrix θ admits a closed-form update $\theta^* = Z^\top A Z$. To update Z , we employ a block coordinate descent approach that updates rows sequentially. For each node, we evaluate all possible community assignments by solving a univariate subproblem, which reduces to minimizing a degree-4 polynomial. The overall per-iteration complexity of our algorithm is $\mathcal{O}(rn\langle d \rangle)$, where $\langle d \rangle$ is the average node degree of the graph.

Like existing DCBM inference methods, our method also requires an initialization. To address this, we propose an efficient initialization method based on the separability property of Z . In the noiseless case, $A = WZ^\top$ with $W = Z\theta$ where Z is separable: there exists an index set of nodes $K \subset \{1, 2, \dots, n\}$ of size r such that $Z(K, :) = \text{diag}(z) \in \mathbb{R}^{r \times r}$, with a strictly positive vector $z \in \mathbb{R}_+^r$, where $\text{diag}(\cdot)$ denotes the diagonal matrix with z on its diagonal. In simpler terms, Z is separable if for each column of Z , there exists at least one row with a single nonzero entry in that column, i.e. at least one node belongs exclusively to each community. Under this condition, we have:

$$A(:, K) = WZ(K, :)^{\top} = W\text{diag}(z). \quad (2)$$

This means that each column of W is equal to a column of A up to scaling factors. In our case, we can further assume that each community contains more than one node. Consequently, Z actually contains several disjoint diagonal submatrices of size r , up to permutations. This is a more favourable situation, as it means that multiple columns of A are close to each column of W (up to scaling factors). To exploit

this structure, we use the smoothed vertex component analysis method (SVCA) proposed by [4]. SVCA estimates W in polynomial time with high robustness by averaging several carefully selected columns of A for each column of W . Then, we recover Z by solving the following problem:

$$\min_{Z \geq 0} \|A - WZ^\top\|_F^2 \quad \text{s.t.} \quad (Z^\top Z)_{i,j} = 0 \quad \forall i \neq j, \quad (3)$$

which admits a closed-form solution. This procedure yields an initial estimate of Z and thus a first partition of the nodes.

Experiments on synthetic and real benchmark networks show that our method achieves community detection comparable to DCBM inference while reducing total runtime. Moreover, the proposed initialization significantly improves solution quality compared to random initialization, leading inference algorithms to converge to better objective values, and reduces the number of iterations required for convergence across all tested algorithms. For further details, the reader is referred to the full version of this work [1]. We observed that using the Frobenius norm or the KL divergence can lead to better partitions depending on the network. It would be interesting to precisely identify the configurations for which one or the other of these measures is most suitable.

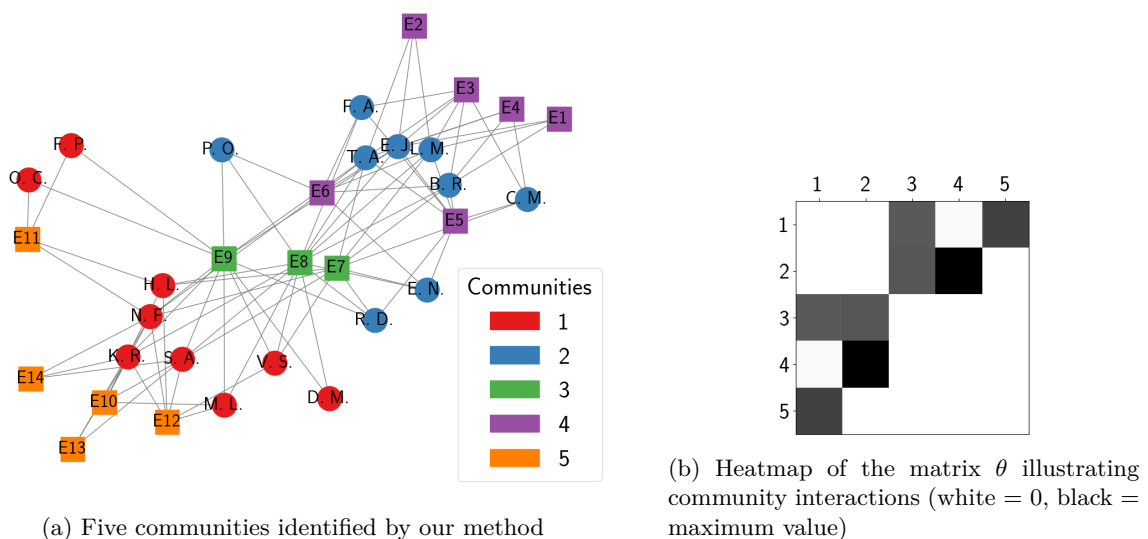


Figure 1: Results of our community detection method on the Southern Women dataset.

To illustrate our work, Figure 1 presents the results of applying our method to the Southern Women dataset. The figure shows the five detected communities and the corresponding matrix θ , which encodes interactions between communities. This dataset describes the participation of women in social events in a southern town in the United States. The network is composed of 32 nodes, 18 for women and 14 for events. Edges connect a woman to an event if she attended it. We observe that θ clearly reflects the underlying bipartite structure of the network.

References

- [1] Alexandra Dache, Arnaud Vandaele, and Nicolas Gillis. Matrix factorization framework for community detection under the degree-corrected block model. *arXiv preprint arXiv:2601.06262*, 2026.
- [2] N. Gillis. *Nonnegative Matrix Factorization*. SIAM, Philadelphia, 2020.
- [3] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107, 2011.
- [4] N. Nadisic, N. Gillis, and C. Kervazo. Smoothed separable nonnegative matrix factorization. *Linear Algebra and its Applications*, 676:174–204, 2023.