The Pairwise Prony Algorithm: Efficient Inference of Stochastic Block Models with Prescribed Subgraph Densities

Anonymous Authors[1]

Abstract

We present an elegant and flexible algorithm that provides the parameters of the simplest stochastic block model (SBM) for a given set of prescribed subgraph densities, from which one can sample networks with negligible computational overhead. The method generalizes the classical method of Prony to the pairwise data of networks. The class of inferred models are at the intersection of exponential random graph models (ERGMs), which are characterized in terms of maximum entropy, and of exchangeable random graphs (i.e., graphons). We show that the required subgraph densities can be efficiently computed for both dense and sparse networks, and provide an implementation of our algorithm in python. Our method provides standardized null models for statistical analysis of network data, including for the challenging case of a single observed graph.

1. Motivation

Statistical significance is generally meaningless without a properly specified alternative. The use of judiciously chosen distributions as standardized null models facilitates reproducibility and meaningful comparisons. For example:

"outliers differ by more than 3 standard deviations"
"false-positive rate for a χ^2 with 5 degrees of freedom"

While networks enjoy myriad summary statistics, their corresponding null models suffer from a lack of consensus. Yet, it is well-known that every distribution of infinitely exchangeable graphs is completely determined by the subgraph densities of all finite subgraphs (Lovász, 2012). This class of distributions contains all graphons and stochastic block models (SBMs) (Diaconis & Janson, 2008).

In classical statistics, the method of moments (Pearson, 1894; Prony, 1795) offers a powerful technique for inferring exchangeable distributions, in particular mixture models (Gordon et al., 2021). For network data, subgraph densities are the equivalent of the classical moments (Bickel et al., 2011). While it is possible to estimate a (finite number) of subgraph densities consistently and at a guaranteed asymptotic rate from a single observed graph as the graph grows (Zhang & Xia, 2022), estimating these densities is only a first step in the inference process. Currently there is no general method for translating them into an estimate of a graphon or stochastic block model. Instead, most inference algorithms used in practice (e.g., Gao et al. (2015); Latouche & Robin (2016); Peixoto (2014)) rely on some combination of maximum likelihood estimation, MCMC sampling, variational inference, clustering of the nodes, or other heuristics.

This work introduces a method that solves several problems simultaneously: efficiently estimating parameters without laborious and delicate numerical fitting, and efficiently sampling from the model without the often-encountered problem of degeneracy that plagues many intuitively attractive network models (Karwa et al., 2016). We generalize a classical algorithm for inferring latent sources (sometimes known as Prony's method) to graph data. The method is an example of an algorithm that recovers a sparse signal from noisy data, and is related to compressed sensing (Sauer, 2018) and the notion of a matrix pencil (Markus, 1988).

Our method proceeds in two steps. The first step is essentially an application of the classical Prony's method to estimate properties of individual latent blocks, such as their normalized degrees and their relative sizes. The second step is entirely new; by leveraging the properties inferred in the first step, it uses a generalization of Prony's method to infer properties of *pairs* of the latent blocks, such as the connection properties of a stochastic block model.

While many algorithms for fitting stochastic block models (e.g., Celisse et al. (2012); De Nicola et al. (2022); Deng et al. (2023)) and graphons (e.g, Airoldi et al. (2013); Li et al. (2022)) have been proposed, ours is unique in that it requires essentially zero computational overhead once the subgraph densities have been estimated.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the SPIGM workshop at ICML 2023. Do not distribute.

2. Notation and Definitions

We use lower-case bold symbols to denote vectors, and upper-case bold symbols to denote matrices and higher-order tensors. Non-bold versions of these symbols refer to particular entries of the bold versions, indicated by their subscripts. We use $\circ$ to denote Hadamard product (i.e., element-wise multiplication), and $\bullet$ for inner product.

A graph G is defined by a set of vertices/nodes $V(G)$ and a set of edges $E(G) \subseteq V(G) \times V(G)$ denoting pairwise connections between nodes. For simplicity, we focus on undirected, unweighted graphs, with no loops or multiple edges. Indexing the $N = |V(G)|$ nodes by the positive integers $[N]$, we G can represented by listing its edges (i, j), where $1 \leq i < j \leq N$. Or, equivalently, by its N-by-N adjacency matrix $\mathbf{A}$, where $A_{ij} = A_{ji} = 1$ if there is an edge between nodes i and j, and 0 otherwise.

A stochastic block model $\text{SBM}(\boldsymbol{\pi}, \mathbf{B})$ is defined by: $\boldsymbol{\pi}$, a probability vector of length K, which assigns each node to one of K (unobserved) latent blocks; and $\mathbf{B}$, a K-by-K connectivity matrix, whose entries $B_{kk'}$ give the probability of an edge connecting a node in block k to a node in block k'. As we are considering undirected graphs, $\mathbf{B}$ is symmetric.

Distributions over graphs can be characterized by their homomorphism subgraph densities $\mu(g)$. Indexed by (some family of) subgraphs g, they are the graphical analogue of the classical moments of a distribution. For a SBM, one can compute $\mu(g)$ by summing over all possible assignments of the vertices of g to the K blocks of $\text{SBM}(\boldsymbol{\pi}, \mathbf{B})$:

$$\underbrace{\mu(g)\Big|_{\text{SBM}(\boldsymbol{\pi},\mathbf{B})}}_{\substack{\text{homomorphism density } \mu \text{ of a subgraph } g \text{ in an} \\ \text{SBM given by } \boldsymbol{\pi} \text{ and } \mathbf{B} \\ \text{with } K \text{ blocks}}} = \underbrace{\sum_{\varphi:V(g)\to[K]}^{|V(g)|^K}}_{\substack{\text{sum over all maps } \varphi \\ \text{from vertices in } g \text{ to} \\ \text{the } K \text{ blocks}}} \left[\underbrace{\left(\prod_{i \in V(g)} \pi_{\varphi(i)} \right)}_{\substack{\text{probability of that} \\ \text{vertex assignment}}} \times \underbrace{\left(\prod_{(i,j)\in E(g)} B_{\varphi(i)\varphi(j)} \right)}_{\substack{\text{probability of the} \\ \text{corresponding edges}}} \right] \quad (1)$$

Sampling a graph from $\text{SBM}(\boldsymbol{\pi}, \mathbf{B})$ proceeds in two steps. First, for each node $n \in [N]$, sample its latent block $k(n) \in [K]$ independently from $\boldsymbol{\pi}$. Then, for each pair of nodes (n, n'), include an edge between them independently with probability $B_{k(n)k(n')}$. Thus, a stochastic block model defines a distribution over graphs with N nodes for each choice of N, and can be identified with the limit of this sequence of distributions as $N \to \infty$.

3. From Subgraph Densities to an SBM

We now describe how to recover the parameters of an SBM, given access to only its subgraph densities. In appendix A, we describe how to infer these subgraph densities from a graph sampled from such an SBM.

3.1. Classical Coin Collecting

Before explain our Pairwise Prony method, let us consider the simpler case of a mixture model for exchangeable sequences of binary variables. Also known as the Bernoulli mixture model, such a distribution can be thought of as the outcomes of (some number of) flips of (some number of) biased coins, where each coin is sampled i.i.d. from a (possibly unequal) mixture of K different biases. To recover this distribution, one must infer the K latent biases b_k, as well as the fraction π_k of coins with each bias. Note that one does not initially know which of the coins have the same latent bias (otherwise the inference problem would be trivial). The moments of this distribution are the expectation of powers of these biases:

$$\langle b^r \rangle = \sum_k \pi_k b_k^r \quad (2)$$

The mixture proportion $\boldsymbol{\pi}$ and the biases $\mathbf{b}$ can be systematically inferred from these moments using the standard Prony's method. In short, construct two matrices $\mathbf{C}$ and $\mathbf{C}'$, with entries $C_{ij} = \langle b^{i+j} \rangle$ and $C'_{ij} = \langle b^{i+j+1} \rangle$

$$\mathbf{C} = \begin{bmatrix} \langle b^0 \rangle & \cdots & \langle b^{K-1} \rangle \\ \vdots & \ddots & \vdots \\ \langle b^{K-1} \rangle & \cdots & \langle b^{2K-2} \rangle \end{bmatrix} \quad (3)$$

$$\mathbf{C}' = \begin{bmatrix} \langle b^1 \rangle & \cdots & \langle b^{K} \rangle \\ \vdots & \ddots & \vdots \\ \langle b^{K} \rangle & \cdots & \langle b^{2K-1} \rangle \end{bmatrix} \quad (4)$$

While not immediately obvious, it is easy to show that the eigenvalues of $\mathbf{C}'\mathbf{C}^{-1}$ are the entries of $\mathbf{b}$, from which one can obtain the associated entries of $\boldsymbol{\pi}$. In the next section, we show why this is the case, while superficially replacing the biases $\mathbf{b}$ of each coin type with the average normalized degrees $\mathbf{d}$ of each block of nodes.

3.2. Distilling the Degree Distribution

The first step of the Pairwise Prony method is to apply this standard Prony's Method to the obtain the normalized degrees $\mathbf{d}$ of the blocks:

$$d_k = \sum_j \pi_j B_{jk} \quad (5)$$

Much like the coin biases from before giving the expected fraction of heads for a coin of a given type, the normalized degrees give the expected fraction of other nodes that share an edge with a node in a given block. As before, we define two matrices $C^{\mathbf{\circ}}{}_{ij} = \langle d^{i+j} \rangle$ and $C^{\mathbf{d}}{}_{ij} = \langle d^{i+j+1} \rangle$, where

$$\langle d^r \rangle = \sum_k \pi_k d_k^r \quad (6)$$

Note that, while π_k and d_k are initially unknown, the resulting moments are observable — they are precisely the homomorphism densities of the star subgraphs

$$\langle d^0 \rangle = \mu(\,\bullet\,) = 1 \qquad \langle d^1 \rangle = \mu(\,\bullet\!-\!\bullet\,)$$
$$\langle d^2 \rangle = \mu(\,\bullet\!-\!\bullet\!-\!\bullet\,) \qquad \langle d^3 \rangle = \mu(\,\bullet\!-\!\bullet\!-\!\bullet\!-\!\bullet\,) \tag{7}$$

As mentioned in the previous section, the eigenvalues of $\mathbf{C}^{\mathbf{d}}(\mathbf{C}^{\boldsymbol{\circ}})^{-1}$ give the normalized degrees of the K blocks. To see why this is the case, notice that $\mathbf{C}^{\boldsymbol{\circ}}$ can be written as a sum of K rank-1 matrices (weighted by $\boldsymbol{\pi}$)

$$\begin{aligned}\mathbf{C}^{\boldsymbol{\circ}} &= \sum_k \pi_k \begin{bmatrix} d_k^0 & \cdots & d_k^{K-1} \\ \vdots & \ddots & \vdots \\ d_k^{K-1} & \cdots & d_k^{2K-2} \end{bmatrix} \\ &= \sum_k \pi_k \begin{bmatrix} d_k^0 \\ \vdots \\ d_k^{K-1} \end{bmatrix} \begin{bmatrix} d_k^0 \\ \vdots \\ d_k^{K-1} \end{bmatrix}^\top \\ &= \mathbf{V}\,\mathrm{diag}(\boldsymbol{\pi})\,\mathbf{V}^\top \end{aligned} \tag{8}$$

where $\mathbf{V}$ is a matrix with entries $V_{jk} = d_k^{j-1}$

$$\mathbf{V} = \begin{bmatrix} d_1^0 & \cdots & d_K^0 \\ \vdots & \ddots & \vdots \\ d_1^{K-1} & \cdots & d_K^{K-1} \end{bmatrix} \tag{9}$$

By decomposing $\mathbf{C}^{\mathbf{d}}$ in the same manner,

$$\mathbf{C}^{\mathbf{d}} = \mathbf{V}\,\mathrm{diag}(\boldsymbol{\pi}\mathbf{d})\,\mathbf{V}^\top \tag{10}$$

we find that $\mathbf{C}^{\mathbf{d}}(\mathbf{C}^{\boldsymbol{\circ}})^{-1}$ can be diagonalized as follows

$$\begin{aligned}\mathbf{C}^{\mathbf{d}}(\mathbf{C}^{\boldsymbol{\circ}})^{-1} &= \big(\mathbf{V}\,\mathrm{diag}(\boldsymbol{\pi}\mathbf{d})\,\mathbf{V}^\top\big)\big(\mathbf{V}\,\mathrm{diag}(\boldsymbol{\pi})\,\mathbf{V}^\top\big)^{-1} \\ &= \mathbf{V}\,\mathrm{diag}(\mathbf{d})\,\mathbf{V}^{-1} \end{aligned} \tag{11}$$

Thus, its spectrum is indeed the normalized degrees $\mathbf{d}$ of the K latent blocks:

$$\mathrm{eigval}\Big(\mathbf{C}^{\mathbf{d}}(\mathbf{C}^{\boldsymbol{\circ}})^{-1}\Big) = \Big\{d_k\Big\}_{k\in[K]} \tag{12}$$

(Note that this procedure relies on the fact that the normalized degrees of the blocks are unique, otherwise $\mathbf{V}$ will not be invertible.) After obtaining the entries of $\mathbf{d}$, one can construct $\mathbf{V}$ and solve a linear system of equations for the corresponding entries of $\boldsymbol{\pi}$:

$$\sum_k V_{jk}\pi_k = \langle d^{j-1} \rangle \tag{13}$$

3.3. Gluing Graphs: An Algebra

Before moving to the second step of our method, it is useful to introduce the notion of (partially-)labelled homomorphism densities. Together with the gluing product, these densities form an algebra that is well-suited for describing the Pairwise Prony method. Tables summarizing these operations are provided in appendix C.

Recall equation (1), the definition of (unlabelled) homomorphism subgraph densities of an SBM. For a given distribution $\mathrm{SBM}(\boldsymbol{\pi}, \mathbf{B})$, each subgraph g corresponds to a scalar density $\mu(g)$. Here we introduce (singly-)labelled homomorphism densities, which take as input a subgraph g with a designated "labelled" vertex $v \in V(g)$, and return K scalars, one for each of the K blocks to which the labelled vertex v is mapped:

$$\underbrace{\mu_k(g; v)\big|_{\mathrm{SBM}(\boldsymbol{\pi},\mathbf{B})}}_{\substack{\text{homomorphism density}\\ \mu \text{ of a subgraph } g \text{ with}\\ \text{labelled vertex } v \text{ in block } k}} = \underbrace{\sum_{\substack{\varphi: V(g)\to[K]\\ \varphi(v)=k}}^{|V(g)\setminus v|^K}}_{\substack{\text{sum over all maps } \varphi \text{ that}\\ \text{send vertex } v \text{ to block } k}} \Bigg[\underbrace{\Bigg(\prod_{i\in V(g)\setminus v} \pi_{\varphi(i)}\Bigg)}_{\substack{\text{probability of that assignment}\\ \text{of unlabelled vertices}}} \times \underbrace{\Bigg(\prod_{(i,j)\in E(g)} B_{\varphi(i)\varphi(j)}\Bigg)}_{\substack{\text{probability of the}\\ \text{corresponding edges}}}\Bigg] \tag{14}$$

The sum is now over maps φ that send the labelled vertex v to a particular block k, and the probability of vertex assignment includes only the remaining unlabelled vertices. All the edges $E(g)$ contribute to the product exactly as in equation (1). A labelled vertex with a single edge corresponds to the normalized degrees of the K blocks:

$$\bullet\!-\!\circ \longleftrightarrow \mathbf{d} \tag{15}$$

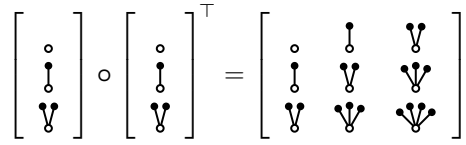
Labelled subgraphs may be combined via an operation known as the gluing product,(Lovász, 2012) obtained by taking their disjoint union, then merging the labelled vertices:

$$\begin{aligned} \bullet\!-\!\circ \circ \bullet\!-\!\circ &= \bullet\!-\!\bullet\!-\!\circ \longleftrightarrow \mathbf{d}\circ\mathbf{d} = \mathbf{d}^2 \\ \bullet\!-\!\circ \circ \bullet\!-\!\bullet\!-\!\circ &= \bullet\!-\!\bullet\!-\!\bullet\!-\!\circ \longleftrightarrow \mathbf{d}\circ\mathbf{d}^2 = \mathbf{d}^3 \\ \bullet\!-\!\bullet\!-\!\circ \circ \bullet\!-\!\bullet\!-\!\circ &= \text{star}_4 \longleftrightarrow \mathbf{d}^2\circ\mathbf{d}^3 = \mathbf{d}^5 \end{aligned} \tag{16}$$

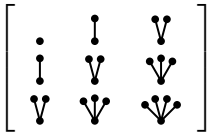
As this corresponds algebraically to entrywise multiplication of vectors, we use the same notation $\circ$ for the gluing product as for the Hadamard product. Finally, to convert a labelled subgraph into an observable subgraph density, the labelled vertex may be "unlabelled" by taking the dot product with $\boldsymbol{\pi}$

$$\text{star}_4 \longleftrightarrow \mathbf{d}^3 \bullet \boldsymbol{\pi} = \langle d^4 \rangle \tag{17}$$

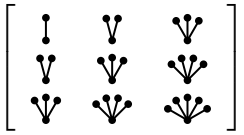
Before describing how to recover the entries of $\mathbf{B}$, let us summarize the method for recovering the normalized degrees from the previous section using the language of labelled graphs. For an SBM with K blocks, define a vector $\mathbf{v}$ of (at least) K labelled subgraphs, and take the outer product of this vector with itself where entries are combined via the gluing product. For example:

$$\mathbf{v} \circ \mathbf{v}^\top =$$ 

While the labelled subgraph densities are not directly observable, their unlabelled counterparts are, and the entries of $\mathbf{C}^\circ$ (equation (4), left) are given by the densities of these unlabeled versions

$$\mathbf{C}^\circ = \left(\mathbf{v} \circ \mathbf{v}^\top\right) \bullet \boldsymbol{\pi} =$$ 

(The glyph containing an unlabelled node and no edges evaluates to 1.) To obtain the second matrix $\mathbf{C}^{\mathbf{d}}$ (equation (4), right), glue the labeled subgraph  to all entries prior to unlabelling

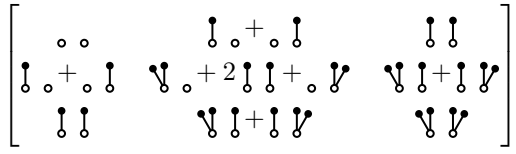
$$\mathbf{C}^{\mathbf{d}} = \left(\left(\mathbf{v} \circ \mathbf{v}^\top\right) \circ$$ $$\right) \bullet \boldsymbol{\pi} =$$ 

Indeed, the reason that the spectrum of $\mathbf{C}^\circ(\mathbf{C}^{\mathbf{d}})^{-1}$ gives the normalized degrees $\mathbf{d}$ of the latent blocks is precisely *because* we glued an extra copy of the corresponding labelled subgraph in the construction of $\mathbf{C}^{\mathbf{d}}$. The next section follows a similar recipe.

3.4. Extracting the Edge Expectations

The previous matrices $\mathbf{C}^\circ$ and $\mathbf{C}^{\mathbf{d}}$ were obtained by taking a dot product that sums over the K blocks. This allowed us to obtain properties of a single block (e.g., its degree d_k). A main insight of this work is that Prony's method can also be used to obtain properties of pairs of blocks (i.e., their connection probability $B_{kk'}$) by using the gluing algebra of *bilabelled* subgraphs. A bilabelled subgraph $(g; u, v)$ has two (distinct) labelled vertices, and their gluing product is given by taking the disjoint union then separately merging the vertices labelled u and the vertices labelled v.

For an SBM with K blocks, there are $K + \binom{K}{2}$ degrees of freedom for the entries of $\mathbf{B}$, so we require a vector $\mathbf{v}$ of at least $K + \binom{K}{2}$ bilabelled subgraphs. To this end, we use the symmetric polynomials (with exponent at most K) in the two variables  and  , corresponding to the normalized degrees of the two blocks. For example:

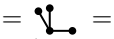
$$\mathbf{v} \circ \mathbf{v}^\top =$$ 

In this example for $K = 2$, the second row of $\mathbf{v}$ has been symmetrized with respect to the two labelled vertices, resulting in a formal linear combination of elements in the gluing algebra.

As before, $\mathbf{C}^{\circ\circ}$ is obtained by unrooting the entries, while for the construction of $\mathbf{C}^{\mathbf{B}}$, we glue the labelled subgraph  , which corresponds to entries of $\mathbf{B}$. Continuining with the $K = 2$ blocks SBM example:

$$\mathbf{C}^{\circ\circ} =$$ (18)

$$\mathbf{C}^{\mathbf{B}} =$$ (19)

Note that, once unlabelled, isomorphic graphs are equivalent, e.g.:  $=$  $=$  . As promised, the eigenvalues of $\mathbf{C}^{\mathbf{B}}(\mathbf{C}^{\circ\circ})^{-1}$ are precisely the entries of $\mathbf{B}$

$$\operatorname{eigval}\left(\mathbf{C}^{\mathbf{B}}(\mathbf{C}^{\circ\circ})^{-1}\right) = \left\{B_{kk'}\right\}_{k \le k'} \tag{20}$$

Moreover, the structure of their corresponding eigenvectors allows us to select specific entries $B_{kk'}$

$$\operatorname{eigvec}\left(\mathbf{C}^{\mathbf{B}}(\mathbf{C}^{\circ\circ})^{-1}\right) = \left\{ \begin{bmatrix} 1 \\ d_k + d_{k'} \\ d_k d_{k'} \end{bmatrix} \right\}_{k \le k'} \tag{21}$$

Thus, for each pair of blocks k and k' with normalized degrees d_k and $d_{k'}$, we can estimate $B_{kk'}$

$$B_{kk'} = \frac{\mathbf{v}^\top(\mathbf{C}^{\mathbf{B}}(\mathbf{C}^{\circ\circ})^{-1})\mathbf{v}}{\mathbf{v}^\top\mathbf{v}} \qquad \mathbf{v} = \begin{bmatrix} 1 \\ d_k + d_{k'} \\ d_k d_{k'} \end{bmatrix}$$

In appendix B, we discuss how to add more subgraph to this method, and show results of applying our method in appendix D.

4. Concluding Context

The power of statistics lies in its ability to summarize large quantities of data with appropriate summary statistics. While the map from stochastic block models to subgraph densities is essentially a counting exercise, the map from subgraph densities to stochastic block models is rather nontrivial. This work offers such an inverse map.

The method we presented can be easily extended in a variety of ways (directed edges, weighted edges, directed weighted edges, etc). While we aimed to make this particular instantiation as clear as possible, the broader message we would like to convey is that the space of clever inferential methods is by no means exhausted, and inspiration can be found even in stepping-stones over two centuries old.

References

Abbe, E. and Sandon, C. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 670–688. IEEE, 2015.

Airoldi, E. M., Costa, T. B., and Chan, S. H. Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. *Advances in Neural Information Processing Systems*, 26, 2013.

Ambroise, C. and Matias, C. New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Stat. Soc. Ser. B.*, 74:3–35, 2012.

Anandkumar, A., Ge, R., Hsu, D., and Kakade, S. A tensor spectral approach to learning mixed membership community models. In *Conference on Learning Theory*, pp. 867–881. PMLR, 2013.

Bickel, P. J., Chen, A., and Levina, E. The method of moments and degree distributions for network models. *Ann. Statist.*, 39(5):2280–2301, 2011.

Bravo-Hermsdorff, G., Gunderson, L. M., Maugis, P.-A., and Priebe, C. E. A principled (and practical) test for network comparison. *arXiv preprint arXiv:2107.11403*, 2021.

Celisse, A., Daudin, J.-J., and Pierre, L. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847 – 1899, 2012.

Channarond, A., Daudin, J.-J., and Robin, S. Classification and estimation in the stochastic blockmodel based on the empirical degrees. 2012.

De Nicola, G., Sischka, B., and Kauermann, G. Mixture models and networks: The stochastic blockmodel. *Statistical Modelling*, 22(1-2):67–94, 2022.

Deng, J., Huang, D., Chang, X., and Zhang, B. Subsampling-based modified bayesian information criterion for large-scale stochastic block models. *arXiv preprint arXiv:2304.06900*, 2023.

Diaconis, P. and Janson, S. Graph limits and exchangeable random graphs. *Rendiconti di Matematica, Serie VII*, 28: 33–61, 2008.

Gao, C., Lu, Y., and Zhou, H. H. Rate-optimal graphon estimation. 2015.

Gordon, S., Mazaheri, B. H., Rabani, Y., and Schulman, L. Source identification for mixtures of product distributions. In *Conference on Learning Theory*, pp. 2193–2216. PMLR, 2021.

Gunderson, L. M. and Bravo-Hermsdorff, G. Introducing graph cumulants: What is the variance of your social network? *arXiv:2002.03959*, 2020.

Karwa, V., Petrović, S., and Bajić, D. Dergms: Degeneracy-restricted exponential random graph models. *arXiv preprint arXiv:1612.03054*, 2016.

Latouche, P. and Robin, S. Variational bayes model averaging for graphon functions and motif frequencies inference in w-graph models. *Statistics and Computing*, 26:1173–1185, 2016.

Lauritzen, S., Rinaldo, A., and Sadeghi, K. Random networks, graphical models and exchangeability. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 80(3):481–508, 2018.

Li, Y., Fan, X., Chen, L., Li, B., and Sisson, S. A. Smoothing graphons for modelling exchangeable relational data. *Machine Learning*, 111(1):319–344, 2022.

Lovász, L. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

Markus, A. S. *Introduction to the Spectral Theory of Polynomial Operator Pencils*. American Mathematical Society, 1988.

Newman, M. *Networks*. Oxford university press, 2018.

Orbanz, P. Subsampling large graphs and invariance in networks. *arXiv:1710.04217*, 2017.

Pearson, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894.

Peixoto, T. P. Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1):011047, 2014.

Prony, G. Essai experiméntal et analytique sur les lois de la dilatabilité des fluides élastiques, et sur celles de la force expansive de l'eau et de la vapeur de l'alkool, à différents températures. *Journal de l'École Polytechnique (Paris)*, 1:24–76, 1795.

Sauer, T. Prony's method: an old trick for new problems. 2018.

White, J. G., Southgate, E., Thomson, J. N., Brenner, S., et al. The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1–340, 1986.

Wolfe, P. J. and Olhede, S. C. Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*, 2013.

Zhang, Y. and Xia, D. Edgeworth expansions for network moments. *The Annals of Statistics*, 50(2):726–753, 2022.

A. From an Observed Graph to an SBM

In order to apply the Pairwise Prony method to an observed graph, we need to obtain unbiased estimators of the homomorphism densities used to construct the $\mathbf{C}$ matrices. In particular, for a subgraph g and a (large) graph G, we want to compute the fraction of *injective* maps φ from $V(g)$ to $V(G)$, such that if $(u, v) \in E(g)$, then $(\varphi(u), \varphi(v)) \in E(G)$.

A.1. Quickly Counting (Injective) Homomorphisms

We start with the adjacency matrix $\mathbf{A}$, the identity matrix $\mathbf{I}$, and the all-ones matrix $\mathbf{1}$. Then, we perform our only computationally intensive matrix multiplication: $\sum_k A_{ik}A_{kj} = D_{ij} + \Lambda_{ij}$, where $\mathbf{D}$ is the diagonal matrix of node degrees, and $\mathbf{\Lambda}$ is the (traceless) "two-hop adjacency" matrix. From these, we can recursively count all the relevant subgraphs using only entrywise multiplication.

We index our subgraphs with a tuple of non-negative integers (ℓ, c, r), corresponding to: the number of edges incident to only the "left" node i, the number of two-hop paths between the "left" and "right" nodes, and the number of edges incident to only the "right" node j. First, we add the two-hop paths:

$$\mathbf{M}^{(0,0,0)} = \mathbf{1} - \mathbf{I} \tag{22}$$

$$\mathbf{M}^{(0,c+1,0)} = \mathbf{M}^{(0,c,0)} \circ \big(\mathbf{\Lambda} - c\mathbf{1}\big) \tag{23}$$

Define the matrix of "left" degrees to be $\mathbf{L} = \mathbf{D} \bullet \big(\mathbf{1} - \mathbf{I}\big) - \mathbf{A}$ (i.e., the degree of node i if node $j \neq i$ were deleted, and zero when $i = j$), and the matrix of "right" degrees to be its transpose $\mathbf{R} = \mathbf{L}^\top$. Next, we add single edges to node i, then to node j:

$$\mathbf{M}^{(\ell+1,c,0)} = \mathbf{M}^{(\ell,c,0)} \circ \big(\mathbf{L} - (\ell + c)\mathbf{1}\big) \tag{24}$$

$$\mathbf{M}^{(\ell,c,r+1)} = \mathbf{M}^{(\ell,c,r)} \circ \big(\mathbf{R} - (r + c)\mathbf{1}\big) - \ell\mathbf{M}^{(\ell-1,c+1,r-1)} \tag{25}$$

Finally, if an edge connecting the left and right nodes is to be included, we put a mark on the middle integer:

$$\mathbf{M}^{(\ell,c',r)} = \mathbf{M}^{(\ell,c,r)} \circ \mathbf{A} \tag{26}$$

To obtain the counts of a subgraph g in the entire graph, simply sum the entries of the corresponding matrix $\mathbf{M}^{(\ell,c,r)}$. This is analogous to the "unlabelling" operation from before.

$$\Big[\!\Big[\mathbf{M}^{(\ell,c,r)}\Big]\!\Big] = \sum_i \sum_j \mathbf{M}^{(\ell,c,r)} \tag{27}$$

For example, $[\![\mathbf{A}]\!]$ is twice the number of edges, as the injective homomorphisms count both orientations. The injective homomorphism densities are obtained by dividing this count by the number of injective maps

$$\mu\Big(g^{(\ell,c,r)}\Big) = \frac{\Big[\!\Big[\mathbf{M}^{(\ell,c,r)}\Big]\!\Big]}{N\big(N-1\big)\cdots\big(N-|V(g)|+1\big)} \tag{28}$$

B. Beyond Degree Correlations: Cycles Do Better

The method described in the main text can be thought of as a "basic" minimum working example of a much more general family of Pairwise Prony methods. The subgraph densities in $\mathbf{C}^{\circ}$ and $\mathbf{C}^{\mathbf{d}}$ are sensitive only to the degree distribution, and the subgraphs in $\mathbf{C}^{\circ\circ}$ and $\mathbf{C}^{\mathbf{B}}$ contain information about the degree-degree correlations. While this is sufficient to recover the parameters of an SBM when the normalized degrees of the blocks are well-separated, the method can sometimes be made more robust by adding additional subgraphs.

In particular, define the "two-hop" matrix of connection probabilities $\mathbf{\Lambda} = \mathbf{B} \bullet_{\boldsymbol{\pi}} \mathbf{B}$, where $\bullet_{\boldsymbol{\pi}}$ denotes the matrix product weighted by $\boldsymbol{\pi}$, ie:

$$\Lambda_{ij} = \sum_k B_{ik}\pi_k B_{kj} \tag{29}$$

These bilabelled subgraphs interact with the others via gluing and unlabelling (see table 3).

We can then use this to add more columns to $\mathbf{C}^{\circ\circ}$ and $\mathbf{C}^{\mathbf{B}}$:

$$\mathbf{C}^{\circ\circ} = \left[\begin{array}{ccc|ccc} \cdot\,\cdot & 2\,\cdot\!-\!\cdot & \cdot\;\cdot \\ 2\,\cdot\!-\!\cdot & 2\,\cdot\!-\!\cdot\!<\!+\,2\,\cdot\;\cdot & 2\,\cdot\!-\!\cdot\;\cdot \\ \cdot\;\cdot & 2\,\cdot\!-\!\cdot\;\cdot & \cdot\!-\!\cdot\!-\!\cdot \end{array}\right. \left.\begin{array}{ccc} <\!\cdot\!> & 2\,<\!\cdot\!> & \cdot\!-\!<\!\cdot\!>\!-\!\cdot \\ 2\,<\!\cdot\!> & 2\,<\!\cdot\!>\!<\!+\,2\,\cdot\!-\!<\!\cdot\!>\!-\!\cdot & 2\,\cdot\!-\!<\!\cdot\!>\!-\!\cdot \\ <\!\cdot\!> & 2\,\cdot\!-\!<\!\cdot\!>\!-\!\cdot & \cdot\!-\!<\!\cdot\!>\!-\!\cdot \end{array}\right] \tag{30}$$

$$\mathbf{C}^{\mathbf{B}} = \left[\begin{array}{ccc|ccc} \cdot\!-\!\cdot & 2\,\cdot\!-\!\cdot\!-\!\cdot & \cdot\!-\!\cdot\;\cdot \\ 2\,\cdot\!-\!\cdot\!-\!\cdot & 2\,\cdot\!-\!\cdot\!-\!<\!+\,2\,\cdot\!-\!\cdot\;\cdot & 2\,\cdot\!-\!\cdot\!-\!\cdot\;\cdot \\ \cdot\!-\!\cdot\;\cdot & 2\,\cdot\!-\!\cdot\!-\!\cdot\;\cdot & \cdot\!-\!\cdot\!-\!\cdot\!-\!\cdot \end{array}\right. \left.\begin{array}{ccc} <\!\cdot\!> & 2\,<\!\cdot\!> & \cdot\!-\!<\!\cdot\!>\!-\!\cdot \\ 2\,<\!\cdot\!> & 2\,<\!\cdot\!>\!<\!+\,2\,\cdot\!-\!<\!\cdot\!>\!-\!\cdot & 2\,\cdot\!-\!<\!\cdot\!>\!-\!\cdot \\ <\!\cdot\!> & 2\,\cdot\!-\!<\!\cdot\!>\!-\!\cdot & \cdot\!-\!<\!\cdot\!>\!-\!\cdot \end{array}\right] \tag{31}$$

Remarkably, we can use these larger matrices in exactly the same way as before (using the Moore-Penrose pseudoinverse). As shown in figure 1 in appendix D, including these additional subgraphs can make the recovery of the SBM more robust.

C. Guide for Gluing Graphs

glyph	symbol	meaning
$\cdot\!-\!\cdot$	$\mathbf{d} = \mathbf{B} \bullet \boldsymbol{\pi}$	vector of the K normalized degrees
$\cdot\!<\!\cdot\!>\!\cdot$	$\mathbf{d}^2 = \mathbf{d} \circ \mathbf{d}$	entrywise multiplication
$\cdot\!-\!\cdot\!-\!\cdot$	$\mathbf{d}^3 \bullet \boldsymbol{\pi}$	homomorphism density of subgraph

Table 1. Operations using homomorphism densities of singly-labelled subgraphs.

glyph	symbol	meaning
$\cdot\!-\!\cdot\;\cdot$	$\mathbf{d}\mathbf{1}^\top = \mathbf{L}$	row matrix of (left) normalized degrees
$\cdot\;\cdot\!-\!\cdot$	$\mathbf{1}\mathbf{d}^\top = \mathbf{R}$	column matrix of (right) normalized degrees
$\cdot\!-\!\cdot$	$\mathbf{B}$	matrix of connection probabilities
$\cdot\!-\!\cdot\!-\!\cdot$	$\mathbf{L}^2 \circ \mathbf{B} \circ \mathbf{R}$	entrywise multiplication
$\cdot\!-\!\cdot\!-\!\cdot$	$\boldsymbol{\pi} \bullet \left(\mathbf{L}^2 \circ \mathbf{B} \circ \mathbf{R}\right) \bullet \boldsymbol{\pi} = \mu($ $)$	observable subgraph density

Table 2. Operations using homomorphism densities of doubly-labeled subgraphs.

glyph	symbol	meaning
$<\!\cdot\!>$	$\mathbf{B} \bullet_{\boldsymbol{\pi}} \mathbf{B} = \boldsymbol{\Lambda}$	two-hop connection probability
$<\!\cdot\!>\!-\!\cdot$	$\boldsymbol{\Lambda} \circ \mathbf{B}$	entrywise multiplication
$<\!\cdot\!>\!-\!\cdot\!-\!\cdot$	$\boldsymbol{\pi} \bullet \left(\boldsymbol{\Lambda} \circ \mathbf{B}\right) \bullet \boldsymbol{\pi} = \mu($ $)$	homomorphism density of triangles

Table 3. Incorporating the "two-hop" connection probability $\boldsymbol{\Lambda}$ into the gluing algebra.

D. Some Simulations

Here we provide a few more simulations for recovering the parameters of 2-by-2, 3-by-3, and 4-by-4 stochastic block models. The code implementing the method will be made available upon acceptance.

To best illustrate the usefulness of the method, we chose stochastic block models with communities that have similar normalized degree densities and rather idiosyncratic inter-community connection probabilities:

$K = 2$ (64 iterations)

$\boldsymbol{\pi}$	$\mathbf{B}$		$\mathbf{d}$
0.25	0.850	0.375	0.49375
0.75	0.375	0.400	0.39375

$K = 3$ (32 iterations)

$\boldsymbol{\pi}$	$\mathbf{B}$			$\mathbf{d}$
0.25	0.9	0.3	0.6	0.600
0.25	0.3	0.7	0.5	0.500
0.50	0.6	0.5	0.3	0.425

$K = 4$ (16 iterations)

$\boldsymbol{\pi}$	$\mathbf{B}$				$\mathbf{d}$
0.25	0.8	0.4	0.6	0.8	0.65
0.25	0.4	0.8	0.7	0.3	0.55
0.25	0.6	0.7	0.2	0.5	0.50
0.25	0.8	0.3	0.5	0.2	0.45

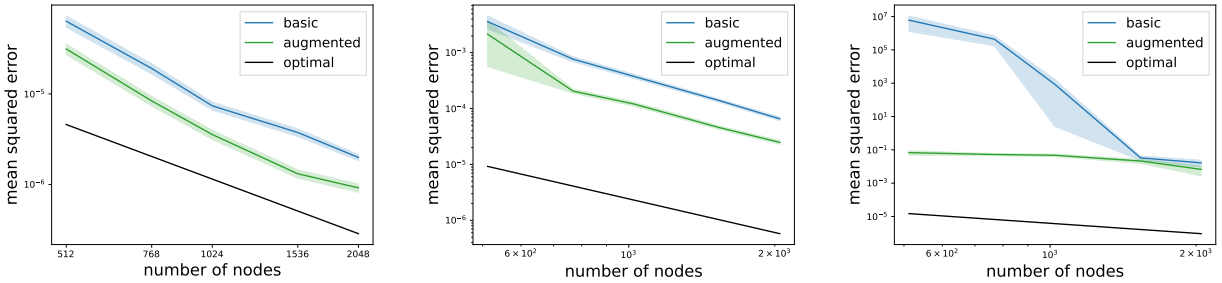


Figure 1. Recovery of the parameters of SBMs using our proposed method without adding extra subgraph densities (green) and adding the "two-hop" density (blue). Left figure corresponds to SBMs with 2 blocks, middle figure to SBMs with 3 blocks, and right figure to SBMs with 4 blocks. Vertical axis denotes the average squared error of the probability of a random dyad, i.e. $\boldsymbol{\pi}_{\text{true}}^{\top}(\mathbf{B}_{\text{true}} - \mathbf{B}_{\text{infer}})^2\boldsymbol{\pi}_{\text{true}}$. Shading denotes one standard error. The black curve denotes the expected squared error if the latent blocks of the nodes were provided.

In addition, we performed a fit to the C. Elegans connectome (White et al., 1986) containing 297 nodes and 2148 edges, obtaining the following parameters for a 4-block SBM, obtaining:

$\boldsymbol{\pi}$	$\mathbf{B}$			
0.0041	0.6891	0.1024	0.3927	0.5367
0.0483	0.1024	0.6314	0.2134	0.0770
0.5628	0.3927	0.2134	0.0663	0.0184
0.3849	0.5367	0.0770	0.0184	0.0056