

## MIT Open Access Articles

*Maximum Likelihood Embedding of  
Logistic Random Dot Product Graphs*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** O'Connor, Luke et al. "Maximum Likelihood Embedding of Logistic Random Dot Product Graphs." Proceedings of the AAAI Conference on Artificial Intelligence, 34, 4 (April 2020): 5289-5297 © 2020 The Author(s)

**As Published:** 10.1609/AAAI.V34I04.5975

**Publisher:** Association for the Advancement of Artificial Intelligence (AAAI)

**Persistent URL:** <https://hdl.handle.net/1721.1/130525>

**Version:** Original manuscript: author's manuscript prior to formal peer review

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Maximum Likelihood Latent Space Embedding of Logistic Random Dot Product Graphs

Luke O'Connor, Muriel Médard and Soheil Feizi

**Abstract**—A latent space model for a family of random graphs assigns real-valued vectors to nodes of the graph such that edge probabilities are determined by latent positions. Latent space models provide a natural statistical framework for graph visualizing and clustering. A latent space model of particular interest is the Random Dot Product Graph (RDPG), which can be fit using an efficient spectral method; however, this method is based on a heuristic that can fail, even in simple cases. Here, we consider a closely related latent space model, the Logistic RDPG, which uses a logistic link function to map from latent positions to edge likelihoods. Over this model, we show that asymptotically exact maximum likelihood inference of latent position vectors can be achieved using an efficient spectral method. Our method involves computing top eigenvectors of a normalized adjacency matrix and scaling eigenvectors using a regression step. The novel regression scaling step is an essential part of the proposed method. In simulations, we show that our proposed method is more accurate and more robust than common practices. We also show the effectiveness of our approach over standard real networks of the karate club and political blogs.

**Index Terms**—Latent space models, Stochastic block models, Maximum likelihood



## 1 INTRODUCTION

CLUSTERING over graphs is a classical problem with applications in systems biology, social sciences, and other fields [1], [2], [3], [4]. Although most formulations of the clustering problem are NP-hard [5], several approaches have yielded useful approximate algorithms. The most well-studied approach is spectral clustering. Most spectral methods are not based on a particular generative network model; alternative, model-based approaches have also been proposed, using loopy belief propagation [14], variational Bayes [15], Gibbs sampling [16], and semidefinite programming [22], [23].

Many spectral clustering methods are derived by proposing a discrete optimization problem, and relaxing it to obtain a continuous, convex optimization whose solution is given by the eigenvectors of a normalized adjacency matrix or Laplacian. A post-processing step, typically  $k$ -means, is used to extract clusters from these eigenvectors [6], [7]. Different matrix normalizations/transformations include Modularity [1], Laplacian [9], normalized Laplacian [10], [11], [26], and Bethe Hessian [12]. These methods are often characterized theoretically in the context of the Stochastic Block Model (SBM) [30], a simple and canonical model of community structure. Theoretical bounds on the detectability of network structure for large networks have been established for stochastic block models [14], [19], [20], [21]. Several spectral methods have been shown to achieve this recovery threshold [8], [12], [17]. Strong theoretical and empirical results have also been obtained using SDP-based [22], [23] and Belief Propagation-based [14] methods, which often have higher computational complexity than spectral methods. A threshold has also been discovered for perfect

clustering, i.e. when community structure can be recovered with zero errors [23], [24].

An alternative approach to the clustering problem is to invoke a latent space model. Each node is assigned a latent position  $v_i$ , and the edges of the graph are drawn independently with probability  $p_{ij} = g(v_i \cdot v_j)$  for some  $g(\cdot)$ . Hoff et al. [31] considered two latent space models, the first of which was a distance model,

$$P_{i,j} = l(-\|v_i - v_j\| - \mu + \beta x_{ij}), \quad l(x) = 1/(1 + \exp(-x))$$

where edge probabilities depend on the Euclidean distance between two nodes.  $x_{ij}$  is a fixed covariate term, which is not learned. Their second model is called a projection model:

$$P_{i,j} = l\left(\frac{v_i \cdot v_j}{\|v_j\|} + \beta x_{ij} - \mu\right). \quad (1)$$

Hoff et al. suggest to perform inference using an MCMC approach over both models. Focusing on the distance model, they have also extended their approach to allow the latent positions to themselves be drawn from a mixture distribution containing clusters [32], [33]. Efforts to improve the computational efficiency have been made in references [34], [35], and with a related mixed membership blockmodel in reference [15].

Young et al. [28] introduced the Random Dot Product Graph (RDPG), in which the probability of observing an edge between two nodes is the dot product between their respective latent position vectors. The RDPG model can be written as

$$P_{i,j} = v_i \cdot v_j.$$

This model is related to the projection model of Hoff et al. (1). The RDPG provides a useful perspective on spectral clustering, in two ways. First, it has led to theoretical advances, including a central limit theorem for eigenvectors of an adjacency matrix [29] and a justification for  $k$ -means

---

• *S. Feizi is the corresponding author. L. O'Connor and S. Feizi contributed equally to this work. L. O'Connor is with Harvard University. S. Feizi is with Stanford University. M. Médard is with Massachusetts Institute of Technology (MIT).*

clustering as a post processing step for spectral clustering [25]. Second, as a natural extension of the SBM, the RDPG describes more comprehensively the types of network structures, such as variable degrees and mixed community memberships, that can be inferred using spectral methods.

Let  $V$  be the  $n \times d$  matrix of latent positions, where  $n$  is the number of nodes and  $d$  is the dimension of the latent vectors ( $d \leq n$ ). Sussman et al. [25] proposed an inference method over the RDPG based on the heuristic that the first eigenvectors of  $A$  will approximate the singular vectors of  $V$ , as  $E(A) = VV^T$ . They also characterized the distribution of the eigenvectors of  $A$  given  $V$ . This heuristic can fail, however, as the first eigenvector of the adjacency matrix often separates high-degree nodes from low-degree nodes, rather than separating the communities. This problem occurs even in the simplest clustering setup: a symmetric SBM with two clusters of equal size and density. Therefore, we were motivated to develop a more robust inference approach.

In this paper, we consider a closely related latent space model, the *Logistic RDPG*, which uses a logistic link function mapping from latent positions to edge probabilities. Like the previously-studied RDPG, the logistic RDPG includes most SBMs as well as other types of network structure, including a variant of the degree corrected SBM. The logistic RDPG is also similar to the projection model, which uses a logistic link function but models a directed graph (with  $p_{i,j} \neq p_{j,i}$ ). Over this model, we show that the maximum likelihood latent-position inference problem admits an asymptotically exact spectral solution. Our method is to take the top eigenvectors of the mean-centered adjacency matrix and to scale them using a logistic regression step. This result is possible because over the logistic model specifically, the likelihood function separates into a linear term that depends on the observed network and a nonlinear penalty term that does not. Because of its simplicity, the penalty term admits a Frobenius norm approximation, leading to our spectral algorithm. A similar approximation is not obtained using other link functions besides the logistic link.

We show that the likelihood of the approximate solution approaches the maximum of the likelihood function when the graph is large and the latent-position magnitudes go to zero. The asymptotic regime is not overly restrictive, as it encompasses many large SBMs at or above the detectability threshold [14]. We compare the performance of our method in the graph clustering problem with spectral methods including the Modularity method [1], the Normalized Laplacian [9] and the Bethe Hessian [12], and the SDP-based methods [22], [23]. We show that our method outperforms these methods over a broad range of clustering models. We also show the effectiveness of our approach over real networks of karate club and political blogs.

## 2 LOGISTIC RANDOM DOT PRODUCT GRAPHS

In this section, we introduce the logistic RDPG, describe our inference method, and show that it is asymptotically equivalent to the maximum likelihood inference.

### 2.1 Definitions

Let  $\mathcal{A}$  be the set of adjacency matrices corresponding to undirected, unweighted graphs of size  $n$ .

**Definition 1 (Stochastic Block Model).** The Stochastic Block Model is a family of distributions on  $\mathcal{A}$  parameterized by  $(k, c, Q)$ , where  $k$  is the number of communities,  $c \in [k]^n$  is the community membership vector and  $Q \in [0, 1]^{k \times k}$  is the matrix of community relationships. For each pair of nodes, an edge is drawn independently with probability

$$P_{i,j} := Pr(A_{ij} = 1 | c_i, c_j, Q) = Q_{c_i, c_j}.$$

Another network model that characterizes low dimensional structures is the *Random Dot Product Graph* (RDPG). This class of models includes many SBMs.

**Definition 2 (Random Dot Product Graph).** The Random Dot Product Graph with link function  $g(\cdot)$  is a family of distributions on  $\mathcal{A}$  parameterized by an  $n \times d$  matrix of latent positions  $V \in \mathbb{R}^{n \times d}$ . For each pair of nodes, an edge is drawn independently with probability

$$P_{i,j} := Pr(A_{ij} = 1 | V) = g(v_i \cdot v_j), \quad (2)$$

where  $v_i$ , the  $i$ -th row of the matrix  $V$ , is the latent position vector assigned to node  $i$ .

The RDPG has been formulated using a general link function  $g(\cdot)$  [28]. The linear RDPG, using the identity link, has been analyzed in the literature because it leads to a spectral inference method. We will refer to this model as either the linear RDPG or as simply the RDPG. In this paper, we consider the Logistic RDPG:

**Definition 3 (Logistic RDPG).** The logistic RDPG is the RDPG with link function:

$$g(x) = l(x - \mu), \quad l(x) := \frac{1}{1 + e^{-x}}, \quad (3)$$

where  $\mu$  is the offset parameter of the logistic link function.

Note that this model is similar to the projection model of Hoff et al. [31] (1). The projection model is for a directed graph, with  $P_{i,j} \neq P_{j,i}$  owing to the division by  $\|v_j\|$ .

**Remark 1.** The parameter  $\mu$  in the logistic RDPG controls the sparsity of the network. If the latent position vector lengths are small, the density of the graph is

$$\frac{1}{n(n-1)} \sum_{i < j} \mathbb{E}(A_{ij}) \approx l(-\mu). \quad (4)$$

A logistic RDPG with this property is called *centered*. In Section 2.2, we show that asymptotically exact maximum likelihood inference of latent positions over the centered logistic RDPG can be performed using an efficient spectral algorithm.

For a general RDPG, the ML inference problem is:

**Definition 4 (ML inference problem for the RDPG).** Let  $A \in \mathcal{A}$ . The ML inference problem over the RDPG is:

$$\max_X \sum_{i \neq j} A_{ij} \log g(X_{i,j}) + (1 - A_{ij}) \log(1 - g(X_{i,j})), \quad (5)$$

$$X = VV^T, \quad V \in \mathbb{R}^{n \times d}.$$

Note that in the undirected case, this objective function is twice the log likelihood.

**Remark 2.** A convex semidefinite relaxation of Optimization (5) can be obtained for some link functions as follows:

$$\begin{aligned} \max_X \quad & \sum_{i \neq j} A_{ij} \log g(X_{ij}) + (1 - A_{ij}) \log(1 - g(X_{ij})), \\ & X \succeq 0, \end{aligned} \quad (6)$$

where  $X \succeq 0$  means that  $X$  is psd. For example, this optimization is convex for the logistic link function (i.e.,  $g(x) = l(x - \mu)$ ) and for the linear link function (i.e.,  $g(x) = x$ ). However, this optimization can be slow in practice, and it often leads to excessively high-rank solutions.

## 2.2 Maximum-Likelihood Inference of Latent Position Vectors

Here, we present an efficient, asymptotically exact spectral algorithm for the maximum-likelihood (ML) inference problem over the logistic RDPG, subject to mild constraints. We assume that the number of dimensions is given. In practice, these parameters are often set manually, but approaches have been proposed to automatically detect the number of dimensions [40].

The proposed maximum-likelihood inference of latent position vectors for the logistic RDPG is described in Algorithm 1. In the following, we sketch the derivation of this algorithm in a series of lemmas. Proofs for these assertions are presented in Section 7.

First, we simplify the likelihood function of Optimization (5) using the logistic link function. Let  $F(X)$  be the log-likelihood, and let the link function be  $g(x) = l(x - \mu)$ . Then:

$$\begin{aligned} F(X) &:= \sum_{i,j} A_{ij} \log l(X_{ij} - \mu) + (1 - A_{ij}) \log(1 - l(X_{ij} - \mu)) \\ &= \sum_{i,j} A_{ij} \log \frac{l(X_{ij} - \mu)}{1 - l(X_{ij} - \mu)} + \log(1 - l(X_{ij} - \mu)) \\ &= \sum_{i,j} A_{ij} (X_{ij} - \mu) + \log(1 - l(X_{ij} - \mu)). \end{aligned} \quad (7)$$

We have used that  $\log(l(x)/(1 - l(x))) = x$ . The maximum likelihood problem takes the following form (for given  $\mu$ ):

$$\begin{aligned} \max_X \quad & Tr(AX) + \sum_{i \neq j} \log\left(1 - \frac{1}{1 + e^{-(X_{ij} - \mu)}}\right), \\ & X = VV^T, \quad V \in \mathbb{R}^{n \times d}. \end{aligned} \quad (8)$$

The objective function has been split into a linear term that depends on the adjacency matrix  $A$  and a penalty term that does not depend on  $A$ . This simplification is what leads to tractable optimization, and it is the reason that the logistic link is needed; using e.g. a linear link, an optimization of this form is not obtained.

We define a penalty function  $f(x)$  that keeps only the quadratic and higher-order terms in the penalty term of (8). Let

$$f(x) := -(h(x) - h(0) - h'(0)x), \quad h(x) = \log(1 - l(x - \mu)).$$

Now,  $h'(0) = -l(-\mu)$ . Let

$$B := A - l(-\mu)\mathbf{1}_{n \times n}$$

in order to re-write Optimization (8) as:

$$\begin{aligned} \max_X \quad & Tr(BX) - \sum_{i \neq j} f(X_{ij}), \\ & X = VV^T, \quad V \in \mathbb{R}^{n \times d}. \end{aligned} \quad (9)$$

Note that for a centered RDPG with average density  $\bar{A}$ ,  $\mu = l^{-1}(\bar{A})$ , and  $B$  is the mean-centered adjacency matrix  $A - \bar{A}\mathbf{1}_{n \times n}$ . In the next step, we convert the penalty term in the objective function into a constraint:

**Lemma 1.** Suppose that  $X^*$  is the optimal solution to Optimization (5). Let

$$h^* := \frac{1}{n(n-1)} \sum_{i,j} f(X_{ij}^*).$$

Then  $X^*$  is also the solution to the following optimization:

$$\begin{aligned} \max_X \quad & Tr(BX) \\ & X = VV^T, \quad V \in \mathbb{R}^{n \times d} \\ & \frac{1}{n(n-1)} \sum_{i \neq j} f(X_{ij}) \leq h^*. \end{aligned} \quad (10)$$

In the following key lemma, we show that the inequality constraint of Optimization (10) can be replaced by its second order Taylor approximation.

**Lemma 2.** For any  $\epsilon > 0$  and  $\gamma \geq 1$ , there exists  $\delta > 0$  such that for any graph whose ML solution  $X^*$  satisfies

$$h^* \leq \delta \quad \text{and} \quad \max_i X_{ii}^* \leq \frac{\gamma}{n} \sum_i X_{ii}^*, \quad (11)$$

the following bound is satisfied. Let  $B$  be the mean centered adjacency matrix of the chosen graph. Let  $s^* \in \mathbb{R}$  be the optimal value of the following optimization, obtained at  $X = X^*$ :

$$\begin{aligned} \max_X \quad & Tr(BX), \\ & X = VV^T, \quad V \in \mathbb{R}^{n \times d} \\ & \frac{1}{n(n-1)} \sum_{i \neq j} f(X_{ij}) \leq h^*, \end{aligned} \quad (12)$$

Let  $\tilde{s}$  be the optimal value of the following optimization:

$$\begin{aligned} \max_X \quad & Tr(BX), \\ & X = VV^T, \quad V \in \mathbb{R}^{n \times d} \\ & \frac{a_2}{n(n-1)} \sum_{i \neq j} X_{ij}^2 \leq h^*. \end{aligned} \quad (13)$$

where  $a_2 := f''(-\mu)$ . Then

$$\tilde{s} \geq (1 - \epsilon)s^*.$$

The parameter  $h^*$  is related to the average length of latent-position vectors ( $X_{ii}^*$ ). If these lengths approach zero,  $h^*$  approaches zero, for a fixed  $\gamma$ . An implication of this constraint is that the logistic RDPG must be approximately centered. Thus, there is a natural choice for the parameter  $\mu$  for the purpose of inference:

$$\hat{\mu} = -l^{-1}\left(\frac{1}{n(n-1)} \|A\|_F\right). \quad (14)$$

---

**Algorithm 1** ML Inference for the logistic RDPG
 

---

**Require:** Adjacency matrix  $A$ , number of dimensions  $d$ . (optional) number of clusters  $k$

Form the mean-centered adjacency matrix  $B := A - 1/(n(n-1))\|A\|$ .

Compute  $d$  eigenvectors of  $B$  with largest eigenvalues:  $e_1, \dots, e_d$ .

Let  $X_i = e_i e_i^T$  for  $1 \leq i \leq d$ .

Perform logistic regression of the entries of  $A$  lying above the diagonal on the corresponding entries of  $X_1, \dots, X_d$ , estimating coefficients  $\lambda_1^*, \dots, \lambda_d^*$  subject to the constraint that  $\lambda_i \geq 0 \quad \forall i$ .

Let  $V$  be the matrix formed by concatenating  $\sqrt{\lambda_1^*} e_1, \dots, \sqrt{\lambda_d^*} e_d$ . Return  $V$ .

(optional) Perform  $k$ -means on  $V$ , and return the inferred clusters.

---

This estimator of  $\mu$  can be viewed as the maximum-likelihood estimator specifically over the centered logistic RDPG. With this choice of  $\mu$ ,  $B$ , the mean-centered adjacency matrix, can be written as

$$B = A - \frac{1}{n(n-1)}\|A\|_F.$$

Note that changing the constant in the inequality constraint of Optimization (13) only changes the scale of the solution, since the shape of the feasible set does not change. Thus, in this optimization we avoid needing to know  $h^*$  a priori (as long as the conditions of Lemma 2 are satisfied).

Next we show that that the solution to Optimization (13) can be recovered up to a linear transformation using spectral decomposition:

**Lemma 3.** Let  $\tilde{X}$  be the optimal solution to Optimization (13). Let  $e_1, \dots, e_d$  be the first  $d$  eigenvectors of  $B$ , corresponding to the largest eigenvalues. Then  $e_1, \dots, e_d$  are identical to the non-null eigenvectors of  $\tilde{X}$ , up to rotation.

Once the eigenvectors of  $\tilde{X}$  are known, it remains only to recover the corresponding eigenvalues. Instead of recovering the eigenvalues of  $\tilde{X}$ , we find the eigenvalues that maximize the likelihood, given the eigenvectors of  $\tilde{X}$ . Let  $\tilde{X}_i = e_i e_i^T$ . Then, the maximum-likelihood estimate of  $\lambda_1, \dots, \lambda_d$  conditional on  $X_1, \dots, X_d = \tilde{X}_1, \dots, \tilde{X}_d$  can be written as follows:

$$\lambda^* := \operatorname{argmax}_{\lambda=(\lambda_1, \dots, \lambda_d)} \sum_{i \neq j} \log P(A_{ij} | \tilde{X}_1, \dots, \tilde{X}_d, \lambda, \mu). \quad (15)$$

**Lemma 4.** Optimization (15) can be solved by logistic regression of the entries of  $A$  on the entries of  $X_1^*, \dots, X_d^*$ , with the constraint that the coefficients are nonnegative, and with intercept  $\mu$ .

These lemmas can be used to show the asymptotic optimality of Algorithm 1:

**Theorem 1.** For all  $\epsilon > 0$  and  $\gamma > 1$ , there exists  $\delta > 0$  that satisfies the following. For any graph with size  $n$  and adjacency matrix  $A$ , suppose that  $X^*$  is the solution to the optimization

$$\max_X P(A|X), \\ X = VV^T, \quad V \in \mathbb{R}^{n \times d}.$$

Let

$$h^* := \frac{1}{n(n-1)} \sum_{i \neq j} f(X_{ij}^*).$$

If

$$h^* < \delta \quad \text{and} \quad \max_i X_{ii}^* \leq \frac{\gamma}{n} \sum_i X_{ii}^*, \quad (16)$$

then

$$\frac{P(A|X = X_*)}{P(A|X = X^*)} > 1 - \epsilon,$$

where  $X_*$  is the solution obtained by Algorithm 1.

Our algorithm is asymptotically exact in the sense that the likelihood ratio between our solution and the true maximum converges uniformly to one as the average latent position length shrinks. Importantly, the convergence is uniform over arbitrarily large graphs; therefore, this regime contains most interesting large network models, such as an SBM with large communities that cannot be perfectly recovered. Coupling this algorithm with a  $k$ -means post processing step leads to a clustering method with robust performance under different network clustering setups.

This result is stronger than the statement that an approximate objective function approximates the likelihood function at the optimum of the likelihood function. Such a result, which can be obtained for many link functions (such as an affine-linear link), is not useful because it does not follow that the optimum of the approximate function lies near the optimum of the true likelihood function. Indeed, for a linear approximation, it has no optimum since the objective function is unbounded. In order to obtain the stronger statement that the likelihood at the optimum of the approximation is large, it is necessary to use a quadratic approximation. For link functions besides the logistic link, the quadratic term in the likelihood function depends on  $A$ , and a spectral optimization method cannot be obtained.

The condition in Theorem 1 that the lengths of optimal latent vectors are sufficiently small is not restrictive for large networks. Consider a sequence of increasingly-large SBMs with two clusters of fixed relative sizes, and a convergent sequence of admissible connectivity matrices whose average density is fixed. There are three asymptotic regimes for the community structure: (1) in which the structure of the network is too weak to detect any clusters at all; (2) in which the communities can be partially recovered, but some mis-assignments will be made; and (3) in which the communities can be recovered perfectly. The true latent position lengths go to zero in regimes (1) and (2) as well as in part of regime (3) [23]. Theorem 1 requires maximum-likelihood latent position lengths, rather than true position lengths, to go to zero. If this is the case, and if maximum likelihood achieves the optimum thresholds for partial recovery and perfect recovery, then our method will as well.

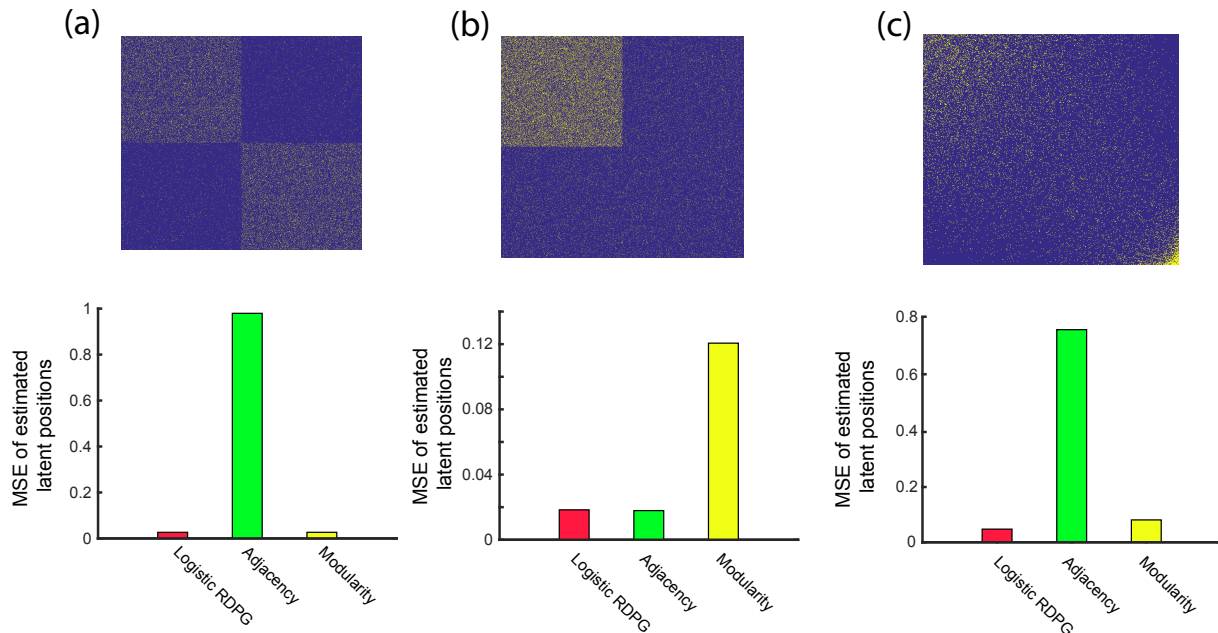


Fig. 1. Normalized mean squared error (one minus squared correlation) of inferred latent positions for two SBMs (a-b) and a non-SBM logistic RDPG (c). The top eigenvectors of the adjacency matrix  $A$  and the modularity matrix  $M$  do not characterize the community structure in panel (a) and in panel (b), respectively. Note that in practice, a different eigenvector could be selected or multiple eigenvectors could be used. In panel (c), the top eigenvector of  $A$  does not recover the latent structure. In contrast, our method successfully recovers the underlying latent position vectors in all cases.

### 3 PERFORMANCE EVALUATION OVER SYNTHETIC NETWORKS

In this section, we compare the performance of our proposed method (Algorithm 1) with existing methods. First, we assess the performance of our algorithm against existing methods in inference of latent position vectors of two standard SBMs depicted in Figure 1. The network demonstrated in panel (a) has two dense clusters. In this case, the first eigenvector of the modularity matrix  $M$  leads to a good estimation of the latent position vector while the first eigenvector of the adjacency matrix  $A$  fails to characterize this vector. This is because the first eigenvector of the adjacency matrix correlates with node degrees. The modularity transformation regresses out the degree component and recovers the community structure. However, the top eigenvector of the modularity matrix fails to identify the underlying latent position vector when there is a single dense cluster in the network, and the community structure *is* correlated with node degrees (Figure 1-b). This discrepancy highlights the sensitivity of existing heuristic inference methods in different network models (the Modularity method has not previously been considered a latent-position inference method, but we believe that its appropriate to do so). In contrast, our simple normalization allows the underlying latent position vectors to be accurately recovered in both cases. We also verified in panel (c) that our method successfully recovers latent positions for a non-SBM logistic RDPG. In this setup, the adjacency matrix’s first eigenvector again correlates with node degrees, and the modularity normalization causes an improvement. We found it remarkable that such a simple normalization (mean centering) enabled such significant improvements; using more sophisticated normalizations such

as the Normalized Laplacian and the Bethe Hessian, no improvements over were observed (data not shown).

Second, we assessed the ability of our method to detect communities generated from the SBM. We compared against the following existing spectral network clustering methods:

- Modularity (Newman, 2006). We take the first  $d$  eigenvectors of the modularity matrix  $M := A - v^T v / 2|\mathcal{E}|$ , where  $v$  is the vector of node degrees and  $|\mathcal{E}|$  is the number of edges in the network. We then perform  $k$ -means clustering on these eigenvectors.
- Normalized Laplacian (Chung, 1997). We take second- through  $(d + 1)$ st- last eigenvectors of  $L_{sym} := D^{-1/2}(D - A)D^{-1/2}$ , where  $D$  is the diagonal matrix of degrees. We then perform  $k$ -means clustering on these eigenvectors.
- Bethe Hessian (Saade et al., 2014). We take the second- through  $(d + 1)$ st- last eigenvectors of

$$H(r) := (r^2 - 1)\mathbf{1}_{n \times n} - rA + D,$$

where  $r^2$  is the density of the graph as defined in [12].

- Unnormalized spectral clustering (Sussman et al., 2012). We take the first  $d$  eigenvectors of the adjacency matrix  $A$ , and perform  $k$ -means clustering on these eigenvectors.
- Spectral clustering on the mean-centered matrix  $B$ . We take the first  $d$  eigenvectors of the matrix  $B$  and perform  $k$ -means on them, without a scaling step.

Note that in our evaluation we include spectral clustering on the mean-centered adjacency matrix  $B$  without subsequent eigenvalue scaling of Algorithm 1 to demonstrate that the scaling step computed by logistic regression is essential to

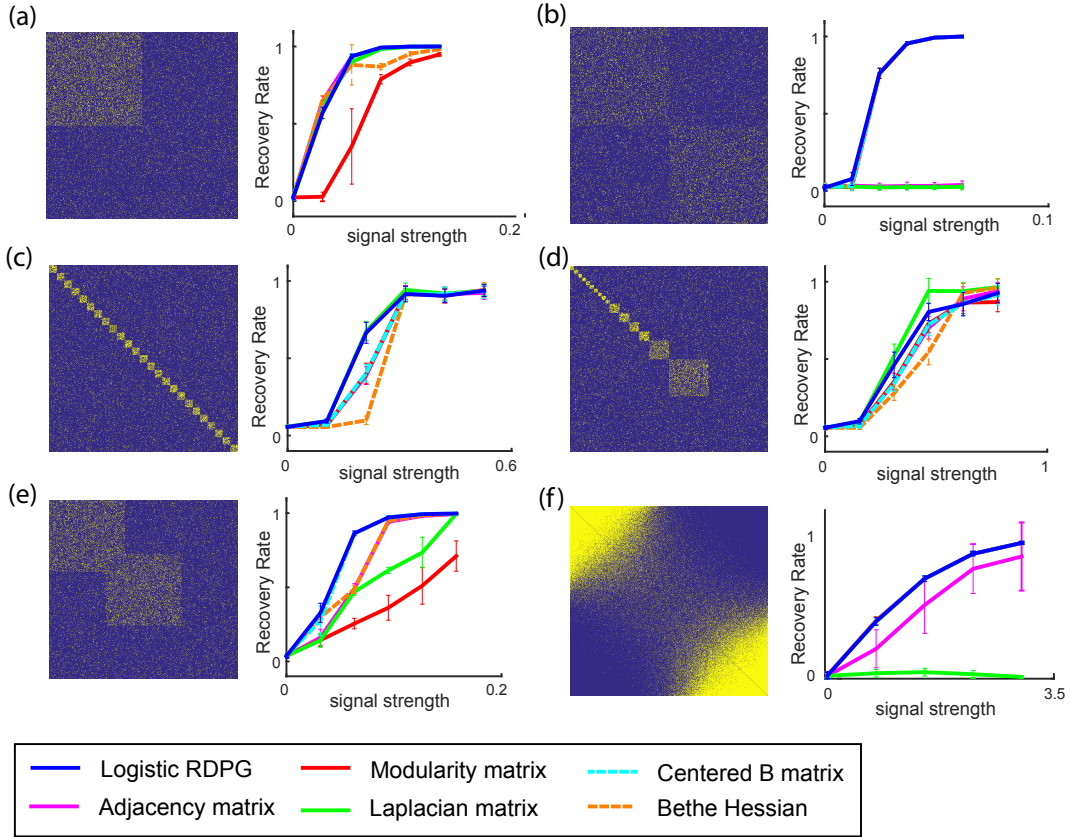


Fig. 2. Performance comparison of our method (logistic RDPG) against spectral clustering methods in different clustering setups. Panels (a)-(e) illustrate networks that can be characterized by SBM, while panel (f) illustrates a non-SBM network model. The scale of the  $x$  axis is different in panel (f) than the rest of the panels. Our proposed method performs consistently well, while other methods exhibit sensitive and inconsistent performance in different network clustering setups. Note that in some cases, such as for the Laplacian in panel (b), performance is improved by using a different eigenvector or by using a larger number of eigenvectors.

the performance of the proposed algorithm. When  $d = 1$ , the methods are equivalent. We also compare the performance of our method against two SDP-based approaches, the method proposed by Hajek et al. (2015) and the SDP-1 method proposed by Amini et al. (2014). For all methods we assume that the number of clusters  $k$  is given.

In our scoring metric, we distinguish between clusters and communities: For instance, in Figure 2-e, there are two clusters and four communities, comprised of nodes belonging only to cluster one, nodes belonging only to cluster two, nodes belonging to both clusters, and nodes belonging to neither. The score that we use is a normalized Jaccard index, defined as:

$$\frac{\max_{\sigma \in S_k} \sum_{l=1}^k \frac{|C_l \cap \hat{C}_{\sigma(l)}|}{|C_l|} - 1}{k - 1} \quad (17)$$

where  $C_l$  is the  $l$ -th community,  $\hat{C}_l$  is the  $l$ -th estimated community, and  $S_k$  is the group of permutations of  $k$  elements. Note that one advantage of using this scoring metric is that it weighs differently-sized clusters equally (it does not place higher weights on larger communities.).

Figure 2 presents a comparison between our proposed method and existing spectral methods in a wide range of clustering setups. Our proposed method performs consistently well, while other methods exhibit sensitive and

inconsistent performance in different network clustering setups. For instance, in the case of two large clusters (b) the second-to-last eigenvector of the Normalized Laplacian fails to correlate with the community structure; in the case of having one dense cluster (a), the Modularity normalization performs poorly; when there are many small clusters (c), the performance of the Bethe Hessian method is poor. In each case, the proposed method performs at least as well as the best alternate method, except in the case of several different-sized clusters (d), when the normalized Laplacian performs marginally better. In the case of overlapping clusters (e), our method performs significantly better than all competing methods. Spectral clustering on  $B$  without the scaling step also performs well in this setup; however, its performance is worse in panels (c-d) when  $d$  is larger, highlighting the importance of our logistic regression step.

The values of  $k$  and  $d$  for the different simulations were:  $k = 2, d = 1$ ;  $k = 2, d = 1$ ;  $k = 25, d = 24$ ;  $k = 17, d = 16$ ;  $k = 4, d = 2$ ;  $k = 2, d = 1$  for panels (a)-(f), respectively. The values of  $d$  are chosen based on the number of dimensions that would be informative to the community structure, if one knew the true latent positions. All networks have 1000 nodes, with background density 0.05.

While spectral methods are most prominent network clustering methods owing to their accuracy and efficiency, other approaches have been proposed, notably including

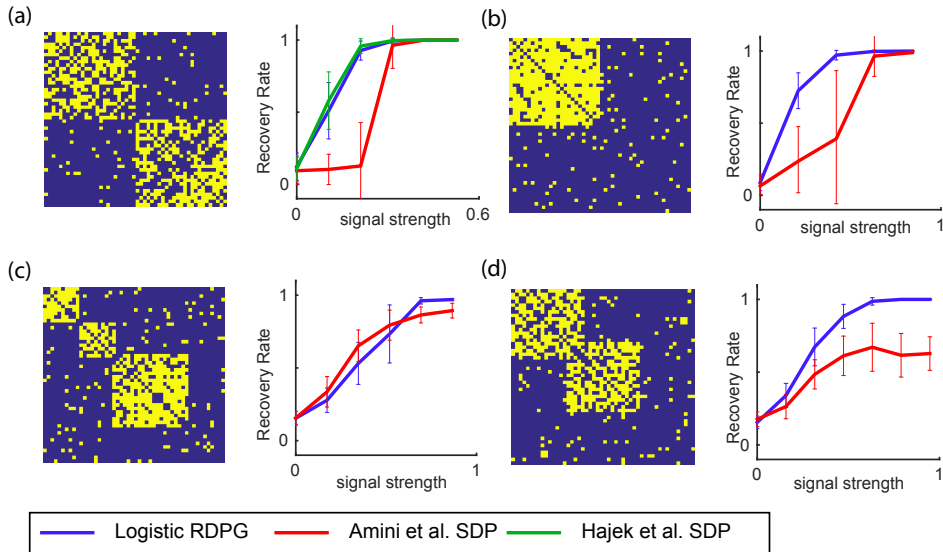


Fig. 3. Performance comparison of our method (logistic RDPG) against semidefinite programming-based clustering methods. Hajek et al.’s method is designed for the case of having two equally sizes partitions, thus it is not included in Panels b-d.

SDP-based methods, which solve relaxations of the maximum likelihood problem over the SBM. We compare the performance of our method with the SDP-based methods proposed by Hajek et al. (2015) and Amini et al. (2014) (Figure 3). In the symmetric SBM, meaning the SBM with two equally-dense, equally-large communities, we find that our method performs almost equally well as the method of Hajek et al. (2015), which is a simple semidefinite relaxation of the likelihood in that particular case. Our method also performs better than the method of Amini et al., which solves a more complicated relaxation of the SBM maximum-likelihood problem in the more general case (Figure 3).

#### 4 PERFORMANCE EVALUATION OVER REAL NETWORKS

To assess the performance of the Logistic RDPG over well-characterized real networks, we apply it to two well-known real networks. First, we consider the Karate Club network [44]. Originally a single karate club with social ties between various members, the club split into two clubs after a dispute between the instructor and the president. The network contains 34 nodes with the average degree 4.6, including two high degree nodes corresponding to the instructor and the president. Applying our method to this network, we find that the first eigenvector separates the two true clusters perfectly (Figure 4-a).

In the second experiment, we consider a network of political blogs, whose edges correspond to links between blogs [45]. This network, which contains 1221 nodes with nonzero degrees, is sparse (the average total degree is 27.4) with a number of high degree nodes (60 nodes with degrees larger than 100). The nodes in this network have been labeled as either *liberal* or *conservative*. We apply our method to this network. Figure 4-b shows inferred latent positions of nodes of this network. As it is illustrated in this figure, nodes with different labels have been separated in the latent space.

Note that some nodes are placed near the origin, indicating that they cannot be clustered confidently; this is occurred owing to their low degrees as the correlation between node degrees and distances from the origin was 0.95.

#### 5 CODE

We provide code for the proposed method in the following link: <https://github.com/SoheilFeizi/spectral-graph-clustering>

#### 6 DISCUSSION

In this paper, we developed a spectral inference method over logistic Random Dot Product Graphs (RDPGs), and we showed that the proposed method is asymptotically equivalent to the maximum likelihood latent-position inference. Previous justifications for spectral clustering have usually been either consistency results [25], [26] or partial-recovery results [17], [18]; to the best of our knowledge, our likelihood-based justification is the first of its kind for a spectral method. This type of justification is satisfying because maximum likelihood inference methods can generally be expected to have optimal asymptotic performance characteristics; for example, it is known that maximum likelihood estimators are consistent over the SBM [38], [39]. It remains an important future direction to characterize the asymptotic performance of the MLE over the Logistic RDPG.

We have focused in this paper on the network clustering problem; however, latent space models such as the Logistic RDPG can be viewed as a more general tool for exploring and analyzing network structures. They can be used for visualization [41], [42] and for inference of partial-membership type structures, similar to the mixed-membership stochastic blockmodel [15]. Our approach can also be generalized to multi-edge graphs, in which the number of edges between two nodes is binomially distributed. Such data is emerging in areas including systems biology, in the form of cell type and tissue specific networks [43].



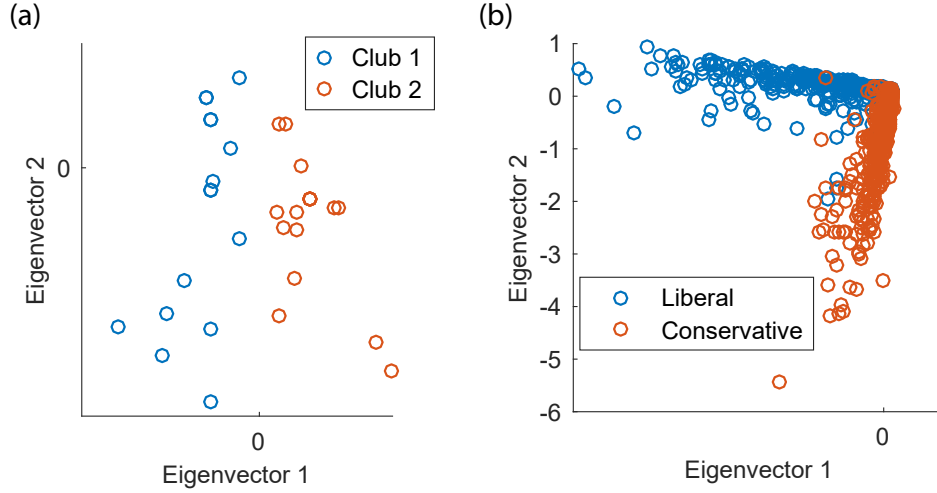


Fig. 4. Estimated latent positions for nodes of two real networks. (a) Karate club social network. (b) Political blogs network.

## 7 PROOFS

**Proof 1 (Proof of Lemma 1).** If not, then the optimal solution to Optimization (10) would be a better solution to Optimization (5) than  $X^*$ .

**Proof 2 (Proof of Lemma 2).** The origin is in the feasible set for both optimizations. For each optimization, the objective function value satisfies  $Tr(rCX) = rTr(CX)$ . Thus, the optimum is either at the origin (if there is no positive solution) or at the boundary of the feasible set. If the optimum is at the origin, we have  $s^* = \tilde{s} = 0$ . If not, let  $X$  be any solution to  $\frac{1}{n^2} \sum_{i,j} f(X_{ij}) = h$ . Let  $r = \|X\|_F$ , and let  $r' = \sqrt{h/a_2}$ . Claim: fixing  $\gamma$ ,  $r/r' \rightarrow 1$  uniformly as  $h^* \rightarrow 0$ .

Define

$$F_X(a) := \frac{1}{n^2} \sum_{i,j} f(aX_{ij}/\|X\|_F)$$

for  $a > 0$ . In addition, since  $r'$  has been defined such that the quadratic term of  $F_X(r')$  is  $a_2 \sum_{i,j} (\frac{r'}{r} X_{ij})^2 = h^*$ , we have

$$F_X(r') = h^* + \frac{1}{n^2} O\left(\sum_{i,j} \left(\frac{r'}{r} X_{ij}\right)^3\right). \quad (18)$$

Moreover, the Taylor series for  $f(\cdot)$  converges in a neighborhood of zero. Because of the constraint

$$\max_{i,j} X_{ij}^* = \max_i X_{ii}^* \leq \frac{\gamma}{n} \sum_i X_{ii}^*$$

we can choose  $\delta$  such that every entry  $X_{ij}$  falls within this neighborhood. This constraint also implies

$$\frac{1}{n^2} \sum X_{ij}^3 \leq \frac{1}{n^2} \sum |X_{ij}|^3 = O\left(\left(\frac{1}{n^2} \|X\|_F\right)^3\right).$$

Substituting this into (18), we have

$$F_X(r') = h^* + \frac{1}{n^2} O(r'^3). \quad (19)$$

Therefore, we have

$$\frac{|F_X(r) - F_X(r')|}{F_X(r')} = \frac{|h^* + O(r'^3) - h^*|}{h^*} = O(r'). \quad (20)$$

Note that  $f(\cdot)$  is convex function with  $f'(x) > 0$  for all  $x > 0$  and  $f'(x) < 0$  for all  $x < 0$ . Thus  $F_X$  is increasing, convex, and zero-valued at the origin: for any  $a, b > 0$ ,

$$\frac{|a - b|}{b} < \frac{|F_X(a) - F_X(b)|}{F_X(b)}. \quad (21)$$

Thus  $\frac{|r-r'|}{r'} = O(r')$  and  $\frac{r}{r'} = 1 + O(r')$ .

Let  $r_s$  be the norm of the arg max to optimization (12); because the objective function is linear we have that  $\tilde{s} \geq \frac{r'}{r_s} s^*$ . Let  $r_t$  be the distance to the intersection of the boundary of the feasible set with the ray from the origin through the arg max to optimization (13); then  $s^* \geq \frac{r_t}{r} \tilde{s}$ . We have shown that both ratios tend uniformly to one. This completes the proof.

**Proof 3 (Proof of Lemma 3).** First, suppose we have prior knowledge of eigenvalues of  $X^*$ . Denote its nonzero eigenvalues by  $\lambda_1^*, \dots, \lambda_d^*$ . Then we would be able to recover the optimal solution to Optimization (13) by solving the following optimization

$$\begin{aligned} \max_X \quad & Tr(BX) \\ & \lambda_i = \lambda_i^* \quad 1 \leq i \leq d \\ & rank(X) = d \end{aligned} \quad (22)$$

Note that the Frobenius norm of a matrix is determined by eigenvalues of the matrix as follows:

$$\|X\|_F^2 = Tr(XX^T) = Tr(X^2) = \sum \lambda_i^2. \quad (23)$$

Thus we can drop the Frobenius norm constraint in (13). Let  $X$  be an  $n \times n$  psd matrix, whose non-null eigenvectors are the columns of a matrix  $E \in \mathbb{R}^{n \times d}$ , and whose respective eigenvalues are  $\lambda_1, \dots, \lambda_d$ . Let  $V := E \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$ , so that  $X = VV^T$ . Rewrite the objective function as

$$Tr(BX) = Tr(V^T B V) = \sum_{i=1}^d \lambda_i e_i^T B e_i.$$

Therefore  $\tilde{X} = EE^T$  and  $X^* = VV^T$ .

**Proof 4 (Proof of Lemma 4).** The upper-triangular entries of  $A$  are independent Bernoulli random variables conditional on  $X$  and  $\mu$ , with a logistic link function. The coefficients should be nonnegative, as  $X$  is constrained to be positive semidefinite.

**Proof 5 (Proof of Theorem 1).** By Lemma 1, we have that the solution to optimization (10) is equal to the log-likelihood, up to addition of a constant. By Lemma 2, we have that for a fixed  $\gamma$ , as  $h^* \rightarrow 0$ , the quotient  $s^*/\tilde{s}$  converges uniformly to one, where  $s^*$  is the solution to (10) and  $\tilde{s}$  is the solution to optimization (13). The convergence is uniform over the choice of  $B$  that is needed for Theorem 1. Because  $s^*$  and  $\tilde{s}$  do not diverge to  $\pm\infty$ , this also implies that  $s^* - \tilde{s}$ , and therefore the log-likelihood ratio, converges uniformly to zero. By Lemma 3, the non-null eigenvectors of the arg max of optimization (13) are equivalent (up to rotation) to the first eigenvectors of  $B$ . Finally, by Lemma 4, the eigenvalues that maximize the likelihood can be recovered using a logistic regression step. By Lemma 2, the theorem would hold if we recovered the eigenvalues solving the approximate optimization (13). By finding the eigenvalues that exactly maximize the likelihood, we achieve a likelihood value at least as large.

## REFERENCES

- [1] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. condmat/0112110, pp. 8271–8276, 2001.
- [2] S. Butenko, *Clustering challenges in biological networks*. World Scientific, 2009.
- [3] N. Mishra, R. Schreiber, I. Stanton, and R. E. Tarjan, "Clustering social networks," in *Algorithms and Models for the Web-Graph*. Springer, 2007, pp. 56–67.
- [4] C.-H. Lee, M. N. Hoehn-Weiss, and S. Karim, "Grouping interdependent tasks: Using spectral graph partitioning to study system modularity and performance," *Available at SSRN*, 2014.
- [5] S. E. Schaeffer, "On the np-completeness of some graph cluster measures," *arXiv preprint cs/0506100*, 2008.
- [6] Ng, Andrew Y., Michael I. Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm." *Advances in neural information processing systems 2* (2002): 849-856.
- [7] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [8] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical review E*, vol. 74, no. 3, 2006.
- [9] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, "The laplacian spectrum of graphs," *Graph theory, combinatorics, and applications*, vol. 2, pp. 871–898, 1991.
- [10] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.
- [11] Shi, Jianbo, and Jitendra Malik. "Normalized cuts and image segmentation." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.8 (2000): 888-905.
- [12] A. Saade, F. Krzakala, and L. Zdeborová, "Spectral clustering of graphs with the bethe hessian," in *Advances in Neural Information Processing Systems*, 2014, pp. 406–414.
- [13] Hagen, Lars, and Andrew B. Kahng. "New spectral methods for ratio cut partitioning and clustering." *Computer-aided design of integrated circuits and systems, iee transactions on* 11.9 (1992): 1074-1085.
- [14] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physical Review E*, vol. 84, no. 6, p. 066106, 2011.
- [15] Airolidi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008), "Mixed Membership Stochastic Blockmodels", *The Journal of Machine Learning Research*, 9, 1981-2014.
- [16] Snijders, T., and Nowicki, K. (1997), "Estimation and Prediction for Stochastic Blockmodels for Graphs With Latent Block Structure", *Journal of Classification*, 14, 75-100.
- [17] Krzakala, Florent, Christopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborov, and Pan Zhang. "Spectral redemption in clustering sparse networks." *Proceedings of the National Academy of Sciences* 110, no. 52 (2013): 20935-20940.
- [18] Nadakuditi, Raj Rao, and Mark EJ Newman. "Graph spectra and the detectability of community structure in networks." *Physical review letters* 108, no. 18 (2012): 188701.
- [19] Mossel, Elchanan, Joe Neeman, and Allan Sly. "Stochastic block models and reconstruction." *arXiv preprint arXiv:1202.1499* (2012).
- [20] Massouli, Laurent. "Community detection thresholds and the weak Ramanujan property." *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. ACM, 2014.
- [21] Mossel, Elchanan, Joe Neeman, and Allan Sly. "A proof of the block model threshold conjecture." *arXiv preprint arXiv:1311.4115* (2013).
- [22] A. A. Amini and E. Levina, "On semidefinite relaxations for the block model," *arXiv preprint arXiv:1406.5647*, 2014.
- [23] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming: Extensions," *arXiv preprint arXiv:1502.07738*, 2015.
- [24] Abbe, Emmanuel, Afonso S. Bandeira, and Georgina Hall. "Exact recovery in the stochastic block model." *arXiv preprint arXiv:1405.3267* (2014).
- [25] D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe, "A consistent adjacency spectral embedding for stochastic blockmodel graphs," *Journal of the American Statistical Association*, vol. 107, no. 499, pp. 1119–1128, 2012.
- [26] Rohe, Karl, Sourav Chatterjee, and Bin Yu. "Spectral clustering and the high-dimensional stochastic blockmodel." *The Annals of Statistics* (2011): 1878-1915.
- [27] Kraetzl, Miro, Christine Nickel, and Edward R. Scheinerman. "Random dot product graphs: A model for social networks". Preliminary Manuscript, 2005.
- [28] S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," in *Algorithms and models for the web-graph*. Springer, 2007, pp. 138–149.
- [29] A. Athreya, V. Lyzinski, D. J. Marchette, C. E. Priebe, D. L. Sussman, and M. Tang, "A central limit theorem for scaled eigenvectors of random dot product graphs," *arXiv preprint arXiv:1305.7388*, 2013.
- [30] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.
- [31] Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. "Latent space approaches to social network analysis." *Journal of the American Statistical Association* 97.460 (2002): 1090-1098.
- [32] Shortreed, Susan, Mark S. Handcock, and Peter Hoff. "Positional estimation within a latent space model for networks." *Methodology* 2.1 (2006): 24-33.
- [33] Handcock, Mark S., Adrian E. Raftery, and Jeremy M. Tantrum. "Model-based clustering for social networks." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170.2 (2007): 301-354.
- [34] Salter-Townshend, Michael, and Thomas Brendan Murphy. "Variational Bayesian inference for the latent position cluster model." *Analyzing Networks and Learning with Graphs Workshop at 23rd annual conference on Neural Information Processing Systems (NIPS 2009)*, Whister, December 11 2009. 2009.
- [35] Friel, Nial, Caitriona Ryan, and Jason Wyse. "Bayesian model selection for the latent position cluster model for Social Networks." *arXiv preprint arXiv:1308.4871* (2013).
- [36] T. Qin and K. Rohe, "Regularized spectral clustering under the degree-corrected stochastic blockmodel," in *Advances in Neural Information Processing Systems*, 2013, pp. 3120–3128.
- [37] S. Van Dongen and A.J. Enright, "Metric distances derived from cosine similarity and pearson and spearman correlations," *arXiv preprint arXiv:1208.3145*, 2012.
- [38] Bickel, Peter, David Choi, Xiangyu Chang, and Hai Zhang. "Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels." *The Annals of Statistics* 41, no. 4 (2013): 1922-1943.
- [39] Celisse, Alain, Jean-Jacques Daudin, and Laurent Pierre. "Consistency of maximum-likelihood and variational estimators in the stochastic block model." *Electronic Journal of Statistics* 6 (2012): 1847-1899.

- [40] Zelnik-Manor, Lihi, and Pietro Perona. "Self-tuning spectral clustering." In *Advances in neural information processing systems*, pp. 1601-1608. 2004.
- [41] Hall, Kenneth M. "An r-dimensional quadratic placement algorithm." *Management science* 17, no. 3 (1970): 219-229.
- [42] Koren, Yehuda. "Drawing graphs by eigenvectors: theory and practice." *Computers & Mathematics with Applications* 49.11 (2005): 1867-1888.
- [43] Neph, S. et al. "Circuitry and dynamics of human transcription factor regulatory networks". *Cell* 150, 1274-1286 (2012).
- [44] Zachary, W. "An information flow model for conflict and fission in small groups". *Journal of anthropological research*, 33(4):452-473, (1977).
- [45] L. A. Adamic and N. Glance. "The political blogosphere and the 2004 us election: divided they blog". In *Proceedings of the 3rd international workshop on Link discovery*, page 36. ACM, (2005).