EURECOM
Department of X
Campus SophiaTech
CS 50193
06904 Sophia Antipolis cedex
FRANCE

Research Report RR-16-315

Characterization of L^1-norm Statistic for Anomaly Detection in Erdős Rényi Graphs

March 16^{th}, 2016
Last update March 16^{th}, 2016

Arun Kadavankandy, Laura Cottatellucci, and Konstantin Avrachenkov

Email : arun.kadavankandy@inria.fr, laura.cottatellucci@eurecom.fr, konstantin.avrachenkov@inria.fr

[1]EURECOM's research is partially supported by its industrial members: BMW Group Research and Technology, IABG, Monaco Telecom, Orange, Principaut de Monaco, SAP, ST Microelectronics, Symantec.

Characterization of L^1-norm Statistic for Anomaly Detection in Erdős Rényi Graphs

Arun Kadavankandy, Laura Cottatellucci, and Konstantin Avrachenkov

Abstract

We devise statistical tests to detect the presence of an embedded Erdős-Rényi (ER) subgraph inside a random graph, which is also an ER graph. We make use of properties of the asymptotic distribution of eigenvectors of random graphs to detect the subgraph. This problem is related to the planted clique problem that is of considerable interest.

Index Terms

Subgraph detection, Erdos-Renyi, Detection and Estimation

Contents

1 Introduction and Notation **1**

2 The subgraph detection problem and Problem Statement **3**

3 Algorithm and Analysis **4**

3.1 Spectral statistics under $\mathcal{H}_0$ 4

3.2 Eigenvalue and eigenvector properties under $\mathcal{H}_1$ 7

3.2.1 Eigenvector distribution under $\mathcal{H}_1$ 9

3.2.2 Distribution of χ under $\mathcal{H}_1$ 17

4 Simulations **18**

5 Conclusions and Future Work **19**

List of Figures

1 CDF of χ under $\mathcal{H}_0$ 19
2 CDF of χ under $\mathcal{H}_1$. 19

1 Introduction and Notation

We study the problem of deciding whether a given realization of a random graph contains an extraneous denser subgraph embedded within it. This falls in the general framework of graph anomaly detection (see the works [1,2] and references therein). Specifically, we consider a special case of the above problem where the random graph is an Erdős Rényi (ER) graph and the embedded subgraph is also an ER graph with a larger density of edges. We note here that this problem is different from classifying the nodes as is done in the works on subgraph detection [3, 4], since we address the problem of detecting the presence of an embedded subgraph; we do not attempt to locate nodes of the subgraph in the given graph. We also mention here the related problem of community detection where the community sizes usually scale linearly with respect to the graph size, and the density of edges in each community is larger than the intercommunity edge density [5, 6]. In our work we consider a graph with a single small community embedded within, whose size scales much slower than linearly with the size of the graph. A special case of the subgraph detection problem is the clique detection problem as considered in [4, 7, 8].

Our work is based on the fact that when there is no embedded subgraph, the modularity matrix of the random graph is a symmetric matrix with independent upper triangular entries with zero mean. The eigenvectors of such a matrix have been shown to be approximately Haar distributed [9, 10], under certain conditions on the moments of the entries. This means that a typical eigenvector of the modularity matrix is delocalized, meaning its L^1-norm is large. Note that the L^1-norm of a unit vector $\mathbf{v}$ satisfies $1 \leq \|\mathbf{v}\|_1 \leq \sqrt{n}$, where the upper bound corresponds to the case of complete delocalization, i.e., all the entries of the vector are of the same order of magnitude, and the lower bound corresponds to the completely localized case, i.e., only one entry is non-zero. On the other hand, when there is a subgraph embedded onto the random graph, we hypothesize that there will exist an eigenvector that is "localized", i.e., a fraction of components possess most of the mass of the eigenvector. This idea has been used in the literature to do community detection based on k-means clustering of the dominant eigenvectors [11], [12]. Delocalization properties of eigenvectors of random matrices under a variety of distributions have been studied recently in a series of works [13–15].

Anomaly detection based on norms has been studied empirically in [1, 2]. There the authors look for the presence of an eigenvector whose L^1-norm is much smaller than a fixed threshold that depends on the mean and variance of the L^1-norms of all the eigenvectors of the modulariy matrix estimated empirically, and declare a subgraph to be present if there exists such an eigenvector. In our work we provide theoretical validation for anomaly detection based on the L^1-norm of only the dominant eigenvector, and show that it is possible to detect the anomaly in this way. We find the distributions of the test statistic with and without the embedded subgraph for a specific setting where both the subgraph and the background graph are independent ER random graphs.

Our contribution is threefold. We derive the distribution of the dominant eigenvector components of the modularity matrix[1] when there is an embedded subgraph. We use this result to derive the asymptotic distribution of the L^1-norm of this eigenvector. We also look at the case where there is no subgraph embedded and use the properties of the eigenvectors of Wigner matrices as explored in [9, 16], to derive the L^1-norm of the eigenvectors when there is no subgraph embedded. Using these distributions we then devise a statistical test to detect the presence of the extraneous subgraph.

In the following we present relevant notational conventions followed throughout the paper.

Notation:

A vector is denoted in bold lower case ($\mathbf{x}$), a matrix in bold upper case ($\mathbf{A}$), and their components as x_i and A_{ij}. Also, $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots x_n^2}$, is the L^2-norm of $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \ldots |x_n|$ is its L^1-norm. For a real symmetric matrix $\mathbf{A}$, $\|\mathbf{A}\|$ denotes its spectral radius, i.e., the maximum eigenvalue in absolute value. We denote the standard Euclidean basis vectors as $\mathbf{e}_i$, a unit vector with all zero components except the i^{th} component, which is equal to 1, and $\mathbf{1}_n \in \mathbb{R}^n$ denotes an $n \times 1$ vector whose components are all equal to 1. Also, $\mathbf{J}_n$ denotes an $n \times n$ matrix whose entries are all equal to 1, i.e., $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^T$. We do not distinguish between a random variable and its realization and this is usually clear from the context.

Also note the conventional asymptotic notations: $f(n) = \mathrm{O}(g(n)), f(n) = \mathrm{o}(g(n)), f(n) = \Omega(g(n))$ denote respectively that $\lim_{n\to\infty} \frac{f(n)}{g(n)} \leq K$, $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0, \lim_{n\to\infty} \frac{g(n)}{f(n)} \leq K$, for any two functions $f(n)$ and $g(n)$ of n. Also $f(n) = \Theta(g(n))$, if $f(n) = \Omega(g(n))$, and $f(n) = \mathrm{O}(g(n))$. We also use the notation $o_p(1)$ and $O_p(1)$, to denote random variables that vanish in probability and are bounded in probability, respectively. A sequence random variables $x_n = \mathrm{o}_p(1)$, if $x_n \xrightarrow{p} 0$. Generic constants independent of n may be denoted K, k, C, c and are arbitrary and may change from line to line. The symbol $\sim$ denotes "has the distribution" for a random variable. The abbreviation $w.p.$ denotes " with probability". Probabilistic operators such as distributions and expectations are given subscripts to specify the hypothesis under which they hold; for example $\mathbb{E}_{\mathcal{H}_1}$ denotes expectation w.r.t the distribution under hypothesis $\mathcal{H}_1$. We use the common notation $\mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ to denote the multivariate normal distribution in $\mathbb{R}^n$ with mean vector $\boldsymbol{\mu}_n$, and covariance matrix $\boldsymbol{\Sigma}_n$. We sometimes use the notation Var for the vairance operator of a random variable, $\mathbb{E}$ denotes expectation and $\mathbb{P}$ denotes probability where the space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ is implicit.

In section 2 we formulate the detection problem, first in general terms; and then in the more specific case studied in this paper. In section 3, we present our anomaly detection algorithm, which is a hypothesis test problem with the probability of false alarm fixed. In section 3.1, we describe the spectral properties of the

[1] By Modularity matrix we mean the adjacency matrix of the graph from which we subtract the edge probability of the background graph.

modularity matrix $\boldsymbol{\mathcal{A}}$ under $\mathcal{H}_0$, and characterize the distribution of the L^1-norm of its eigenvectors. Proposition 1 gives the main result on the asymptotic distribution of χ under $\mathcal{H}_0$. In section 3.2 we analyze the spectral properties under $\mathcal{H}_1$, and in Theorem 2, derive a Central Limit Theorem (CLT) for the individual components of the dominant eigenvector of $\boldsymbol{\mathcal{A}}$. Using this distribution we compute the approximate asymptotic distribution of the L^1-norm statistic under $\mathcal{H}_1$ in section 3.2.2. Finally in section 5 we describe our conclusions and directions for future research.

2 The subgraph detection problem and Problem Statement

In this section we formulate the general problem of subgraph detection and later describe the specific problem we want to analyze. Let $G = (V, E)$ denote the observed graph, where V is the set of vertices, with cardinality $|V| = n$, and $E \subset V \times V$ is the set of edges. When there is no embedded subgraph, $G = G_b$, where $G_b = (V, E_b)$ is the background graph with E_b used to denote the edge set of the background graph. Let us denote the subgraph by $G_s = (V_s, E_s)$ with $V_s \subset V$, and $|V_s| = m$. When there is an embedded subgraph we have $E = E_b \cup E_s$. We desire to perform the following detection problem based on an observation of the graph G,

$$\mathcal{H}_0 : E = E_b \tag{1}$$
$$\mathcal{H}_1 : E = E_b \cup E_s. \tag{2}$$

In other words, the null hypothesis $\mathcal{H}_0$ corresponds to the case when there is no embedded subgraph, and all the edges of the observed graph belong to the background graph, and the hypothesis $\mathcal{H}_1$ corresponds to the case where the edges of the observed graph belong to either the background graph or the subgraph.

In this work we focus on a specific case of the above problem where both the background graph and the embedded subgraph are independently drawn from an ER graph ensemble. For simplicity of mathematics we allow self-loops, but in general this does not impact the results to a large extent. We assume $G_b = \mathcal{G}(n, p_b)$, and $G_s = \mathcal{G}(m, p_s)$, where $\mathcal{G}(l, q)$ denotes the class of ER random graphs of size l and edge probability q. Under $\mathcal{H}_1$, the probability of two nodes within V_s being connected in G is therefore $p_1 = 1 - (1 - p_b)(1 - p_s) = p_b + p_s - p_b p_s$ and elsewhere the edge probability is p_b. Under $\mathcal{H}_0$, the edge probability is uniformly p_b. Without loss of generality we assume that $V_s = \{1, 2, \dots m\}$.

It can be observed under $\mathcal{H}_1$ the graph is probabilistically equivalent to a Stochastic Block Model (SBM) with two communities of size m and $n-m$, within community link probabilities $p_1 = p_b + p_s - p_b p_s$ and $p_2 = p_b$; and outlink probability $p_0 = p_b$. Properties of SBM have been studied extensively in several works in the literature under assumption of linearly increasing block sizes; see e.g. [17, 18].

The adjacency matrix $\mathbf{A}$ of G is given as below

$$A_{ij} = A_{ji} \sim \begin{cases} \mathcal{B}(p_a) & \text{if } i, j \leq m \\ \mathcal{B}(p_b) & \text{otherwise} \end{cases} \tag{3}$$

where $\mathcal{B}(p)$ denotes the Bernoulli distribution that is 1 with probability p; $p_a = p_1$ under $\mathcal{H}_1$ and p_b under $\mathcal{H}_0$. Notice that p_b, p_s and m in general scale with the graph size n; the constraints on the actual scaling with respect to n will be made explicit when the results are given. We also define $\boldsymbol{\mathcal{A}} = \mathbf{A} - p_b\mathbf{J}_n$. Since we are considering undirected graphs, $\mathbf{A}$ is symmetric with independent upper diagonal entries and the same holds for $\boldsymbol{\mathcal{A}}$. Being a symmetric matrix it admits a spectral decomposition such that $\boldsymbol{\mathcal{A}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathbf{T}}$, where $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n\end{bmatrix}$, is an orthonormal matrix whose columns are made of the normalized eigenvectors with respective eigenvalues $\Lambda_{ii} = \lambda_i$, in decreasing order without loss of generality (wlog), $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

3 Algorithm and Analysis

In what follows we focus on the following algorithm. It is similar to the algorithm introduced in [2] based on finding the eigenvector of $\boldsymbol{\mathcal{A}}$ with the least L^1 norm.

Algorithm: Subgraph Detection

- *Input*: Adjacency matrix $\mathbf{A}$, background probability $p_b, \mu_{(0)}$, the mean of χ under $\mathcal{H}_0$ and $\sigma^2_{(0)}$, its variance under $\mathcal{H}_0$. Fix probability of false alarm p_{FA}.
- Construct the matrix $\boldsymbol{\mathcal{A}} = \mathbf{A} - p_b\mathbf{J}$
- Compute the eigenvector $\mathbf{u}_1$ corresponding to eigenvalue λ_1, and find $\chi = \|\mathbf{u}_1\|_1$.
- Find τ, such that (s.t.) $\mathbb{P}_{\mathcal{H}_0}\{\chi < \tau\} = p_{FA}$, i.e., $\tau = \mu_{(0)} + \sigma_{(0)}\Phi^{-1}(p_{FA})$
- If $\chi < \tau$, declare $\mathcal{H}_1$, otherwise $\mathcal{H}_0$,

where Φ is the Cumulative Density Function (CDF) of $\mathcal{N}(0, 1)$.

3.1 Spectral statistics under $\mathcal{H}_0$

Under $\mathcal{H}_0$, $\boldsymbol{\mathcal{A}}$ is a symmetric matrix with independent centered upper triangular entries as given below

$$\mathcal{A}_{ij} = \mathcal{A}_{ji} = \begin{cases} 1 - p_b & \text{w.p. } p_b \\ -p_b & \text{w.p. } 1 - p_b \end{cases}$$

i.e., the components of $\boldsymbol{\mathcal{A}}$ are independent on and above the diagonal, with zero mean, and variance $p_b(1-p_b)$. Thus the matrix $\boldsymbol{\mathcal{A}}$ under $\mathcal{H}_0$ is a standard Wigner matrix. Its spectral properties such as the empirical spectral distribution and the spectral radius are well-studied in the literature under different scaling laws on p_b, see e.g., [18, 19]. The eigenvectors of Wigner matrices are approximately Haar-distributed on the space of unitary matrices on $\mathbb{R}^{n\times n}$ as suggested by partial results on universality of eigenvector statistics [9, 10]. Therefore, a typical eigenvector $\mathbf{u}_i$ is approximately uniformly distributed on the hypersphere, $\mathbf{S}^{n-1} = \{\mathbf{s} : \|\mathbf{s}\| = 1\}$, in the L^2- (euclidean) space. A unit vector on the hypersphere can be modelled as a Gaussian eigenvector normalized to have unit L^2- norm, i.e., $\frac{\mathbf{x}}{\|\mathbf{x}\|}$, with $\mathbf{x}$ being a $\mathbb{R}^n$ Gaussian vector with covariance matrix $\mathbf{I}$, i.e., $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I})$. We assume the following fact, which is a widely held conjecture about the asymptotic distribution of the eigenvectors of a Wigner matrix. This holds exacly for Wigner matrices with gaussian entries such as the Gaussian Unitary ensemble and the Gaussian Orthogonal Ensemble []anderson2009zeitouni.

Observation 1 *(Haar distribution of Eigenvectors of a Wigner matrix) A typical eigenvector $\mathbf{u}_i$ of $\boldsymbol{\mathcal{A}}$ under hypothesis $\mathcal{H}_0$ is distributed uniformly on the hypersphere on $S^{(n-1)}$. The distribution of a typical eigenvector $\mathbf{u}_i$ is identical to the distribution of $\mathbf{x}/\|\mathbf{x}\|$, where $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$.*

Let us define $g(\mathbf{x}) = \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|}$. Below we derive a central limit theorem for $g(\mathbf{x})$, when $\mathbf{x} \sim \mathcal{N}(\mu, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal matrix in $\mathbb{R}^{n\times n}$ such that $\Sigma_{ii} = \mathbb{E}x_i^2 = \sigma_i^2$ i.e., the components x_i have mean μ_i and variance σ_i^2. We derive this general result that will be useful later in the paper. For now we derive the specific case where $\mu = 0$, and $\Sigma = \mathbf{I}$.

Lemma 1 *(Central Limit Theorem for $\|\mathbf{x}\|_1/\|\mathbf{x}\|$) Let $\mathbf{x}$ be a Gaussian random vector with independent and identically distributed (i.i.d.) components, then $g(\mathbf{x})$ satisfies a central limit theorem with the limit distribution being Gaussian with mean $\mu_0 = \sqrt{\frac{n}{\alpha_2}}\alpha_1$ and variance $\sigma_0^2 = \frac{1}{\alpha_2}\left(C_{11} + (\frac{\alpha_1}{2\alpha_2})^2 C_{22} - \frac{\alpha_1}{\alpha_2}C_{12}\right)$, where $\alpha_1 = \mathbb{E}(|x_1|)$, $\alpha_2 = \mathbb{E}(|x_1|^2)$, $C_{11} = Var(|x_1|), C_{22} = Var(|x_1|^2), C_{12} = \mathbb{E}((|x_1| - \mathbb{E}(|x_1|))(|x_1|^2 - \mathbb{E}(|x_1|^2)))$, i.e., $g(\mathbf{x}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mu_0, \sigma_0^2))$.*

Proof: Consider the two dimensional vector $\mathbf{z}_i = \begin{pmatrix} |x_i| \\ |x_i|^2 \end{pmatrix}$, and $\mathbf{z}^{(n)} = \sum_{i=1}^n \mathbf{z}_i$. Note that $\mathbf{z}_i$ are i.i.d. random vectors in $\mathbb{R}^2$, with mean $\mathbf{m} = \mathbf{E}\mathbf{z}_i = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$, and covariance matrix

$$\mathbf{C} = \begin{bmatrix} \mathbb{E}|x_i|^2 - (\mathbb{E}|x_i|)^2 & \mathbb{E}|x_i|^3 - \mathbb{E}|x_i|^2\mathbb{E}|x_i| \\ \mathbb{E}|x_i|^3 - \mathbb{E}|x_i|^2\mathbb{E}|x_i| & \mathbb{E}|x_i|^4 - (\mathbb{E}|x_i|^2)^2 \end{bmatrix}.$$

Hence, by applying the multidimensional CLT , see [23], we conclude that the distribution of $\mathbf{r}^{(n)} = \frac{1}{\sqrt{n}}\left(\mathbf{z}^{(n)} - n\mathbf{m}\right)$ converges to $\mathcal{N}(0, \mathbf{C})$. Now the function

$g(\mathbf{x})$ can be represented as a function of the vector $\mathbf{z}^{(n)}$, which we denote as g for brevity. By the Skorohod representation theorem see [23] there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ where we can construct a sequence of random vectors $\mathbf{r}^{(n)}$ that converges in the almost sure sense to the random vector $\mathbf{r}$ with distribution $\mathcal{N}(\mathbf{0}, \mathbf{C})$. Therefore,

$$\begin{aligned}
g &= \frac{z_1^{(n)}}{\sqrt{z_2^{(n)}}} \\
&= (\sqrt{n}r_1 + n\alpha_1)(\sqrt{n}r_2 + n\alpha_2)^{-1/2} \\
&= \frac{1}{\alpha_2^{1/2}}(r_1 + \sqrt{n}\alpha_1)(1 - \frac{1}{2}\frac{r_2}{\alpha_2\sqrt{n}} + \mathrm{o}_p(n^{-1/2})) \\
&= \frac{1}{\alpha_2^{1/2}}(\sqrt{n}\alpha_1 - \frac{r_2}{2\alpha_2}\alpha_1 + r_1 - \mathrm{O}_p(n^{-1/2}) + \mathrm{o}_p(n^{-1/2})) \\
&= \sqrt{n}\frac{\alpha_1}{\sqrt{\alpha_2}} + \frac{1}{\sqrt{\alpha_2}}(r_1 - \frac{r_2}{2}\frac{\alpha_1}{\alpha_2}) + \mathrm{o}_p(1),
\end{aligned}$$

Therefore we obtain

$$g - \sqrt{n}\frac{\alpha_1}{\sqrt{\alpha_2}} = \frac{1}{\sqrt{\alpha_2}}(r_1 - \frac{\alpha_1}{2\alpha_2}r_2) + \mathrm{o}_p(1). \tag{4}$$

Since the vector $\mathbf{r}^{(n)}$ almost surely converges to the vector $\mathbf{r}$, by the Continuous Mapping Theorem, any continuous function $f(\mathbf{r}^{(n)})$ converges to $f(\mathbf{r}))$ almost surely, where in our case $f(\mathbf{r}) = \frac{1}{\sqrt{\alpha_2}}(r_1 - \frac{\alpha_1}{2\alpha_2}r_2)$. But this is a linear combination of two jointly Gaussian random variables, and hence is also a Gaussian Random Variable (r.v) with mean 0, and variance $\beta_1 + \beta_2\frac{\alpha_1^2}{4} - \alpha_1\beta_{12}$. Also, by the fact that if x_n, y_n are two random variable sequences such that $x_n \to x$ a.s. and $y_n \to y$ in probability, then $x_n + y_n \to x + y$ in probability, the right hand side of (4) is a random variable that converges in probability to a Gaussian random variable with mean 0, and variance $\sigma_{(0)}^2 = \frac{1}{\alpha_2}\left(C_{11} + (\frac{\alpha_1}{2\alpha_2})^2 C_{22} - \frac{\alpha_1}{\alpha_2}C_{12}\right) = 1 - 3/\pi$, and hence g converges to a Gaussian random variable with mean $\mu_{(0)} = \sqrt{n}\frac{\alpha_1}{\sqrt{\alpha_2}} = \sqrt{\frac{2n}{\pi}}$ and variance $\sigma_{(0)}^2$. Now $g(\mathbf{x})$ has the same distribution as g. Therefore $g(\mathbf{x})$ converges in distribution to $\mathcal{N}(\mu_{(0)}, \sigma_{(0)}^2)$. $\square$

Proposition 1 *Under $\mathcal{H}_0, \chi \sim \mathcal{N}(\mu_{(0)}, \sigma_{(0)}^2)$, asymptotically in distribution, where $\mu_{(0)} = \sqrt{\frac{2n}{\pi}}$, and $\sigma_{(0)}^2 = 1 - \frac{3}{\pi}$.*

Proof: The proof uses Approximation **??** and follows from Lemma 1, where $\alpha_1 = \mathbb{E}(|x_1|) = \sqrt{\frac{2}{\pi}}$, $\alpha_2 = \mathbb{E}(|x_1|^2) = 1$, $C_{11} = \mathrm{Var}(|x_1|) = 1 - 2/\pi, C_{22} = \mathrm{Var}(|x_1|^2) = 2, C_{12} = \mathbb{E}((|x_1| - \mathbb{E}(|x_1|))(|x_1|^2 - \mathbb{E}(|x_1|^2))) = \sqrt{\frac{2}{\pi}}$. $\square$

3.2 Eigenvalue and eigenvector properties under $\mathcal{H}_1$

Under hypothesis $\mathcal{H}_1$ the matrix $\boldsymbol{\mathcal{A}}$ is given as below

$$\mathcal{A}_{ij} = \begin{cases} \begin{cases} 1 - p_b & \text{w.p. } p_1 \\ -p_b & \text{w.p. } 1 - p_1 \end{cases}, & \text{if } 1 \leq i, j \leq m, \\ \begin{cases} 1 - p_b & \text{w.p. } p_b \\ -p_b & \text{w.p. } 1 - p_b \end{cases} & \text{if } i > m \text{ or } j > m, \end{cases}$$

Thus under $\mathcal{H}_1$, the matrix $\boldsymbol{\mathcal{A}}$ has a non-zero mean $\overline{\boldsymbol{\mathcal{A}}} = \mathbb{E}_{\mathcal{H}_1}\boldsymbol{\mathcal{A}}$ given by

$$\overline{\boldsymbol{\mathcal{A}}} = \begin{bmatrix} (p_1 - p_b)\mathbf{J}_m & \mathbf{0}_{m\times n-m} \\ \mathbf{0}_{n-m\times m} & \mathbf{0}_{n-m\times n-m} \end{bmatrix}. \tag{5}$$

Also note that for the components $\mathcal{A}_{ij}$, such that $1 \leq i, j \leq m$, the upper diagonal components have the variance of $p_1(1 - p_1)$, and the other components have a variance of $p_b\delta_p$, where $\delta_p := p_1 - p_b$.

This matrix has rank 1, and with a single non-zero eigenvalue $m\delta_p$, with eigenvector $\frac{1}{\sqrt{m}}\begin{bmatrix} \mathbf{1}_m \\ \mathbf{0}_{n-m\times 1} \end{bmatrix}$.

Intuitively, $\overline{\boldsymbol{\mathcal{A}}}$ is the subgraph component, and when the subgraph component is large enough, we can conceivably detect the subgraph from the observed graph, i.e., if the eigenvalue of $\overline{\boldsymbol{\mathcal{A}}}$ is large to be separate enough from the spectrum of $\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}$, we expect to be able to detect the embedded subgraph. A bound on $\|\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}\|$ is obtained by use of the following theorem based on the matrix Bernstein's Lemma [26].

Theorem 1 *Under the condition that* $p_b \gg \frac{\log^2(n)}{n}$,

$$\begin{aligned} \|\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}\| &< \sqrt{12\log(n)\max(\sigma_1^2 m + \sigma_0^2(n-m), \sigma_0^2 n)} \\ &= \sqrt{12\log(n)\sigma^2} \text{ almost surely (a.s.),} \end{aligned} \tag{6}$$

where $\sigma_1^2 = p_1(1-p_1), \sigma_0^2 = p_0(1-p_0)$, *and define* $\sigma^2 := \max(\sigma_1^2 m + \sigma_0^2(n - m), \sigma_0^2 n)$.

Proof:
We need the following lemma on Matrix Bernstein inequality [26].

Lemma 2 *(Matrix Bernstein). Let* $\mathbf{S}_1, \mathbf{S}_2, \dots \mathbf{S}_t$ *be independent random matrices with common dimension* $d_1 \times d_2$. *Assume that each matrix has bounded deviation from its mean, i.e.,*

$$\|\mathbf{S}_k - \mathbb{E}\mathbf{S}_k\| \leq R, \text{ for each } k = 1, \dots n.$$

Form the sum $\mathbf{Z} = \sum_{k=1}^t \mathbf{S}_k$ *and introduce a variance parameter*

$$\sigma_{\mathbf{Z}}^2 = \max\left\{\|\mathbf{E}[(\mathbf{Z} - \mathbb{E}\mathbf{Z})(\mathbf{Z} - \mathbb{E}\mathbf{Z})^H]\|, \|\mathbf{E}[(\mathbf{Z} - \mathbb{E}\mathbf{Z})^H(\mathbf{Z} - \mathbb{E}\mathbf{Z})]\|\right\}.$$

Then

$$\mathbb{P}\{\|\mathbf{Z}-\mathbb{E}\mathbf{Z}\|>t\} \leq (d_1+d_2).\exp\left(\frac{-t^2/2}{\sigma_{\mathbf{Z}}^2+Rt/3}\right), \tag{7}$$

for all $t \geq 0$.

With $\mathbf{Z} := \boldsymbol{\mathcal{A}}$, we can decompose $\mathbf{Z}$ as sums of Hermitian matrices $\mathbf{S}_{i'j'}$, $\mathbf{Z} = \sum_{i'j'} \mathbf{S}_{i'j'}$ such that:

$$(\mathbf{S}_{i'j'})_{ij} = \begin{cases} \mathcal{A}_{i'j'} \text{ if } i = i', j = j' \\ \mathcal{A}_{i'j'} \text{ if } i = j', j = i' \\ 0 \text{ otherwise.} \end{cases} \tag{8}$$

Notice that $\|\mathbf{S}_{i'j'}\mathbf{x}\| = |2x_{i'}x_{j'}\mathcal{A}_{mn}| \leq |x_{i'}^2 + x_{j'}^2|$. Consequently $\|\mathbf{S}_{i'j'}\| \leq 1$, giving $R = 1$. Let $\mathbf{Y} = \mathbf{E}[(\mathbf{Z}-\mathbb{E}\mathbf{Z})^H(\mathbf{Z}-\mathbb{E}\mathbf{Z})]$, then

$$Y_{ij} = \begin{cases} v_1 & \text{if } i = j, i \leq m \\ v_2 & \text{if } i = j, i > m \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $v_1 = \sigma_1^2 m + \sigma_0^2(n-m), v_2 = \sigma_0^2 n$, with $\sigma_1^2 = p_1(1-p_1)$, and $\sigma_0^2 = p_b(1-p_b)$. Therefore $\sigma_{\mathbf{Z}}^2 = \max(v_1, v_2) := \sigma^2$. Thus it follows that

$$\begin{aligned} \mathbf{P}(\|\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}\| \geq c\sigma) &\leq 2n\exp(\frac{-c^2\sigma^2}{2\sigma^2 + c\sigma/3}) \\ &\leq 2n\exp(-c^2/3), \end{aligned}$$

if $\sigma^2 > c\sigma$ or $\sigma > c$. The RHS falls faster than n^{-1} if $c > \sqrt{6\log(n)}$, and thus by an application of Borel-Cantelli Lemma [23] the result follows. ☐

For the above result to hold we require that $\exists N$ s.t. $\forall n > N \max(v_1, v_2) > (6\log(n))^2$.

If p_b does not scale with n, this condition is immediately satisfied. Let us consider the case where the embedded subgraph is a clique, i.e., $p_s = p_1 = 1$. Then $\sigma^2 = \sigma_0^2 n = p_b(1-p_b)n$, and the condition is satisfied when $np_b \gg \log^2(n)$; similarly when both p_1, p_b are decreasing functions of n, the condition is easily verified to be satisfied when $np_b \gg \log^2(n)$.

Definition 1 *(Spectral gap G) We define the spectral gap Δ as the difference between the maximum eigenvalue of the mean matrix and edge of the spectrum*

$$\begin{aligned} G &= m\delta_p - \|\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}\| \\ &\geq m\delta_p - \sqrt{12\log(n)\sigma^2} \\ &= G_0. \end{aligned}$$

By Lemma 4, it holds that a.s.,

$$m\delta_p(1-\Delta) \leq \lambda \leq m\delta_p(1+\Delta), \tag{10}$$

and by Theorem 1, it also holds that a.s.

$$|\lambda_i| \leq \sqrt{12\log(n)(m\delta_p + np_b)} \leq m\delta_p\Delta, \tag{11}$$

for $i \geq 2$ where,

$$\Delta := \frac{\sqrt{12\log(n)np_b}}{m\delta_p}. \tag{12}$$

Note that by Condition 3, $\Delta = \mathrm{o}(1)$.

Note: By more carefully bounding the spectral radius $\|\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}\|$ it must be possible to remove the $\sqrt{\log(n)}$ factor from Δ.

3.2.1 Eigenvector distribution under $\mathcal{H}_1$

We develop a CLT for the components of the dominant eigenvector of the "modularity" matrix $\boldsymbol{\mathcal{A}}$. It is similar in vein to the CLT derived in [21], for the components of the eigenvector of a single dimensional Random Dot Product Graph(RDPG). See [21] for further details. Throughout this section the distributions of the random variables correspond to those under $\mathcal{H}_1$, and this fact is not explicitly noted from here onwards.

We need to characterize the distribution of the *dominant eigenvector*[2] of $\boldsymbol{\mathcal{A}}$, which we denote $\mathbf{u} := \mathbf{u}_1$, corresponding to the eigenvalue $\lambda := \lambda_1$. Observe that the mean matrix $\overline{\boldsymbol{\mathcal{A}}}$ can be written as $\bar{\mathbf{x}}\bar{\mathbf{x}}^T$, where $\bar{\mathbf{x}} = \sqrt{\delta_p}\begin{bmatrix}\mathbf{1}_m^T & \mathbf{0}_{n-m}^T\end{bmatrix}^T$, with a single non-zero eigenvalue $\bar{\lambda} = m\delta_p$ and its eigenvector as $\bar{\mathbf{u}} = \frac{\bar{\mathbf{x}}}{\|\bar{\mathbf{x}}\|}$. Let us define $\mathbf{x}$ as $\mathbf{x} = \lambda^{1/2}\mathbf{u}$, and so $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|_2$. Intuitively when there is a non-diminishing spectral gap G for large n, a random realization of $\mathbf{x}$ would be close to $\bar{\mathbf{x}}$. Therefore the i^{th} component of $\mathbf{x}$ would have a limiting distribution with mean $\bar{x}_i$. We can then derive the limiting distribution of the L^1-norm statistic from the distribution of $\mathbf{x}$. We derive our results under the following conditions.

Condition 1

$$p_b \gg \frac{\log^6(n)}{n}$$

Condition 2

$$mp_1 \leq np_b$$

Condition 3

$$m\delta_p = \Omega((np_b\log(n))^{2/3})$$

[2]By the dominant eigenvalue of a matrix we mean the largest eigenvalue of the matrix, and the dominant eigenvector is the corresponding eigenvector.

Condition 4

$$mp_b = \Omega(1)$$

Notice that Condition 4 also implies that $mp_1 = \Omega(1)$, because $mp_1 > mp_b$.
Discussion of the Conditions:
The condition 2 in essence says that the "signal strength" $m\delta_p$ that we are trying to detect is not so large that it can be detected trivially, by say, an ordering of the degrees of the graph vertices. The second condition 3 is required so that the spectral gap G is large enough to prove the results on the CLT of the eigenvector components presented in this paper. It must be possible to relax this last condition by more sophisticated techniques. This we reserve for a future work.

We present below our main theorem on the CLT of the components of the dominant eigenvectors.

Theorem 2 *Under Conditions 2 and 3 the following CLT holds true for the entries of the unnormalized eigenvector* $\mathbf{x} = \lambda^{1/2}\mathbf{u}$, *where* $\mathbf{u}$ *is the eigenvector corresponding to the eigenvalue* λ *of* $\boldsymbol{\mathcal{A}}$ *under* $\mathcal{H}_1$.

$$\sqrt{\frac{m\delta_p}{p_1(1-p_1)}}\left(x_i - \sqrt{\delta_p}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{13}$$

for $1 \le i \le m$, *and*

$$\sqrt{\frac{m\delta_p}{p_b(1-p_b)}} x_i \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{14}$$

for $1 + m \le i \le n$.

Proof:
Define $\gamma_i = \sqrt{\frac{m\delta_p}{p_1(1-p_1)}}$ for $1 \le i \le m$ and $\gamma_i = \sqrt{\frac{m\delta_p}{p_b(1-p_b)}}$ for $m+1 \le i \le n$. Notice that $x_i = \frac{1}{\lambda^{1/2}}[\boldsymbol{\mathcal{A}}\mathbf{u}]_i$ and $\bar{x}_i = \frac{1}{\bar{\lambda}^{1/2}}\left[\overline{\boldsymbol{\mathcal{A}}}\bar{\mathbf{u}}\right]_i = \sqrt{\delta_p}$ for $1 \le i \le m$ and $\bar{x}_i = 0$ for $m+1 \le i \le n$. Here $[\mathbf{z}]_i$ denotes the i^{th} component of vector $\mathbf{z}$. We can write

$$\gamma_i(x_i - \bar{x}_i) \coloneqq T_1 + T_2 + T_3.$$

We treat each of the above three terms separately as below.

- We show that $T_1 = \gamma_i\left(\frac{1}{\lambda^{1/2}}[\boldsymbol{\mathcal{A}}(\mathbf{u} - \bar{\mathbf{u}})]_i\right) \to 0$ in probability, in Lemma 6.
- We show $T_2 = \gamma_i\left(\frac{1}{\lambda^{1/2}}[\boldsymbol{\mathcal{A}}\bar{\mathbf{u}} - \overline{\boldsymbol{\mathcal{A}}}\bar{\mathbf{u}}]_i\right)$ satisfies a CLT and is asymptotically distributed as $\mathcal{N}(0,1)$, in Lemma 3.
- Finally we show that $T_3 = \gamma_i\left((\frac{1}{\lambda^{1/2}} - \frac{1}{\bar{\lambda}^{1/2}})[\overline{\boldsymbol{\mathcal{A}}}\bar{\mathbf{u}}]_i\right) \to 0$, for $1 \le i \le m$in probability in Lemma 4, by showing a concentration result for the dominant eigenvalue λ. Notice that $T_3 = 0$ for $i > m$.

The result then follows by an application of Slutsky's thereom [23]. $\square$

Lemma 3 *Under Conditions 2 and 3 the following CLT holds for the entries of* $\mathbf{y}$.

$$\sqrt{\frac{m\delta_p}{p_1(1-p_1)}}\left(y_i - \frac{\|\bar{\mathbf{x}}\|\bar{x}_i}{\lambda^{1/2}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{15}$$

for $1 \leq i \leq m$, *and*

$$\sqrt{\frac{m\delta_p}{p_b(1-p_b)}} y_i \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{16}$$

for $1 + m \leq i \leq n$.

Proof:
We prove (15) and the proof for (16) follows along the same lines. Observe that

$$\begin{aligned} y_i - \frac{\|\bar{\mathbf{x}}\|\bar{x}_i}{\lambda^{1/2}} &= \frac{1}{\lambda^{1/2}} \sum_{j=1}^{n} \mathcal{A}_{ij}\bar{u}_j - \frac{\|\bar{\mathbf{x}}\|\bar{x}_i}{\lambda^{1/2}} \\ &= \frac{1}{\lambda^{1/2}} \left(\sum_{j=1}^{n} \mathcal{A}_{ij}\bar{x}_j / \|\bar{\mathbf{x}}\| - \bar{x}_i \|\bar{\mathbf{x}}\| \right) \\ &= \frac{1}{\lambda^{1/2}\|\bar{\mathbf{x}}\|} \left(\sum_{j=1}^{m} \mathcal{A}_{ij}\bar{x}_j - \bar{x}_i \|\bar{\mathbf{x}}\|^2 \right) \qquad (17) \\ &= \frac{1}{\lambda^{1/2}\|\bar{\mathbf{x}}\|} \left(\sum_{j=1}^{m} (\mathcal{A}_{ij} - \bar{x}_i\bar{x}_j)\bar{x}_j \right), \end{aligned}$$

where in (17) we used the fact that $\bar{x}_i = 0$, for $i > m$. We need the following concentration lemma for the eigenvalue λ, based on the Bauer-Fike lemma ([24]).

Lemma 4 *Under Condition 3,* $\lambda \to m\delta_p$ *a.s. as* $n \to \infty$.

Proof: By Bauer-Fike Lemma ([24]) and Theorem 1,

$$\begin{aligned} |\lambda - m\delta_p| &\leq \sqrt{6\log(n)(mp_1 + (n-m)p_b)} \\ &= \sqrt{6\log(n)(m\delta_p + np_b)}. \end{aligned}$$

Therefore,

$$\begin{aligned} \left|\frac{\lambda}{m\delta_p} - 1\right| &\leq \sqrt{\frac{6\log(n)m\delta_p}{(m\delta_p)^2} + \frac{6\log(n)np_b}{(m\delta_p)^2}} \\ &\leq \sqrt{\frac{12\log(n)np_b}{(m\delta_p)^2}} \qquad (18) \\ &= \frac{\sqrt{12\log(n)np_b}}{m\delta_p}, \end{aligned}$$

which implies $\lambda \to m\delta_p$, a.s., by Condition 3, where in (18) we used the fact that $m\delta_p < np_b$, which follows from Condition 2. $\square$

Notice $\|\bar{\mathbf{x}}\| = \sqrt{\bar{x}_1^2 + \bar{x}_2^2 + \bar{x}_3^2 + \dots \bar{x}_n^2} = \sqrt{m\delta_p}$ deterministically. Thus we obtain

$$\begin{aligned}\sqrt{\frac{m\delta_p}{p_1(1-p_1)}}(y_i - \frac{\|\bar{\mathbf{x}}\|\bar{x}_i}{\lambda^{1/2}}) &= \frac{\sqrt{m\delta_p}}{\lambda^{1/2}\sqrt{m\delta_p(p_1(1-p_1))}}\left(\sum_{j=1}^{m}(\mathcal{A}_{ij} - \bar{x}_i\bar{x}_j)\bar{x}_j\right) \\ &= \frac{\sqrt{m\delta_p}}{\sqrt{mp_1(1-p_1)}\lambda^{1/2}}\left(\sum_{j=1}^{m}(\mathcal{A}_{ij} - \delta_p)\right),\end{aligned}$$

since $\bar{x}_i = \sqrt{\delta_p}$ for $1 \le i \le m$. We invoke the Lindeberg Central Limit Theorem [23] to determine the asymptotic distribution of the above.

Theorem 3 *(Lindeberg Central Limit Theorem) Suppose that for each* n,

$$X_{n1}, X_{n2}, \dots X_{nr_n}$$

are independent, with $\mathbb{E}X_{nk} = 0, \sigma_{nk}^2 = \mathbb{E}X_{nk}^2$*, and define* $s_n^2 = \sum_{k=1}^{r_n}\sigma_{nk}^2$*. Define* $S_n = \sum_{k=1}^{r_n} X_{nk}$*. Then* $S_n/s_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$*, if*

$$\lim_{n\to\infty}\sum_{k=1}^{r_n}\frac{1}{s_n^2}\mathbb{E}X_{nk}^2\mathbb{I}\{|X_{nk}| \ge \epsilon s_n\} = 0, \tag{19}$$

$\forall \epsilon > 0$.

Now take $S_n = \sum_{j=1}^{m}(\mathcal{A}_{ij} - \delta_p)\sqrt{\delta_p}$, then $X_{nk} := (\mathcal{A}_{ij} - \delta_p)\sqrt{\delta_p}$, and $\mathbb{E}X_{nk} = 0$, and $\sigma_{nk}^2 = \mathbb{E}X_{nk}^2 = \delta_p p_1(1-p_1)$, giving $s_n = m\delta_p p_1(1-p_1)$. Then the left hand side of condition (19) becomes

$$\lim_{n\to\infty}\frac{m}{m\delta_p p_1(1-p_1)}\mathbb{E}X_{nk}^2\mathbb{I}\{|X_{nk}|/\sqrt{m\delta_p p_1(1-p_1)} \ge \epsilon\},$$

because X_{nk} are i.i.d. random variables. The above is equivalent to

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{E}\left(\frac{X_{nk}}{\sqrt{\delta_p p_1(1-p_1)}}\right)^2\mathbb{I}\left\{\frac{|X_{nk}|}{\sqrt{\delta_p p_1(1-p_1)}} \ge \epsilon\sqrt{m}\right\} \\ &:= \lim_{n\to\infty}\mathbb{E}\tilde{X}_{nk}^2\mathbb{I}\left\{|\tilde{X}_{nk}| \ge \epsilon\sqrt{m}\right\},\end{aligned} \tag{20}$$

where $\tilde{X}_{nk} = X_{nk}/\sqrt{\delta_p p_1(1-p_1)}$ is given as

$$\tilde{X}_{nk} = \begin{cases}\frac{1-p_1}{\sqrt{p_1(1-p_1)}} & \text{w.p. } p_1 \\ \frac{-p_1}{\sqrt{p_1(1-p_1)}} & \text{w.p. } 1-p_1.\end{cases}$$

Therefore we can write (20) as

$$\frac{1-p_1}{p_1}\mathbb{I}\{\sqrt{\frac{1-p_1}{p_1 m}} \geq \epsilon\} + \frac{p_1}{1-p_1}\mathbb{I}\{\sqrt{\frac{p_1}{mp_1}} \geq \epsilon\}.$$

Clearly, if $mp_1 = \Omega(1), \exists N$, s.t. the above is zero $\forall n > N$, and $\epsilon > 0$. Hence Linderberg condition is satisfied, and we obtain that

$$\lim_{n\to\infty} \frac{1}{\sqrt{m\delta_p p_1(1-p_1)}} \sum_{j=1}^{m} (\mathcal{A}_{ij} - \delta_p)\sqrt{\delta_p} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

or equivalently,

$$\lim_{n\to\infty} \frac{1}{\sqrt{mp_1(1-p_1)}} \sum_{j=1}^{m} (\mathcal{A}_{ij} - \delta_p) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{21}$$

Thus by applying Slutsky's theorem with Lemma 4 and (21) we obtain the result for $1 \leq i \leq m$.

Similarly, for $m+1 \leq i \leq n$,

$$\begin{aligned}\sqrt{\frac{m\delta_p}{p_b(1-p_b)}} y_i &= \sqrt{\frac{m\delta_p}{p_b(1-p_b)}} \lambda^{-1/2} \sum_{j=1}^{m} \mathcal{A}_{ij}\bar{u}_j, \\ &= \sqrt{\frac{m\delta_p}{\lambda}} \frac{1}{\sqrt{mp_b(1-p_b)}} \sum_{j=1}^{m} \mathcal{A}_{ij} \\ &\xrightarrow{\mathcal{D}} \mathcal{N}(0,1),\end{aligned}$$

where the proof follows from another application of Theorem 3, Lemma 4 and Slutsky's Theorem, provided that $mp_b = \Omega(1)$, which follows from Condition 4.

To complete the proof of Theorem 2, we need to first derive an entry-wise error bound between the eigenvector $\bar{\mathbf{u}}$ of $\overline{\boldsymbol{\mathcal{A}}}$ and the dominant eigenvector $\mathbf{u}$, of $\boldsymbol{\mathcal{A}}$ which we present in the following lemma.

Armed with the results we have thus far, we are now prepared to prove the main central limit theorem in the paper, a CLT for each individual component of the non-normalized dominant eigenvector $\mathbf{x}$ of $\boldsymbol{\mathcal{A}}$.

In order to prove Theorem 2, we need an error bound between $\mathbf{u}$ and $\bar{\mathbf{u}}$. To derive this we use the traditional Davis-Kahan theorem from [27], which we quote below.

Theorem 4 *(Davis-Kahan Theorem [27]) Let $\mathbf{C}$ and $\mathbf{D}$ be two Hermitian operators, and let S_1, S_2 be any two subsets of $\mathbb{R}$ such that the distance between the two subsets, $d(S_1, S_2) = \delta > 0$. Let $\mathbf{E} = P_{\mathbf{C}}(S_1)$, the projection matrix on to the* *space spanned by the eigenvectors of* $\mathbf{C}$ *whose eigenvalues fall in* S_1*, and similarly,* $\mathbf{F} = P_{\mathbf{D}}(S_2)$*. Then, for every unitarily invariant matrix norm* [3] $|||\,|||$*,*

$$|||\mathbf{EF}||| \leq \frac{c}{\delta}|||\mathbf{C} - \mathbf{D}|||$$

where c *is a fixed constant. In fact,* $c = \pi/2$.

Using the above, we derive the following result.

Lemma 5 *Let* $\mathbf{u}, \bar{\mathbf{u}}$ *and* Δ *be as defined above. Then a.s.,*

$$\|\mathbf{u} - \bar{\mathbf{u}}\|_2 \leq \frac{c\Delta}{1 - 2\Delta},$$

where c *is a constant independent of* n.

Proof:
In the notation of Theorem 4, choose $\mathbf{C} := \overline{\boldsymbol{\mathcal{A}}}$, and $\mathbf{D} := \boldsymbol{\mathcal{A}}$. Let us take $S_1 = [-a_n, a_n]$, where $a_n = m\delta_p\Delta$. Then S_1 does not contain the non-zero eigenvalue $\bar{\lambda}$ of $\mathbf{C}$, and hence $\mathbf{E} = \mathbf{P}_{\mathbf{C}}(S_1)$ is the projection matrix on to the orthogonal space of $\bar{\mathbf{u}}$, and therefore, $\mathbf{E} = \mathbf{I} - \bar{\mathbf{u}}\bar{\mathbf{u}}^T$. Let $S_2 = [m\delta_p(1 - \Delta) - \infty)$, such that it only contains the dominant eigenvalue of $\boldsymbol{\mathcal{A}}$, which gives $\mathbf{F} = \mathbf{P}_{\mathbf{D}}(S_2) = \mathbf{u}\mathbf{u}^T$. Demonstrably, δ in Theorem 4 satisfies $\delta > m\delta_p(1-\Delta) - m\delta_p\Delta = m\delta_p(1-2\Delta)$. Also, we choose $|||\,||| := \|\,\|_2$, the induced L^2-norm on matrices, which can be easily shown to be unitarily invariant. From Theorem 1 it holds that $\|\overline{\boldsymbol{\mathcal{A}}} - \boldsymbol{\mathcal{A}}\|_2 \leq m\delta_p\Delta$. Also,

$$\begin{aligned}
\|\mathbf{EF}\|_2 &= \|(\mathbf{I} - \bar{\mathbf{u}}\bar{\mathbf{u}}^T)\mathbf{u}\mathbf{u}^T\|_2 \\
&= \|\mathbf{u}\mathbf{u}^T - \bar{\mathbf{u}}(\bar{\mathbf{u}}^T\mathbf{u})\mathbf{u}^T\|_2 \\
&= \|\,(\mathbf{u} - \alpha\bar{\mathbf{u}})\,\mathbf{u}^T\|_2 \qquad (22)\\
&= \|\mathbf{u} - \alpha\bar{\mathbf{u}}\|_2 \qquad (23)\\
&= (1 - \alpha^2)^{1/2},
\end{aligned}$$

where in (22) we used the notation $\alpha := \bar{\mathbf{u}}^T\mathbf{u}$. In obtaining (23) we used the fact that $\|\mathbf{x}\mathbf{y}^T\|_2 = \|\mathbf{x}\|_2\|\mathbf{y}\|_2$, for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and in the last line we used the fact that $\|\mathbf{u}\|_2 = \|\bar{\mathbf{u}}\|_2 = 1$. Therefore by Theorem 4

$$\begin{aligned}
(1 - \alpha^2)^{1/2} &\leq \frac{\sqrt{2}c\Delta m\delta_p}{m\delta_p(1 - 2\Delta)} \\
&= c\frac{\Delta}{1 - 2\Delta} \qquad (24)
\end{aligned}$$

[3]A unitarily invariant matrix norm is such that $|||\mathbf{UAV}||| = |||\mathbf{A}|||$, for any matrix $\mathbf{A}$, where $\mathbf{U}, \mathbf{V}$ are two unitary matrices

Thus we obtain

$$
\begin{aligned}
\|\mathbf{u} - \bar{\mathbf{u}}\|_2 &= \sqrt{2}(1-\alpha)^{1/2} \\
&< \sqrt{2}(1-\alpha^2)^{1/2} && (25) \\
&\leq c\frac{\Delta}{1-2\Delta}, && (26)
\end{aligned}
$$

where in (25) we used the fact that $\mathbf{u}$ is only fixed up to a scale factor of ± 1, and so α can be chosen to be non-negative, and in (26) we used (24). $\square$

We finally need the following lemma and the subsequent observations.

Lemma 6 *There exists a constant C s.t. $\|\mathbf{y} - \frac{1}{\lambda^{1/2}}\boldsymbol{\mathcal{A}}\mathbf{u}\| \leq C\sqrt{m\delta_p}\Delta^2 = C\frac{\log(n)np_b}{(m\delta_p)^{3/2}}$, a.s.*

Proof: Observe that can write $\boldsymbol{\mathcal{A}} = \sum_{i\geq 2}\lambda_i \mathbf{u}_i\mathbf{u}_i^T + \lambda\mathbf{u}\mathbf{u}^T = \widetilde{\boldsymbol{\mathcal{A}}} + \lambda\mathbf{u}\mathbf{u}^T$, where $\|\widetilde{\boldsymbol{\mathcal{A}}}\|_2 = \max_{i>2}|\lambda_i| \leq m\delta_p\Delta$, a.s. Hence we have

$$
\begin{aligned}
\|\mathbf{y} - \frac{1}{\lambda^{1/2}}\boldsymbol{\mathcal{A}}\mathbf{u}\| &= \frac{1}{\lambda^{1/2}}\|\boldsymbol{\mathcal{A}}(\mathbf{u}-\bar{\mathbf{u}})\|_2 \\
&= \frac{1}{\lambda^{1/2}}\|\left(\widetilde{\boldsymbol{\mathcal{A}}} + \lambda\mathbf{u}\mathbf{u}^T\right)(\mathbf{u}-\bar{\mathbf{u}})\|_2 \\
&\leq \frac{\|\widetilde{\boldsymbol{\mathcal{A}}}(\mathbf{u}-\bar{\mathbf{u}})\|_2}{\lambda^{1/2}} + \lambda^{1/2}\|\mathbf{u}-\bar{\mathbf{u}}\|^2 \\
&\leq c\frac{\sqrt{m\delta_p}\Delta^2}{(1-\Delta)^{1/2}(1-2\Delta)} + \frac{C\sqrt{m\delta_p}\Delta^2(1+\Delta)^{1/2}}{(1-2\Delta)^2} && (27) \\
&\leq C\sqrt{m\delta_p}\Delta^2,
\end{aligned}
$$

a.s., where in (27), we used the bound in Lemma 5.

Notice that the eigenvector components $\mathbf{u}_i$, are exchangeable for $1 \leq i \leq m$,, and similarly for $\mathbf{u}_i, 1+m \leq i \leq n$. (This is clear since we have $\boldsymbol{\mathcal{A}}\mathbf{u} = \lambda\mathbf{u}$, and the distribution of $\mathcal{A}_{ij}$ being the same for $1 \leq i \leq m$, and for $i > m$.)

Lemma 7 *For $1 \leq i \leq m$, we have $\sqrt{\frac{m\delta_p}{p_1(1-p_1)}}|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}| \to 0$, and for $m+1 \leq i \leq n$, we have $\sqrt{\frac{m\delta_p}{p_b(1-p_b)}}|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}| \to 0$, in probability.*

Proof:
For $1 \leq i \leq m$, using Markov inequality,

$$
\begin{aligned}
\mathbb{P}\{\sqrt{\frac{m\delta_p}{p_1(1-p_1)}}\left|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}\right| > \epsilon\} &\leq C\frac{\mathbb{E}m\frac{\delta_p}{p_1}|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}|^2}{\epsilon^2} \\
&= C\frac{\sum_{i=1}^m \mathbb{E}|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}|^2}{\epsilon^2} \qquad (28) \\
&\leq C\frac{\mathbb{E}\|\mathbf{y} - \frac{1}{\lambda^{1/2}}\boldsymbol{\mathcal{A}}\mathbf{u}\|^2}{\epsilon^2} \\
&\leq C\left(\frac{\log(n)np_b}{(m\delta_p)^{3/2}}\right)^2 \to 0, \qquad (29)
\end{aligned}
$$

where (28) follows from $\frac{\delta_p}{p_1} = \frac{p_1 - p_b}{p_1} \leq K$, for some $K, N, n > N$, and exchangeability, and the last step follows from Lemma 6. Similarly for $1 + m \leq i \leq n$,

$$
\begin{aligned}
\mathbb{P}\{\sqrt{\frac{m\delta_p}{p_b(1-p_b)}}|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}| > \epsilon\} &\leq \frac{\mathbb{E}m|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}|^2}{\epsilon^2}\frac{\delta_p}{p_b(1-p_b)} \\
&\leq C\frac{\mathbb{E}(n-m)|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}|^2}{\epsilon^2} \qquad (30) \\
&= C\frac{\sum_{i=1+m}^n \mathbb{E}|y_i - \frac{(\boldsymbol{\mathcal{A}}\mathbf{u})_i}{\lambda^{1/2}}|^2}{\epsilon^2} \qquad (31) \\
&\leq C\frac{\mathbb{E}\|\mathbf{y} - \frac{1}{\lambda^{1/2}}\boldsymbol{\mathcal{A}}\mathbf{u}\|^2}{\epsilon^2} \\
&\leq C\left(\frac{\log(n)np_b}{(m\delta_p)^{3/2}}\right)^2 \to 0,
\end{aligned}
$$

where in (30), we use Condition 2, and in 31, we used exchangeability of $u_i, m + 1 \leq i \leq n$.

Furthermore, we need the following lemma.

Lemma 8 *Under Condition 3,* $\sqrt{\frac{m\delta_p}{p_1(1-p_1)}}\left(\frac{\|\bar{\mathbf{x}}\|}{\bar{\lambda}^{1/2}} - \frac{\|\bar{\mathbf{x}}\|}{\lambda^{1/2}}\right)\bar{x}_i \to 0.$

Proof:
It holds that

$$(1 - \frac{\|\bar{\mathbf{x}}\|}{\lambda^{1/2}}) = \frac{\lambda - \|\bar{\mathbf{x}}\|^2}{(\lambda^{1/2} + \|\bar{\mathbf{x}}\|)\lambda^{1/2}} \qquad (32)$$

Following approach in Athreya,

$$
\begin{aligned}
|\lambda^{1/2} - \|\bar{\mathbf{x}}\|^2| &= |\mathbf{u}^T\boldsymbol{\mathcal{A}}\mathbf{u} - \bar{\mathbf{u}}^T\overline{\boldsymbol{\mathcal{A}}}\bar{\mathbf{u}}| \\
&\leq |\mathbf{u}^T\boldsymbol{\mathcal{A}}\mathbf{u} - \bar{\mathbf{u}}^T\boldsymbol{\mathcal{A}}\bar{\mathbf{u}}| + |\bar{\mathbf{u}}^T\boldsymbol{\mathcal{A}}\bar{\mathbf{u}} - \bar{\mathbf{u}}^T\overline{\boldsymbol{\mathcal{A}}}\bar{\mathbf{u}}| \\
&\leq 2\lambda\|\mathbf{u} - \bar{\mathbf{u}}\|^2,
\end{aligned}
$$

asymptotically almost surely (a.a.s).
From (32) we obtain

$$\left|1 - \frac{\|\bar{\mathbf{x}}\|}{\lambda^{1/2}}\right| \leq \frac{2\lambda\|\mathbf{u}-\bar{\mathbf{u}}\|^2}{m\delta_p(1+(1-\Delta)^{1/2})(1-\Delta^{1/2})} \tag{33}$$

$$\leq c\Delta^2(1+K\Delta) \tag{34}$$

$$\leq c\Delta^2 \tag{35}$$

a.a.s, and hence $\delta_p\sqrt{\frac{m}{p_1(1-p_1)}}(1-\frac{\|\bar{\mathbf{x}}\|}{\lambda^{1/2}}) \leq C\left(\sqrt{\delta_p}\frac{\log(n)np_b}{(m\delta_p)^{3/2}}\right) \to 0$, using Condition 3. □

3.2.2 Distribution of χ under $\mathcal{H}_1$

We use the CLT derived in Thereom 2 to derive an approximate CLT for our test statistic $\chi = \|\mathbf{u}\|_1$ under $\mathcal{H}_1$. The distribution is approximate since we make the assumption that the components of $\mathbf{x}$ are independently distributed and have the Gaussian distribution derived in thereom 2 for finite n as opposed to the asymptotic regime in which Theorem 2 holds.

Proposition 2 *Under the assumption that the components of* $\mathbf{x}$ *are independent and Gaussian with the distribution derived in theorem 2,* $\frac{\chi-\mu_{(1)}}{\sigma_{(1)}}$ *is asymptotically distributed as* $\mathcal{N}(0,1)$.
To simplify the presentation of the formulae we introduce the following notation. Let $r = \frac{m\delta_p^2}{2p_1(1-p_1)}$, $s = \frac{m\delta_p^2}{2p_b(1-p_b)}$. *Also,* $\beta_1 = \sqrt{\frac{\delta_p}{\pi r}}e^{-r} + \sqrt{\delta_p}\left(1-2Q(\sqrt{2r})\right)$, *and* $\beta_2 = \sqrt{\frac{\delta_p}{\pi s}}$. *In addition we also define*

$$E_1 = \frac{1}{\sqrt{\pi}}\left(\frac{\delta_p}{r}\right)^{3/2} M(-\frac{3}{2},\frac{1}{2},-r)$$

$$E_2 = \frac{3}{4}\left(\frac{\delta_p}{r}\right)^2 M(-2,1/2,-r)$$

where $M(a,b,z)$ *is the confluent hypergeometric gamma function [29]. Then*

$$\mu_{(1)} = \frac{N_{\alpha_1}}{\sqrt{N_{\alpha_2}}}$$

and

$$\sigma_{(1)}^2 = \frac{1}{N_{\alpha_2}}\left(C_{11} + \left(\frac{N_{\alpha_1}}{2N_{\alpha_2}}\right)^2 C_{22} - \frac{N_{\alpha_1}}{N_{\alpha_2}}C_{12}\right),$$

where $N_{\alpha_1} = m\beta_1 + (n-m)\beta_2$, *and* $N_{\alpha_2} = m\left(\delta_p(1+\frac{1}{2r})\right) + (1-\frac{2}{\pi})\frac{\delta_p(n-m)}{2s}$.
Finally,

$$C_{11} = m\left(\delta_p(1+\frac{1}{2r}) - \beta_1^2\right) + (1-\frac{2}{\pi})\frac{\delta_p(n-m)}{2s}$$

$$C_{12} = m\left(E_1 - \beta_1\delta_p(1+\frac{1}{2r})\right) + \frac{n-m}{\sqrt{4\pi}}\left(\frac{\delta_p}{s}\right)^{3/2}$$

$$C_{22} = m\left(E_2 - \delta_p^2(1+\frac{1}{2r})^2\right) + \frac{3(n-m)}{4}\left(\frac{\delta_p}{s}\right)^2$$

The CLT result stated in Proposition 2 is approximate, since in deriving the result we assumed that the components of the scaled dominant eigenvector are Gaussian for finite n, whereas in truth the distribution is only Gaussian in the asymptotic limit. On the other hand, from simulations we see that the distribution indeed matches our prediction. We provide approximate expressions of $\mu_{(1)}$ and $\sigma^2_{(1)}$ derived above, using the fact that $r = \Omega(1)$, and $s = \Omega(1)$. For the parameter values we choose under the Conditions 1,3 and 4, and using asymptotic approximations for the Q-function and $M(a, b, x)$, [29] we can show that for large n,

$$\mu_{(1)} \approx \sqrt{m}\left(1 - \frac{1}{4r} - \frac{\rho}{4s}\right)\left(1 + \frac{\rho}{\sqrt{\pi s}}\right),$$

where $\rho := \frac{n-m}{m}$. For large n, the fractions in the braces are $\mathrm{o}(1)$ implying that the expected value of χ is close to $\sqrt{m} \ll \mu_{(0)}$. This agrees with our intuition that asymptotically the eigenvector $\mathbf{u}$ is localized to the nodes belonging to the subgraph. Similarly using the asymptotic approximation for $M(a, b, x)$ for large x [29], one can show that for large n, and m, δ_p satisfying Condition 3,

$$\sigma^2_{(1)} \approx \frac{1}{2}(1-\frac{2}{\pi})\frac{\rho}{s}(1 - \frac{1}{2r} - \frac{\rho}{2s})$$

Thus we see that $\sigma^2_{(1)} \sim \frac{\rho}{s} = \frac{2(n-m)p_b(1-p_b)}{(m\delta_p)^2} \sim \frac{(n-m)p_b}{(m\delta_p)^2}$. This is interesting because it says that the variance of χ under $\mathcal{H}_1$ is inversely proportional to the strength of the signal $m\delta_p$ and in addition it is inversely proportional to Δ, the spectral gap ratio, indicating that smaller the spectral gap, the harder it is to detect the presence of the subgraph. In addition $\sigma^2_{(1)}$ is several orders of magnitude less than $\mu_{(1)}$ and so the concentration is quite sharp.

4 Simulations

We present simulations to validate the distributions of the statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$. We choose values of m, n, δ_p and p_b so that the Conditions 1, 2, 3 and 4 are satisfied. First we generate an ER graph of size $n = 1500$ and edge probability $p_b = 0.15$, and calculate the dominant eigenvector of its modularity matrix. We compute its L^1-norm and repeat the experiment $1e^4$ times and compute the empirical CDF $F_\chi(\chi)$, which is the solid blue line with "x" marker in figure 1. In the

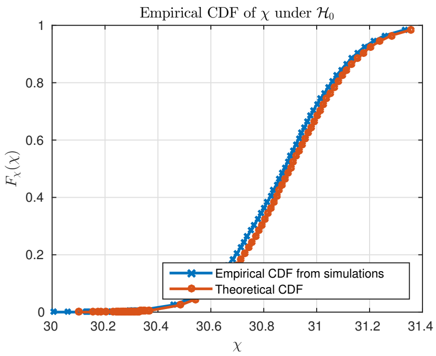

Figure 1: CDF of χ under $\mathcal{H}_0$

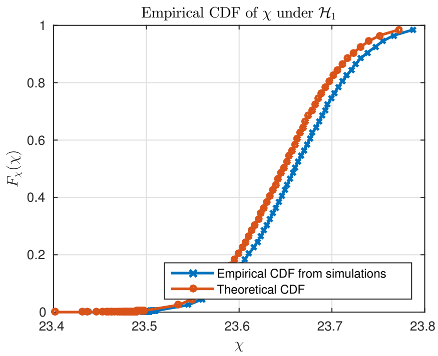

Figure 2: CDF of χ under $\mathcal{H}_1$.

same figure we plot the CDF of a Gaussian r.v with mean $\mu_{(0)}$ and variance $\sigma^2_{(0)}$ (red solid line with "o" marker). This verifies that χ indeed has a distribution close to a Gaussian with the predicted mean and variance. Next we embed a subgraph in this ER graph with $m = 450$ and $\delta_p = 0.25$, and compute the L^1-norm of the dominant eigenvector and repeat the experiment $1e^4$ times to obtain the empirical CDF. The results are plotted in figure 2. We indeed can observe that the empirical CDF (blue solid line with "x" marker), matches quite well with the Gaussian CDF (red solid line with "o" marker whose mean and variance are $\mu_{(1)}$ and $\sigma^2_{(1)}$ respectively, thus corroborating our theoretical findings. Notice that because the distributions are far apart in the parameter regime under consideration, we obtain practically error free detection.

5 Conclusions and Future Work

In this work we study a test statistic χ which is the L^1-norm of the dominant eigenvector of the modularity matrix of the random graph and analyse its distribution in the presence and absence of the anomalous subgraph. We show that the distributions are sufficiently far apart so that error free detection is possible. In the future we would like to improve the scaling of $m\delta_p$ with respect to n. As shown in a few works, detecting subgraph nodes is not possible if this quantity scales slower than $\theta(\sqrt{np_b})$ [3]. We expect that it must be possible to detect the presence of the anomaly under a much more stringent regime, even though we cannot detect the subgraph nodes. In addition we note that the results we derive on the distribution of eigenvector components in this paper may be useful in performing a classification test on the eigenvector components to detect the subgraph nodes.

References

[1] B. Miller, M. Beard, P. Wolfe, and N. Bliss, "A spectral framework for anomalous subgraph detection," *Signal Processing, IEEE Transactions on*, vol. 63,

no. 16, pp. 4191–4206, 2015.

[2] B. Miller, N. Bliss, and P. Wolfe, *Subgraph detection using eigenvector L1 norms*, 2010.

[3] B. Hajek, Y. Wu, and J. Xu, "Recovering a Hidden Community Beyond the Spectral Limit in $O(|E| \log^* |V|)$ Time," *arXiv preprint arXiv:1510.02786*, 2015.

[4] V. Jethava, A. Martinsson, C. Bhattacharyya, and D. Dubhashi, "Lovász ϑ function, svms and finding dense subgraphs," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 3495–3536, 2013.

[5] K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," *The Annals of Statistics*, pp. 1878–1915, 2011.

[6] 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.

[7] N. Alon, M. Krivelevich, and B. Sudakov, "Finding a large hidden clique in a random graph," *Random Structures & Algorithms*, vol. 13, no. 3-4, pp. 457–466, 1998. [Online]. Available: http://dx.doi.org/10.1002/(SICI)1098-2418(199810/12)13:3/4⟨457::AID-RSA14⟩3.0.CO;2-W

[8] R. R. Nadakuditi, "On hard limits of eigen-analysis based planted clique detection," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 129–132.

[9] T. Tao and V. Vu, "Random matrices: Universal properties of eigenvectors," *Random Matrices: Theory and Applications*, vol. 1, no. 01, p. 1150001, 2012.

[10] Z. Bai and G. Pan, "Limiting behavior of eigenvectors of large wigner matrices," *Journal of Statistical Physics*, vol. 146, no. 3, pp. 519–549, 2012.

[11] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[12] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[13] C. Bordenave and A. Guionnet, "Localization and delocalization of eigenvectors for heavy-tailed random matrices," *Probability Theory and Related Fields*, vol. 157, no. 3-4, pp. 885–953, 2013.

[14] L. Erdős, B. Schlein, and H.-T. Yau, "Local semicircle law and complete delocalization for wigner random matrices," *Communications in Mathematical Physics*, vol. 287, no. 2, pp. 641–655, 2009.

[15] M. Rudelson and R. Vershynin, “No-gaps delocalization for general random matrices,” *arXiv preprint arXiv:1506.04012*, 2015.

[16] F. Benaych-Georges, “Eigenvectors of wigner matrices: universality of global fluctuations,” *arXiv preprint arXiv:1104.1219*, 2011.

[17] 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.

[18] K. Avrachenkov, L. Cottatellucci, and A. Kadavankandy, “Spectral properties of random matrices for stochastic block model,” in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on*. IEEE, 2015, pp. 537–544.

[19] X. Ding and T. Jiang, “Spectral distributions of adjacency and laplacian matrices of random graphs,” *The Annals of Applied Probability*, vol. 20, no. 6, 2010.

[20] S. Heimlicher, M. Lelarge, and L. Massoulié, “Community detection in the labelled stochastic block model,” *arXiv preprint arXiv:1209.2910*, 2012.

[21] 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.

[22] P. Mitra, “Entrywise bounds for eigenvectors of random graphs,” *the electronic journal of combinatorics*, vol. 16, no. 1, p. R131, 2009.

[23] P. Billingsley, *Probability and Measure*, 3rd ed. New York, NY: Wiley, 1995.

[24] Y. Saad, *Numerical methods for large eigenvalue problems*. SIAM, 1992, vol. 158.

[25] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, ser. Cambridge studies in advanced mathematics. Cambridge University Press, 2009, vol. 118.

[26] J. A. Tropp, “User-friendly tail bounds for sums of random matrices,” *Foundations of computational mathematics*, vol. 12, no. 4, pp. 389–434, 2012.

[27] R. Bhatia, *Matrix analysis*. Springer Science & Business Media, 2013, vol. 169.

[28] F. Benaych-Georges and A. Guionnet, “Central limit theorem for eigenvectors of heavy tailed matrices,” *Electron. J. Probab*, vol. 19, no. 54, pp. 1–27, 2014.

[29] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1964, vol. 55.

[30] A. Kadavankandy, L. Cottatellucci, and K. Avrachenkov, "Characterization of l^1-norm statistic for anomaly detection in erdös rényi graphs," Eurecom, France, Research Report RR-16-315, March 2016. [Online]. Available: http://www.eurecom.fr/en/publication/list/author/cottatellucci-laura