Characterization of L1-norm Statistic for Anomaly Detection in Erdos Renyi Graphs

Arun Kadavankandy, Laura Cottatellucci, Konstantin Avrachenkov

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

Arun Kadavankandy, Laura Cottatellucci, and Konstantin Avrachenkov

***Abstract*— We describe a test statistic based on the L^1-norm of the eigenvectors of a modularity matrix to detect the presence of an embedded Erdős Rényi (ER) subgraph inside a larger ER random graph. An embedded subgraph may model a hidden community in a large network such as a social network or a computer network. We make use of the properties of the asymptotic distribution of eigenvectors of random graphs to derive the distribution of the test statistic under certain conditions on the subgraph size and edge probabilities. We show that the distributions differ sufficiently for well defined ranges of subgraph sizes and edge probabilities of the background graph and the subgraph. This method can have applications where it is sufficient to know whether there is an anomaly in a given graph without the need to infer its location. The results we derive on the distribution of the components of the eigenvector may also be useful to detect the subgraph nodes.**

***Index Terms*— Subgraph detection, Erdos-Renyi, Detection and Estimation**

I. 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 within the general framework of graph anomaly detection, which has been studied from a signal processing point of view in [1], [2] and the references therein. Graphs can efficiently capture long-range correlations among data-objects in many fields such as physics, social sciences, biology, and information systems. Anomaly detection on graphs is a branch of data mining that focuses on the analysis of such data instances to discover rare occurrences. This fundamental problem has significance in varied applications in domains such as security, finance, politics, marketing, and information and communications technologies. The interested reader is referred to [3] for a presentation of a wide range of real-world applications in telecom, auction, account, opinion, social and computer networks.

Specifically, we consider a special case of the above problem where the random graph is an ER graph and the embedded subgraph is also an ER graph with a greater density of edges. A random ER graph embedded in a large ER graph has been proposed to model terrorist transactions in a large network [4]. More generally, an embedded subgraph may model a hidden community in a larger network such as a social network. This model subsumes the clique detection problem as studied in [5], which is an important problem in Theoretical Computer Science, and also in cryptography [6]. Here the goal is to understand the size of the smallest clique that can be detected by polynomial-time algorithms. See [7], [8] for some work in this direction.

Our goal is different from classifying the subgraph nodes as done in [8], [9], in that we do not attempt to locate nodes of the subgraph. Also note 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 [10], [11].

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 [12], [13], 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 [14], [15]. Delocalization properties of eigenvectors of random matrices under a variety of distributions have been studied recently in a series of works [16]–[18].

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

This work was partly funded by the French Government (National Research Agency, ANR) through the "Investments for the Future" Program reference #ANR-11-LABX-0031-01.

A. Kadavankandy is with INRIA, Sophia Antipolis and University of Nice, arun.kadavankandy@inria.fr. L. Cottatellucci is with Eurecom, France, laura.cottatellucci@eurecom.fr. K. Avrachenkov is with INRIA, Sophia Antipolis, k.avrachenkov@inria.fr

random graphs.

Our contribution is threefold. We derive the distribution of the dominant eigenvector components of the modularity matrix when there is an embedded subgraph. The modularity matrix is the adjacency matrix of the graph with the edge probability of the background graph subtracted. It was introduced in [19] where it is used as a metric to measure the quality of community partitioning in a general graph. 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 [12], [20], 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.

Next 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, $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots x_n^2}$, is the L^2-norm of $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \dots |x_n|$ 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 we use the standard big O, Σ notations. 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 $\boldsymbol{\Sigma}_n$.

In section II we first formulate the general detection problem, and later in the more specific case studied in this paper. In section III, we present our anomaly detection algorithm, which is a hypothesis test problem with the probability of false alarm fixed. In section III-A, we describe the spectral properties of the 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 III-B 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 III-B.2.In section IV we present some simulation results and finally in section V we describe our conclusions and directions for future research.

II. The Anomaly detection problem

In this section we formulate the general problem of anomalous subgraph detection. 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 $E = E_b \cup E_s$. Based on an observation of the graph G, we decide on hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$ where

$$\mathcal{H}_0 : E = E_b \tag{1}$$

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

In our model, 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 in the asymptotic limit when the graph size scales to infinity. 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. [21], [22].

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;

$$p_a = \begin{cases} p_1 & \text{if } \mathcal{H}_1 \\ p_b & \text{if } \mathcal{H}_0. \end{cases}$$

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. Let $\boldsymbol{\mathcal{A}} = \mathbf{A} - p_b \mathbf{J}_n$ be the modularity matrix. 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 the latter 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), $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

III. 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 $\mathcal{N}(0,1)$.

In the following sections we present our mathematical results. We skip the proofs due to space constraints, and the reader is referred to the research report [23].

We need some technical conditions to prove our results, which are given below.

Condition 1:

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

Condition 2:

$$mp_1 \leq np_b$$

Condition 3:

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

where $\delta_p \equiv p_1 - p_0$

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:
We need Condition 1 to make sure that the adjacency matrix components have light tails, so that the assumption of Haar distribution of the eigenvectors of the centralized adjacency matrix in Approximation 1 is valid. Condition 3 is required so that the gap between the dominant eigenvalue λ_1 of $\boldsymbol{\mathcal{A}}$ and the next largest eigenvalue is large enough so that the dominant eigenvector is localized. We use it to prove the CLT for the eigenvector components presented in this paper. We believe that it is possible to relax this condition by more sophisticated techniques, which we reserve for future work. Condition 4 is purely technical and is required to show CLT of the components of the eigenvector corresponding to the vertices outside the embedded subgraph.

A. 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 $\mathcal{H}_0$ belongs to the class of random matrices called Wigner matrices. It consists of symmetric matrices with independent upper triangular entries with zero mean and equal variances [24]. The spectral properties of centered adjacency matrices 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., [22], [25]. The eigenvectors of Wigner matrices have also been studied under general distributions of matrix entries. For finite n it is known that the eigenvectors are Haar distributed for Wigner matrices with Gaussian entries such as the Gaussian Unitary ensemble and the Gaussian Orthogonal Ensemble [24]. This means that a typical eigenvector 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 with independent and identically distributed (i.i.d.) components drawn from $\mathcal{N}(0,1)$, normalized to have unit L^2-norm, i.e., $\mathbf{x}/\|\mathbf{x}\|$, with $\mathbf{x}$ being a $\mathbb{R}^n$ Gaussian vector with covariance matrix $\mathbf{I}$, i.e., $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I})$. In [12], the joint distribution of a subset of scaled eigenvector components was shown to be i.i.d. Gaussian under a restriction on the size of the subset. In [13] and [20] the authors studied spectral functions that depend on the eigenvectors for general Wigner matrices and showed them to converge to a Brownian bridge as n scales to infinity, thus providing evidence that the eigenvectors are Haar distributed. In [26] the Brownian Bridge property was extended to Wigner matrices with heavy tailed entries. Based on these results we make the following approximation for the centered adjacency matrix of an ER graph.

Approximation 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 $S^{(n-1)}$ when $p_b \gg \frac{\log^6(n)}{n}$. 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 $g(\mathbf{x}) = \|\mathbf{x}\|_1/\|\mathbf{x}\|$. Below we derive a central limit theorem for $g(\mathbf{x})$, when $\mathbf{x}$ a Gaussian random vector with i.i.d. components.

Lemma 1: (Central Limit Theorem for $\|\mathbf{x}\|_1/\|\mathbf{x}\|$) Let $\mathbf{x}$ be a Gaussian random vector with 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 $\alpha_1 = \mathbb{E}(|x_1|)$, $\alpha_2 = \mathbb{E}(|x_1|^2)$, $C_{11} = \text{Var}(|x_1|), C_{22} = \text{Var}(|x_1|^2), C_{12} = \mathbb{E}((|x_1| - \mathbb{E}(|x_1|))(|x_1|^2 - \mathbb{E}(|x_1|^2)))$, i.e., $g(\mathbf{x}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mu_0, \sigma_0^2)$).

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

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

$$\mathbb{E}((|x_1| - \mathbb{E}(|x_1|))(|x_1|^2 - \mathbb{E}(|x_1|^2))) = \sqrt{\tfrac{2}{\pi}}. \qquad \square$$

B. Spectral Statitistics under $\mathcal{H}_1$

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

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

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

$$\overline{\boldsymbol{\mathcal{A}}} \coloneqq \mathbb{E}_{\mathcal{H}_1}\boldsymbol{\mathcal{A}} = \begin{bmatrix} (p_1 - p_b)\mathbf{J}_m & \mathbf{0}_{m\times n-m} \\ \mathbf{0}_{n-m\times m} & \mathbf{0}_{n-m\times n-m} \end{bmatrix}. \quad (4)$$

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

The matrix $\overline{\boldsymbol{\mathcal{A}}}$ has rank 1, and with a single non-zero eigenvalue $\bar{\lambda} = m\delta_p$, with eigenvector $\bar{\mathbf{u}} = \frac{1}{\sqrt{m}}\begin{bmatrix} \mathbf{1}_m \\ \mathbf{0}_{n-m\times 1} \end{bmatrix}$. In order to show the CLT presented in Theorem 2 we need a bound on the spectral radius of $\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}$, which is presented below.

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

$$\|\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}\| < \sqrt{12\log(n)\max(\sigma_1^2 m + \sigma_0^2(n-m), \sigma_0^2 n)} \quad (5)$$
$$= \sqrt{12\log(n)\sigma^2} \text{ almost surely (a.s.),}$$

where $\sigma_1^2 = p_1(1-p_1), \sigma_0^2 = p_0(1-p_0)$, and define $\sigma^2 \coloneqq \max(\sigma_1^2 m + \sigma_0^2(n-m), \sigma_0^2 n)$.
For the above result to hold we require that $\exists N$ s.t. $\forall n > N$ $\sigma^2 > (6\log(n))^2$, which can be easily verified to be satisfied when $np_b \gg \log^2(n)$.

Let us define $\Delta \coloneqq \frac{\sqrt{12\log(n)np_b}}{m\delta_p}$. Note that under Condition 3 $\Delta = \mathrm{o}(1)$, implying that $\|\boldsymbol{\mathcal{A}} - \overline{\boldsymbol{\mathcal{A}}}\|/\|\overline{\boldsymbol{\mathcal{A}}}\| \to 0$. This means that asymptotically the spectral characteristics of $\boldsymbol{\mathcal{A}}$ and $\overline{\boldsymbol{\mathcal{A}}}$ are nearly the same. We use this fact to show that $\mathbf{u}_1$ converges to $\bar{\mathbf{u}}$, which is instrumental in the proof of Theorem 2 (Details in the research report [23]).

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 [27], for the components of the eigenvector of a single dimensional Random Dot Product Graph(RDPG). See our technical report 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*[1] of $\boldsymbol{\mathcal{A}}$, which we denote $\mathbf{u} \coloneqq \mathbf{u}_1$, corresponding to the eigenvalue $\lambda \coloneqq \lambda_1$. Observe that the mean matrix $\overline{\boldsymbol{\mathcal{A}}}$ can be written as $\bar{\mathbf{x}}\bar{\mathbf{x}}^T$, where $\bar{\mathbf{x}} = \sqrt{\delta_p}\begin{bmatrix} \mathbf{1}_m^T & \mathbf{0}_{n-m}^T \end{bmatrix}^T$ and $\bar{\mathbf{u}} = \bar{\mathbf{x}}/\|\bar{\mathbf{x}}\|$. Let us define $\mathbf{x}$ as $\mathbf{x} = \lambda^{1/2}\mathbf{u}$, and so $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$. 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 present below our main theorem on the CLT of the components of the dominant eigenvectors.

Theorem 2: Under Conditions 2, 3 and 4 the following CLT holds true for the entries of the scaled dominant eigenvector $\mathbf{x} = \lambda^{1/2}\mathbf{u}$, where $\mathbf{u}$ is the eigenvector corresponding to the eigenvalue λ of $\boldsymbol{\mathcal{A}}$ under $\mathcal{H}_1$.

$$\sqrt{\frac{m\delta_p}{p_1(1-p_1)}}\left(x_i - \sqrt{\delta_p}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad (6)$$

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

$$\sqrt{\frac{m\delta_p}{p_b(1-p_b)}}x_i \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad (7)$$

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

Sketch of Proof of Theorem 2:
For the full proof, please see the research report [23]; we only provide a brief outline here. 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 $\bar{x}_i = \frac{1}{\lambda^{1/2}}[\overline{\boldsymbol{\mathcal{A}}}\bar{\mathbf{u}}]_i = \sqrt{\delta_p}$ 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, using bounds on $\|\mathbf{u} - \bar{\mathbf{u}}\|$ derived using eigenvector perturbation lemmas in [28] derived using Theorem 1.
- 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 $\mathcal{N}(0,1)$, under Condition 4.
- Finally we show that $T_3 = \gamma_i\left((\frac{1}{\lambda^{1/2}} - \frac{1}{\bar{\lambda}^{1/2}})[\overline{\boldsymbol{\mathcal{A}}}\bar{\mathbf{u}}]_i\right) \to 0$, in probability under Condition 3, by showing a concentration result for the dominant eigenvalue λ. Notice that $T_3 = 0$ for $i > m$.

The result thus follows by an application of Slutsky's thereom [29].

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 Theorem 2 for finite n as opposed to the asymptotic regime in which Theorem 2 holds. We expect this to be a good approximation for large n, as our simulations in Section IV indicate.

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

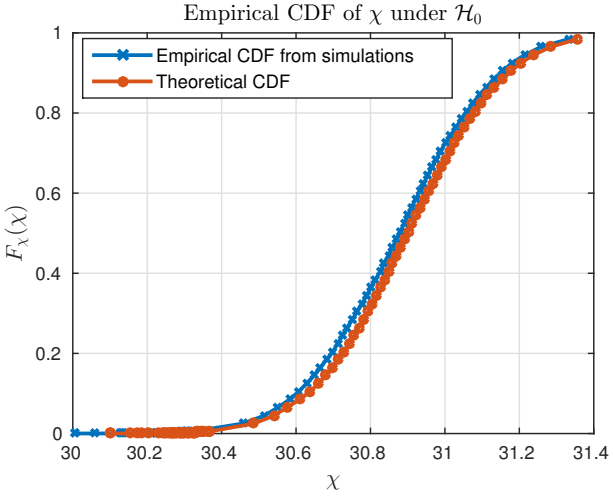

Fig. 1. CDF of χ under $\mathcal{H}_0$

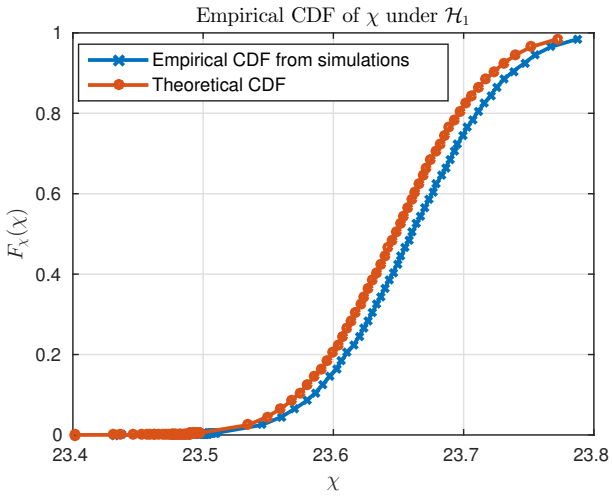

Fig. 2. CDF of χ under $\mathcal{H}_1$.

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 $\mathcal{N}(0,1)$ with μ_1 and $\sigma^2_{(1)}$ given in (8,9) respectively. Note that $Q(x)$ is the Q-function, i.e., the Complimentary Cumulative Density Function (CDF) of the an Random Variable (r.v) 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 [30]. Then

$$\mu_{(1)} = \frac{N_{\alpha_1}}{\sqrt{N_{\alpha_2}}} \tag{8}$$

and

$$\sigma^2_{(1)} = \frac{1}{N_{\alpha_2}}\left(C_{11} + \left(\frac{N_{\alpha_1}}{2N_{\alpha_2}}\right)^2 C_{22} - \frac{N_{\alpha_1}}{N_{\alpha_2}}C_{12}\right), \tag{9}$$

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 $\sigma^2_{(1)}$ in (8) and (9), using the fact that $r = \Omega(1)$, and $s = \Omega(1)$. For the parameter values we choose satisfying Conditions 1,3 and 4, and using asymptotic approximations for the Q-function and $M(a,b,x)$, [30] 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 $o(1)$ 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 [30], 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.

IV. 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 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 10^4 times and compute the empirical CDF $F_\chi(\chi)$, which is the solid blue line with "x" marker in figure 1. In the 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.

V. 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 $\theta(\sqrt{np_b})$ [9]. We would like to investigate the possibility of detecting the presence of the anomaly under a much more stringent regime, where it might not be information-theoretically possible to detect the subgraph nodes.

References

[1] B. Miller, M. Beard, P. Wolfe, and N. Bliss, "A spectral framework for anomalous subgraph detection," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4191–4206, 2015.

[2] B. Miller, N. Bliss, and P. Wolfe, "Subgraph detection using eigenvector l1 norms," in *NIPS 2010*, 2010.

[3] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 626–688, 2015.

[4] T. L. Mifflin, C. Boner, G. A. Godfrey, and J. Skokan, "A random graph model for terrorist transactions," in *Aerospace Conference, 2004. Proceedings. 2004 IEEE*, vol. 5. IEEE, 2004, pp. 3258–3264.

[5] 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

[6] A. Juels and M. Peinado, "Hiding cliques for cryptographic security," *Designs, Codes and Cryptography*, vol. 20, no. 3, pp. 269–280, 2000.

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

[8] 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.

[9] 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.

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

[11] 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.

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

[13] 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.

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

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

[16] 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.

[17] 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.

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

[19] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

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

[21] 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.

[22] K. Avrachenkov, L. Cottatellucci, and A. Kadavankandy, "Spectral properties of random matrices for stochastic block model," in *WiOpt*. IEEE, 2015, pp. 537–544.

[23] 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

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

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

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