

The spectral gap of a random subgraph of a graph

Fan Chung ^{*†}

Paul Horn ^{*}

Abstract

We examine the relationship of a graph G and its random subgraphs which are defined by independently choosing each edge with probability p . Suppose that G has a spectral gap λ (in terms of its normalized Laplacian) and minimum degree d_{\min} . Then we can show that a random subgraph of G on n vertices with edge-selection probability p almost surely has as its spectral gap $\lambda - O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)$.

1 Introduction

Often, when we examine a large graph, perhaps arising from some realistic setting (e.g. webgraphs, biological networks, or some information network derived from a large database), we are unable to see the entire graph. Instead, what we can observe are relatively small subgraphs of the large graph. We are interested, then, in understanding the relationship between the large host graph, and the subgraphs that we actually observe. The basic question becomes whether there are some graph invariants that we can evaluate for the original host graph which lead to good estimates for properties and structures of our observed subgraphs, and vice-versa.

In this paper, we begin by considering a host graph G . Fixing an edge-selection parameter, $p \in [0, 1]$, we consider the family of subgraphs by percolating G with parameter p . That is, our observed subgraph H is a random subgraph of G such that each edge from G is chosen independently with the edge-selection probability p .

An example of the utility of such an approach is in the study of epidemiological models. Here, the host graph represents a contact network where the vertices represent people in some community of interest, and edges denote certain contact or interaction among pairs of people. A disease is often considered to pass through an interaction with probability p . Thus, the observed random subgraph can represent the actual spread of disease through the contact network. Other examples concern various social networks such as telephone or instant messaging networks. A group of friends contacting each other during a specified period of time in a large social network can be viewed as a random subgraph of a large host graph consisting of all contacts between members of the network. The classical Erdős-Rényi model, $G(n, p)$, is also a particular instance of this model. Indeed, it is the special case where the host graph is the complete graph K_n .

In this paper, we consider the relationship of the spectrum of the host graph and that of a random subgraph with edge-selection probability p . The methods that we use here are based on Wigner's high moment methods [9]. Such an approach has been extensively utilized in the early work on random graphs and matrices in numerous research papers including the early work by Füredi and Komlós [5] as well as in some recent work on random sparse graphs in [2] and [8]. However, all the previous work belong to the special case that the host graph is taken to be the complete graph (or the full matrix). Here similar

^{*}University of California, San Diego

[†]Research supported in part by NSF Grants DMS 0457215 and ITR 0426858

techniques are used and modified in order to deal with the spectral gap of a random subgraph of a given host graph.

Since we are dealing with a general graph (with possibly uneven degree distribution), we consider the (normalized) Laplacian (see [1]). For a graph on n vertices, the Laplacian is a symmetric matrix of size $n \times n$ defined as follows: (More details will be given in the next section.)

$$\mathcal{L} = I - D^{-1/2}AD^{-1/2},$$

where A denotes the adjacency matrix and D denotes the diagonal degree matrix. Let

$$0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$$

denote the eigenvalues of \mathcal{L} . We write

$$\lambda = \min\{\lambda_1, 2 - \lambda_{n-1}\}.$$

For example, a random d -regular graph on n vertices almost surely has a spectral gap $\lambda > 1 - (2\sqrt{d-1} + \epsilon)/d$ for any positive constant ϵ as n approaches infinity (see [4]).

When there is a nontrivial spectral gap, i.e., λ is separated from 0, the graph has many nice properties, such as expansion properties and the rapid convergence of random walks on the graph. What we will show here is how to derive a bound for the spectral gap of a random subgraph of G in terms of the probability p of edge-selection and the spectral gap of G .

Theorem 1. *Suppose G is a graph on n vertices with spectral gap λ and minimum degree d_{\min} . A random subgraph H of G with edge-selection probability p almost surely has a spectral gap λ_H satisfying*

$$\lambda_H = \lambda - O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right).$$

An equivalent statement for the above theorem is the following: For $pd_{\min} \geq (\log n)^2/(\log \log n)^3$, we have

$$\lambda_H = \lambda - O\left(\sqrt{\frac{\log n}{pd_{\min}}}\right)$$

and, for $pd_{\min} < (\log n)^2/(\log \log n)^3$, we have

$$\lambda_H = \lambda - O\left(\frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right).$$

As immediate consequences of the above theorem, a random graph almost surely will have the following properties (see [1, 7]):

Corollary 1. *For a graph G on n vertices with spectral gap λ and minimum degree d_{\min} , a subgraph H with edge-selection probability p almost surely satisfies the following properties:*

1. H satisfies the expansion property as follows: For $X \subseteq V(H)$, the number of edges in H leaving X , denoted by $\partial_H(X)$ satisfies

$$\partial_H(X) \geq \left(\lambda - O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)\right) p \text{vol}(X)$$

where $\text{vol}(X)$ is $\sum_{x \in X} d(x)$ and $d(x)$ denotes the degree of x in the host graph G .

2. H satisfies the discrepancy property as follows: For $X, Y \subseteq V(H)$, the number of edges of H between X and Y , denoted by $e_H(X, Y)$, satisfies

$$|e_H(X, Y) - p \frac{\text{vol}(X)\text{vol}(Y)}{\text{vol}(H)}| \leq \left(1 - \lambda - O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)\right) p\sqrt{\text{vol}(X)\text{vol}(Y)}.$$

3. For a random walk on H with transition probability matrix P_H , the total variation distance after t steps from the stationary distribution π , denoted by $\Delta_{TV}(t)$, is bounded above by

$$\begin{aligned} \Delta_{TV}(t) &= \max_{A \subseteq V} \max_{y \in V} \left| \sum_{x \in A} (P_H^t(y, x) - \pi(x)) \right| \\ &\leq e^{-c} \end{aligned}$$

for any $c > 0$ provided t satisfies

$$t \geq \frac{1}{\lambda - O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)} \left(\log \frac{\text{vol}(H)}{\min_x d(x)} + c \right).$$

The above theorem depends on the volume $\text{vol}(X)$ of a subset X in G . However it can also be thought of as a statement depending on the volume of X in H in the following manner. The volume of X in H is denoted by $\text{vol}_H(X) = \sum_{x \in X} d_H(x)$, where $d_H(x)$ is the degree of x in the subgraph H . If the volume of X in G is large, we have good control over the volume of X in H using Chernoff bounds. It is not difficult to prove that if $\text{Vol}(X)$ is sufficiently large, we almost surely have (see [3])

$$|\text{vol}_H(X) - p\text{vol}(X)| \leq g(n)\sqrt{p\text{vol}(X)} \quad (1)$$

for any function $g(n)$ that goes to infinity with n .

2 Preliminaries

Let $G = (V, E)$ be a graph. We denote by H a random graph obtained from G by taking each edge independently with probability p . That is,

$$\Pr(\{u, v\} \in E(H)) = \begin{cases} p & \text{if } \{u, v\} \in E(G), \\ 0 & \text{otherwise.} \end{cases}$$

Let A and A_H denote the adjacency matrix of G and H , respectively. We denote the diagonal matrices whose entries consist of the degrees of the vertices in G and H respectively, by D and D_H . Let $0 = \eta_0 \leq \dots \leq \eta_{m-1}$ denote the eigenvalues of \mathcal{L}_H , and let φ_i for $i = 0, \dots, m-1$ denote a set of orthonormal eigenvectors associated with the η_i (represented here as row vectors). The projection to φ_i , for each i , is $P_i = \varphi_i^* \varphi_i$ where φ^* denotes the transpose of φ . Then we have

$$\mathcal{L}_H = \sum_i \eta_i P_i.$$

We consider, then, the matrix

$$\begin{aligned} M &= I - \mathcal{L}_H - P_0 \\ &= \sum_{i \neq 0} (1 - \eta_i) P_i. \end{aligned}$$

We will use the fact that for any integer k , we have

$$\text{Trace}(M^{2k}) = \sum_{i \neq 0} (1 - \eta_i)^{2k}.$$

Immediately we have the following:

Fact 1. *For any positive integer k*

$$\max_{i \neq 0} |1 - \eta_i| = \|M\| \leq (\text{Trace}(M^{2k}))^{1/(2k)}.$$

We denote the spectral gap of H by $\lambda_H = \max\{\eta_1, 2 - \eta_{n-1}\}$. Hence we have

$$\lambda_H = 1 - \max_{i \neq 0} |1 - \eta_i| = 1 - \|M\|.$$

Let K denote the all ones matrix. We can rewrite M as:

$$\begin{aligned} M &= D_H^{-1/2} A_H D_H^{-1/2} - P_0 \\ &= D_H^{-1/2} A_H D_H^{-1/2} - \varphi_0^* \varphi_0 \\ &= D_H^{-1/2} A_H D_H^{-1/2} - \frac{1}{\text{vol}(H)} D_H^{1/2} K D_H^{1/2}. \end{aligned}$$

Instead of directly dealing with M , we consider the simpler matrix

$$\begin{aligned} C &= p^{-1} D^{-1/2} A_H D^{-1/2} - \frac{1}{p \text{vol}(G)} p D^{1/2} K D^{1/2} \\ &= p^{-1} D^{-1/2} A_H D^{-1/2} - \frac{1}{\text{vol}(G)} D^{1/2} K D^{1/2}. \end{aligned} \tag{2}$$

where one may note that $p \text{vol}(G)$ is the expected volume of H .

In a way, C can be thought of as an estimate for the expectation of M . Our plan is to first carefully consider $\|C\|$ in the next section, and then bound the norm of the difference between M and C in Section 4.

3 A bound on $\|C\|$

In this section we prove the following theorem.

Theorem 2. *Let G be a given graph with spectral gap λ and minimum degree d_{\min} . Let H be a random subgraph of G with edge-selection probability p . Then the matrix C as defined in (2) almost surely satisfies*

$$\|C\| = 1 - \lambda + O\left(\sqrt{\frac{\log n}{p d_{\min}}} + \frac{(\log n)^{3/2}}{p d_{\min} (\log \log n)^{3/2}}\right).$$

Proof. To bound the norm of C , we express C as a sum of two parts:

$$C = B + M'$$

where

$$B = p^{-1} D^{-1/2} A_H D^{-1/2} - D^{-1/2} A D^{-1/2}$$

and

$$M' = D^{-1/2}AD^{-1/2} - \frac{1}{\text{vol}(G)}D^{1/2}KD^{1/2}.$$

Note that M' is equivalent to the matrix M for the graph G . Hence we have

$$\|M'\| = 1 - \lambda.$$

It suffices to show that almost surely we have $\|B\| = O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)$. In other words, we wish to prove that for any $\epsilon > 0$, there is an absolute constant c so that for n sufficiently large, we can bound the following probability as follows:

$$\Pr \left[\|B\| \geq c \left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}} \right) \right] \leq \epsilon.$$

The matrix B is a random matrix, where the entries b_{ij} are independent random variables defined by:

$$b_{ij} = \begin{cases} \frac{1}{p\sqrt{d_i d_j}} - \frac{1}{\sqrt{d_i d_j}} = \frac{1-p}{p\sqrt{d_i d_j}} & \text{if } v_i \sim v_j \in H \\ -\frac{p}{p\sqrt{d_i d_j}} & \text{if } v_i \not\sim v_j \in H \text{ and } v_i \sim v_j \in G \\ 0 & \text{otherwise.} \end{cases}$$

Here d_i denotes the degree of v_i in G . It follows from the definition that the expected value of b_{ij} satisfies

$$\mathbb{E}[b_{ij}] = 0.$$

Now consider the (i, i) th entry of B^{2k} . A typical term of such an entry is of the form

$$b_{i_1 i_2} b_{i_2 i_3} \dots b_{i_{2k-1} i_{2k}}$$

with $i_1 = i_{2k} = i$. Assuming this term is non-zero, this corresponds to a closed walk in G starting and ending at v_i . Taking expectations, we note that $\mathbb{E}[b_{i_1 i_2} b_{i_2 i_3} \dots b_{i_{2k-1} i_{2k}}] \neq 0$ only if each b_{ij} occurs at least twice (since $\mathbb{E}[b_{ij}] = 0$ and all b_{ij} 's are independent). In other words, each edge must occur at least twice in the closed walk. We refer to a such a closed walk, which contributes to the expected trace, as a *surviving walk*.

To determine the expected contribution of a surviving walk to $\text{Trace}(B^{2k})$, we consider the expected value $\mathbb{E}[b_{ij}^m]$. Note that for $m \geq 2$

$$|\mathbb{E}[b_{ij}^m]| \leq \frac{|(1-p)^m p + (-p)^m (1-p)|}{p^m (d_i d_j)^{m/2}} \leq \frac{p}{p^m (d_i d_j)^{m/2}} = \frac{1}{p^{m-1} (d_i d_j)^{m/2}}.$$

The last inequality follows from the easy fact that $|(1-p)^m p + (-p)^m (1-p)| \leq p$ when $p \leq 1$.

To bound the trace of B^{2k} , we must get a handle on the number of surviving walks and their contribution to the trace. Consider a surviving walk of length $2k$ on vertices v_1, \dots, v_{l+1} and let us assume that the vertices are labelled by their first occurrence in the walk. Thus to get to vertex v_i , we must have followed an edge from one of v_1, \dots, v_{i-1} . We define the *exposure sequence* of the walk to be a vector (a_1, \dots, a_i) such that we first travel to vertex v_i from vertex $v_{a_{i-1}}$. Clearly, $a_i \in \{1, \dots, i\}$. Hence there are at most $i!$ possible exposure sequences. We seek to enumerate our surviving walks by their exposure sequences.

Consider a surviving walk on vertices v_1, \dots, v_{l+1} with exposure sequence (a_1, \dots, a_l) . Let us assume that the walk contains edges e_1, \dots, e_k with multiplicities m_1, \dots, m_k , respectively. Then the contribution of the walk to the expected value of the trace is at most

$$\begin{aligned} \mathbb{E}[b_{e_1}^{m_1} \dots b_{e_m}^{m_m}] &= \prod_i \mathbb{E}[b_{e_i}^{m_i}] \leq \prod_{\substack{e_i \\ e_i = \{v_{i1}, v_{i2}\}}} \frac{1}{p^{m_i-1} (d_{i1} d_{i2})^{m_i/2}} \\ &\leq \frac{1}{\left(\prod_{i=1}^l d_{a_i}\right) (pd_{\min})^{2k-l}} \end{aligned}$$

where the d_{a_i} term comes from the fact that there must exist an edge contributing $1/(pd_{a_i})$ to the product since this edge must occur with multiplicity at least 2.

Given a set of vertices $S = \{v_1, \dots, v_{l+1}\}$ and an exposure sequence $\mathbf{e} = (a_1, \dots, a_l)$, let $W(S, \mathbf{e}, k)$ denote the number of surviving walks of length $2k$ on vertices S with exposure sequence \mathbf{e} . We can upper bound the number of surviving walks by the number of surviving walks on these vertices in a complete graph of the same size. Let $W'(k, l)$ denote the number of surviving walks of length $2k$ on the complete graph K_{l+1} such that the vertices are visited in order v_1, \dots, v_{l+1} . (Clearly the labeling does not affect the number of paths; just the fact that the vertices are visited in a particular order.) We note that for a given set S with $|S| = l + 1$, there can be at most $l!$ exposure sequences. Furthermore, for a set S with $l + 1$ vertices and an exposure sequence \mathbf{e} , we have

$$W(S, \mathbf{e}, k) \leq W'(k, l).$$

This inequality is immediate, as each surviving walk on $G|_S$ corresponds injectively to a walk on the complete graph K_{l+1} . Further note that this inequality holds independently of the exposure sequence \mathbf{e} .

Füredi and Komlós [5] gave an upper bound on $W'(k, l)$. Recently this bound was improved by Vu, and it is this new bound we use. Inequality (9) in Vu [8] asserts

Lemma 1.

$$W'(k, l) \leq \binom{2k}{2l} 2^{k+2(k-l)+1} (l+1)^{3(k-l)}$$

We can now bound the expected value of $\text{Trace}(B^{2k})$ by both applying the above bound for $W'(k, l)$ and using the fact that we are counting surviving walks. We use the notation $u \sim v$ to denote that u and v are adjacent in G .

$$\begin{aligned}
\mathbb{E}[\text{Trace}(B^{2k})] &\leq \sum_{l=1}^k \sum_{S=\{v_1, \dots, v_{l+1}\}} \sum_{\mathbf{e}=(a_1, \dots, a_l)} W(S, \mathbf{e}, k) \frac{1}{\left(\prod_{i=1}^l d_{a_i}\right) (pd_{\min})^{2k-l}} \\
&\leq \sum_{l=1}^k \sum_{\mathbf{e}=(a_1, \dots, a_l)} \sum_{v_1} \sum_{v_2 \sim v_{a_1}} \sum_{v_3 \sim v_{a_2}} \cdots \sum_{v_{l+1} \sim v_{a_l}} W(S, \mathbf{e}, k) \frac{1}{\left(\prod_{i=1}^l d_{a_i}\right) (pd_{\min})^{2k-l}} \\
&\leq \sum_{l=1}^k \sum_{\mathbf{e}=(a_1, \dots, a_l)} \sum_{v_1} \sum_{v_2 \sim v_{a_1}} \sum_{v_3 \sim v_{a_2}} \cdots \sum_{v_{l+1} \sim v_{a_l}} W'(k, l) \frac{1}{\left(\prod_{i=1}^l d_{a_i}\right) (pd_{\min})^{2k-l}} \\
&= \sum_{l=1}^k \sum_{\mathbf{e}=(a_1, \dots, a_l)} \sum_{v_1} \sum_{v_2 \sim v_{a_1}} \sum_{v_3 \sim v_{a_2}} \cdots \sum_{v_{l+1} \sim v_{a_{l-1}}} W'(k, l) \frac{1}{\left(\prod_{i=1}^{l-1} d_{a_i}\right) (pd_{\min})^{2k-l}} \\
&\vdots \\
&= \sum_{l=1}^k \sum_{\mathbf{e}=(a_1, \dots, a_l)} \sum_{v_1} W'(k, l) \frac{1}{(pd_{\min})^{2k-l}} \\
&\leq \sum_{l=1}^k \sum_{\mathbf{e}=(a_1, \dots, a_l)} n \binom{2k}{2l} 2^{k+2(k-l)+1} (l+1)^{3(k-l)} \frac{1}{(pd_{\min})^{2k-l}} \\
&\leq \sum_{l=1}^k l! n \binom{2k}{2l} 2^{3k} (l+1)^{3(k-l)} \frac{1}{(pd_{\min})^{2k-l}} \\
&\leq \sum_{l=1}^k l^l n \binom{2k}{2l} 2^{3k} (l+1)^{3(k-l)} \frac{1}{(pd_{\min})^{2k-l}} \\
&\leq \sum_{l=1}^k n 3^{2k} (l+1)^{3k-2l} \frac{1}{(pd_{\min})^{2k-l}} \\
&= \sum_{l=1}^k n 3^{2k} s_{l,k}
\end{aligned}$$

where we define

$$s_{l,k} = \frac{(l+1)^{3k-2l}}{(pd_{\min})^{2k-l}}.$$

For a fixed $\epsilon > 0$, with $\epsilon < 1/4$ we now choose

$$k = \lceil \log n + \log(1/\epsilon) \rceil \tag{3}$$

and set

$$pd_{\min} = \alpha \left(\frac{k}{\log k} \right)^{3/2}.$$

Note that α is a function of k (and hence n). We wish to show the following:

Claim:

$$s_{l,k} \leq \left(\frac{c}{\alpha \min\left\{\alpha, \frac{k^{1/2}}{\log^{3/2} k}\right\}} \right)^k \tag{4}$$

for some absolute constant c .

We let

$$\begin{aligned}
f(l) &:= \frac{s_{l,k}}{s_{l-1,k}} = \frac{(1+l^{-1})^{3k-2l} pd_{\min}}{l^2} \\
&= c_0 \frac{e^{3k/l} pd_{\min}}{l^2}
\end{aligned} \tag{5}$$

where c_0 is upper and lower bounded by some absolute constants. For a given value of pd_{\min} and for the range of $0 \leq l \leq k$, the function $s_{l,k}$ either attains its maximum at l_0 satisfying $l_0 = k$ and $f(k) > 1$, or l_0 is one of the two integers closest to the solution of $f(l) = 1$. Note that for the first case, we have

$$\begin{aligned}
s_{l_0,k} = s_{k,k} &\leq c \left(\frac{k+1}{pd_{\min}} \right)^k \\
&\leq c \left(\frac{k}{\alpha k^{3/2} \log^{-3/2} k} \right)^k \\
&= c \left(\frac{1}{\alpha k^{1/2} \log^{-3/2} k} \right)^k
\end{aligned}$$

which implies (4). We may assume that l_0 is one of the two integers closest to the solution of $f(l) = 1$. Furthermore, for $l < k/(2 \log k)$, we have, from (5),

$$\frac{s_{l,k}}{s_{l-1,k}} > \frac{e^{2 \log k}}{k^2} > 1.$$

Therefore we may assume that

$$l_0 \geq \frac{k}{2 \log k}.$$

There are two possibilities:

Case 1: $l_0 \leq 100k/\log k$. Then

$$\begin{aligned}
s_{l_0,k} &\leq \left(c' \frac{l_0^3}{(pd_{\min})^2} \right)^k \\
&\leq \left(c \frac{k^3 / (\log k)^3}{\alpha^2 (k/\log k)^3} \right)^k \\
&= \left(\frac{c}{\alpha^2} \right)^k
\end{aligned}$$

which implies (4).

Case 2: $l_0 \geq 100k/\log k$. We use the fact that l_0 is one of the two integers closest to the solution of $f(l) = 1$. From equation (5), we have:

$$\begin{aligned}
s_{l_0,k} &\leq \left(\frac{l_0^{3-2l_0/k}}{(pd_{\min})^{2-l_0/k}} \right)^k \\
&\leq \left(c'' \frac{l_0^{3-2l_0/k}}{(l_0^2 e^{-3k/l_0})^{2-l_0/k}} \right)^k \\
&\leq \left(\frac{c''' e^{6k/l_0}}{l_0} \right)^k
\end{aligned} \tag{6}$$

One can check that for the given range of l_0 this satisfies:

$$\begin{aligned}
(6) &\leq \left(\frac{c'''k}{l_0^2}\right)^k \\
&\leq \left(\frac{c'''ke^{3k/l_0}}{l_0^2}\right)^k \\
&\leq \left(\frac{c_0}{\alpha k^{1/2} \log^{-3/2} k}\right)^k
\end{aligned}$$

which again implies (4). In this section, c_0, c, c', c'', \dots are suitably chosen integers. The proof for the claim is completed.

We are now ready to consider bounding the norm of B .

$$\begin{aligned}
\mathbb{E}[\text{Trace}(B^{2k})] &\leq n^2 32^k \max_l s_{l,k} \\
&\leq 2n^2 \left(\frac{32c_0}{\alpha \min\{\alpha, k^{1/2} \log^{-3/2} k\}}\right)^k.
\end{aligned}$$

Since $\mathbb{E}[\|B\|^{2k}] \leq \mathbb{E}[\text{Trace}(B^{2k})]$, we have

$$\mathbb{E}[\|B\|^{2k}] \leq 2n^2 \left(\frac{c}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right)^{2k}.$$

By the previous equation and Markov's equality, we have

$$\begin{aligned}
\Pr\left(\|B\| \geq 2 \frac{c}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right) &= \Pr\left(\|B\|^{2k} \geq 2^{2k} \left(\frac{c}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right)^{2k}\right) \\
&\leq \frac{\mathbb{E}[\|B\|^{2k}]}{2^{2k} \left(\frac{c}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right)^{2k}} \\
&\leq \frac{2n^2 \left(\frac{c}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right)^{2k}}{2^{2k} \left(\frac{c}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right)^{2k}} \\
&\leq \frac{2n^2}{2^{2k}} \\
&\leq \epsilon
\end{aligned}$$

for the given $\epsilon > 0$ (noting this holds as $\epsilon < 1/4$) and our choice of k in (3). Hence we have proved that almost surely we have

$$\|C\| \leq \|M'\| + \|B\| = 1 - \lambda + O\left(\frac{1}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right).$$

In a similar way, we can use the fact that $\|C\| \geq \|M'\| - \|B\|$ to get

$$\|C\| = 1 - \lambda \pm O\left(\frac{1}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right).$$

It is easily verified that

$$O\left(\frac{1}{\min\{\alpha, \alpha^{1/2} k^{1/4} \log^{-3/4} k\}}\right) = O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)$$

which completes the proof of Theorem 2. □

4 Bounding the Spectral Gap

In this section, we plan to give a complete proof for Theorem 1. Namely, we wish to show that for a graph G with spectral gap λ and minimum degree d_{\min} , a random subgraph H obtained from G with edge-selection probability p almost surely has eigenvalues of the Laplacian \mathcal{L}_H of H satisfying:

$$\lambda_H = \max_{i \neq 0} |1 - \eta_i| = 1 - \lambda + O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right).$$

As a matter of notation, we let d_i refer to the degree of vertex v_i in G and d'_i refer to the degree of vertex i in H . We also let a'_{ij} refer to the ij th entry of A_H , the adjacency matrix of H .

To prove Theorem 1, recall that the eigenvalues of the Laplacian \mathcal{L}_H satisfy

$$\max_{i \neq 0} |1 - \eta_i| = \|M\|$$

where $M = D_H^{-1/2} A_H D_H^{-1/2} - \frac{1}{\text{vol}(H)} D_H^{1/2} K D_H^{1/2}$. Now we write:

$$M = E + C + R + S$$

where we define

$$\begin{aligned} E &= D_H^{-1/2} A_H D_H^{-1/2} - \frac{1}{p} D^{-1/2} A_H D^{-1/2} - \frac{p}{\text{vol}(G)} D D_H^{-1/2} K D_H^{-1/2} D + \frac{1}{\text{vol}(G)} D^{1/2} K D^{1/2} \\ R &= \frac{p}{\text{vol}(G)} D D_H^{-1/2} K D_H^{-1/2} D - \frac{1}{p \text{vol}(G)} D_H^{1/2} K D_H^{1/2} \\ S &= \left(\frac{1}{p \text{vol}(G)} - \frac{1}{\text{vol}(H)} \right) D_H^{1/2} K D_H^{1/2} \end{aligned}$$

and C is as defined in (2). Thus

$$\begin{aligned} e_{ij} &= \left(a'_{ij} - \frac{p d_i d_j}{\text{vol}(G)} \right) \left(\frac{1}{\sqrt{d'_i d'_j}} - \frac{1}{p \sqrt{d_i d_j}} \right) \\ r_{ij} &= \frac{1}{p \text{vol}(G)} \frac{p^2 d_i d_j - d'_i d'_j}{\sqrt{d'_i d'_j}} \\ s_{ij} &= \left(\frac{1}{p \text{vol}(G)} - \frac{1}{\text{vol}(H)} \right) \sqrt{d'_i d'_j} \end{aligned}$$

and

$$c_{ij} = \frac{a'_{ij}}{p \sqrt{d_i d_j}} - \frac{1}{\text{vol}(G)} \sqrt{d_i d_j}.$$

Clearly

$$\|M\| \leq \|E\| + \|C\| + \|R\| + \|S\|.$$

Hence, it suffices to establish the appropriate upper bounds for the norms of E, C, R and S .

To bound these, we use the following Chernoff bounds (see, e.g. [3]).

Lemma 2. *For $1 \leq i \leq n$, let X_i be independent random variables satisfying $|X_i| \leq M$. Let $X = \sum_i X_i$. Then we have, for any $a > 0$,*

$$P(|X - E[X]| > a) \leq e^{-\frac{a^2}{2(\text{Var}(X) + M a/3)}}. \quad (7)$$

We will prove the following lemma (whose proof we delay until after the proof of Theorem 1).

Lemma 3. *Assuming that $pd_{\min} \gg \log n$, almost surely every vertex v_i satisfies*

$$d'_i = pd_i(1 + O\left(\sqrt{\frac{\log n}{pd_{\min}}}\right)).$$

Let X_e , for $e \in E(G)$, be the random indicator variable which is 1 if $e \in H$ and 0 otherwise. We can write

$$\text{vol}(H) = \sum_{e \in E(G)} 2X_e. \quad (8)$$

We can show that almost surely

$$|\text{vol}(H) - p\text{vol}(G)| < 2\sqrt{p\text{vol}(H)g(n)} \quad (9)$$

for any function $g(n)$ that goes to infinity as n approaches infinity.

We also have the following lemma (whose proof will be given later).

Lemma 4. *Suppose that $pd_{\min} \gg \log n$. Almost surely the vector χ with $\chi(i) = (d'_i - pd_i)/\sqrt{pd_i}$ satisfies*

$$\|\chi\|^2 \leq (1 + o(1))n.$$

Proof of Theorem 1. We note that we already established a bound on $\|C\|$ in the last section. By Theorem 2, we have that almost surely

$$\|C\| = 1 - \lambda + O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right).$$

For convenience, we define

$$\beta = \sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}.$$

For $\|R\|$, almost surely we have the following by using equation (8) and the Cauchy-Schwartz inequality.

$$\begin{aligned} \|R\| &= \max_{\|y\|=1} \langle y, Ry \rangle \\ &\leq \max_{\|y\|=1} \frac{1}{p\text{vol}(G)} \sum_{i,j} y_i y_j \frac{d'_i(d'_j - pd_j) + (d'_i - pd_i)pd_j}{\sqrt{d'_i d'_j}} \\ &= \frac{1}{p\text{vol}(G)} \max_{\|y\|=1} \left\{ \sum_i \sqrt{d'_i} y_i \sum_j \frac{(d'_j - pd_j)y_j}{\sqrt{d'_j}} + \sum_i \frac{(d'_i - pd_i)y_i}{\sqrt{d'_i}} \sum_j \frac{pd_j y_j}{\sqrt{d'_j}} \right\} \\ &\leq \frac{1}{p\text{vol}(G)} \max_{\|y\|=1} \left\{ \left(\sum_i d'_i \right)^{1/2} \|y\| \left(\sum_j \frac{(d'_j - pd_j)^2}{d'_j} \right)^{1/2} + \|y\| \left(\sum_j \frac{(d'_j - pd_j)^2}{d'_j} \right)^{1/2} \left(\sum_j \frac{(pd_j)^2}{d'_j} \right)^{1/2} \right\} \\ &\leq (2 + o(1)) \sqrt{\frac{n}{p\text{vol}(G)}} \\ &\leq (1 + o(1)) \frac{2}{\sqrt{pd}} \\ &= o(\beta). \end{aligned}$$

For $\|S\|$, by using (9) and the Cauchy-Schwartz inequality we have:

$$\begin{aligned}
\|S\| &= \max_{\|y\|=1} \langle y, Sy \rangle = \max_{\|y\|=1} \sum_{i,j} |y_i y_j| \left(\frac{1}{p \operatorname{vol}(G)} - \frac{1}{\operatorname{vol}(H)} \right) \sqrt{d'_i d'_j} \\
&\leq \left(\frac{1}{p \operatorname{vol}(G)} - \frac{1}{\operatorname{vol}(H)} \right) \max_{\|y\|=1} \sum_{i,j} |y_i \sqrt{d'_i}| |y_j \sqrt{d'_j}| \\
&\leq \frac{2\sqrt{p \operatorname{vol}(G) \log(n)}}{p \operatorname{vol}(H) \operatorname{vol}(G)} \max_{\|y\|=1} \left(\sum_i |y_i \sqrt{d'_i}| \right)^2 \\
&\leq \frac{2\sqrt{p \operatorname{vol}(G) \log(n)}}{p \operatorname{vol}(H) \operatorname{vol}(G)} \left(\sum_i y_i^2 \right) \left(\sum_i d'_i \right) \\
&= O\left(\sqrt{\frac{\log n}{p \operatorname{vol}(G)}} \|y\|^2 \right) \\
&= o\left(\frac{\log n}{\sqrt{p d_{\min}}} \right) \\
&= o(\beta).
\end{aligned}$$

Finally, it remains to bound $\|E\|$. We recall

$$\begin{aligned}
e_{ij} &= \left(a'_{ij} - \frac{p d_i d_j}{\operatorname{vol}(G)} \right) \left(\frac{1}{\sqrt{d'_i d'_j}} - \frac{1}{p \sqrt{d_i d_j}} \right) \\
&= c_{ij} \frac{p \sqrt{d_i d_j} - \sqrt{d'_i d'_j}}{\sqrt{d'_i d'_j}}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
\|E\| &= \max_{\|y\|=1} \langle y, Ey \rangle \\
&= \max_{\|y\|=1} \sum_{i,j} y_i y_j c_{i,j} \frac{\sqrt{d'_i} (\sqrt{d'_j} - \sqrt{p d_j}) + (\sqrt{d'_i} - \sqrt{p d_i}) \sqrt{p d_j}}{\sqrt{d'_i d'_j}}.
\end{aligned}$$

Let $y'_i = y_i (\sqrt{d'_i} - \sqrt{p d_i}) / \sqrt{d'_i}$ and $y''_i = y_i \sqrt{p d_i} / \sqrt{d'_i}$. Then we have almost surely

$$\begin{aligned}
\|E\| &\leq \max_{\|y\|=1} \langle y, C y' \rangle + \langle y', C y'' \rangle \\
&\leq \max_{\|y\|=1} \|C\| \|y'\| + \|C\| \|y'\| \|y''\| \\
&\leq O(\beta).
\end{aligned}$$

This last observation follows from Lemma 3, which implies

$$\|y'\| \leq \|y\| \left(1 - \frac{1}{\sqrt{1 + O(\sqrt{\log n / p d_{\min}})}} \right) = O(\beta)$$

and, $\|y''\| = (1 + o(1)) \|y\| = O(1)$. Note that we have already observed that $\|C\| = O(1)$. Combining these results, we have

$$\begin{aligned}
\max_{i \neq 0} |1 - \eta_i| &= \|M\| \\
&\leq \|E\| + \|C\| + \|R\| + \|S\| \\
&\leq 1 - \lambda + O(\beta).
\end{aligned}$$

In the other direction, the lower bound follows as $\|M\| \geq \|C\| - \|E\| - \|R\| - \|S\| = 1 - \lambda + O(\beta)$.

This gives the following bound on the spectral gap of H , completing the proof of Theorem 1:

$$\lambda_H = 1 - \max_{i \neq 0} |1 - \eta_i| = \lambda - O(\beta).$$

□

It remains to prove Lemmas 3 and 4.

Proof of Lemma 3. For a vertex $v_i \in G$, we can write $d'_i = \sum_{v_j \sim v_i} X_j$ where X_j is the random indicator variable having value 1 if $\{v_i, v_j\} \in E(H)$ and 0 otherwise. Then $E[d'_i] = pd_i$ and $\text{Var}(d'_i) = d_i p(1-p)$. By the Chernoff bounds, we have

$$\Pr(|d'_i - pd_i| > a) \leq \exp\left(-\frac{a^2}{2(d_i p(1-p)) + a/3}\right).$$

Setting $a = 2\sqrt{\log(n)pd_i}$, we have that

$$\begin{aligned} \Pr(|d'_i - pd_i| > a) &\leq \exp\left(-\frac{4pd_i \log(n)}{2pd_i(1-p) + 4\sqrt{pd_i \log(n)}/3}\right) \\ &\leq n^{-2+o(1)} \end{aligned}$$

Thus almost surely, for all i we have $|d'_i - pd_i| \leq 2\sqrt{\log(n)pd_i}$. This can be restated as, for all i ,

$$|d'_i - pd_i| \leq pd_i \sqrt{\frac{\log n}{pd_i}} = pd_i O\left(\sqrt{\frac{\log n}{pd_{\min}}}\right).$$

□

The following proof of Lemma 4 is analogous to Lemma 3.3 in [2].

Proof of Lemma 4. For a vertex $v_i \in G$, let $X_i = (d'_i - pd_i)^2$ and $X = \sum_{i=1}^n \frac{X_i}{pd_i}$. For each i , we can write

$$X_i = \left(\sum_{v_j \sim v_i} X_{ij} - p\right)^2$$

where X_{ij} 's are the indicator random variables of the event that v_i adjacent to v_j in H (as denoted by $v_i \sim v_j$). We define

$$x_{ij} = \begin{cases} X_{ij} - p & \text{if } v_i \sim v_j, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $E[x_{ij}] = 0$, and $X_i = (\sum_j x_{ij})^2$. Also,

$$\begin{aligned} E[X_i] &= \text{Var}(d'_i) = E\left[\sum_{v_i \sim v_j} x_{ij}^2\right] < pd_i \\ E[X_i^2] &= E[(d'_i - pd_i)^4] \\ &= E\left[\left(\sum_{v_i \sim v_j} x_{ij}\right)^4\right] \\ &= \sum_{i \sim j} E[x_{ij}^4] + 6 \sum_{\substack{j \neq k \\ v_j \sim v_i, v_k \sim v_i}} E[x_{ij}^2 x_{ik}^2] \\ &\leq pd_i + 6p^2 d_i^2. \end{aligned}$$

If $v_i \not\sim v_j$ and $v_i \neq v_j$, then X_i and X_j are independent. If $v_i \sim v_j$ and $v_i \neq v_j$, we have

$$\begin{aligned}
\mathbb{E}[X_i X_j] &= \mathbb{E}[(d'_i - pd_i)^2 (d'_j - pd_j)^2] \\
&= \mathbb{E}\left[\left(\sum_{v_k \sim v_i} x_{ik}\right)^2 \left(\sum_{v_l \sim v_j} x_{lj}\right)^2\right] \\
&= \sum_{\substack{v_k \sim v_i, v_l \sim v_j \\ \{i,k\} \neq \{l,j\}}} \mathbb{E}[v_{ik}^2 v_{lj}^2] + \mathbb{E}[v_{ij}^4] \\
&\leq p^2 d_i d_j + p.
\end{aligned}$$

Thus,

$$\begin{aligned}
\text{Var}(X_i) &\leq pd_i + 5p^2 d_i^2 \\
\text{coVar}(X_i, X_j) &\leq \begin{cases} 0 & \text{if } v_i \not\sim v_j, \\ p & \text{otherwise.} \end{cases}
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^n \frac{1}{pd_i} \mathbb{E}[X_i] < n \\
\text{Var}(X) &= \sum_{i=1}^n \frac{1}{p^2 d_i^2} \text{Var}(X_i) + 2 \sum_{i < j \leq n} \frac{1}{w_i w_j} \text{coVar}(X_i, X_j) \\
&\leq \left(5 + \frac{1}{pd_{\min}} + \frac{1}{d}\right)n \\
&\leq 6n.
\end{aligned}$$

Using Chebyshev's inequality, we have, for any $a > 0$,

$$\Pr(|X - \mathbb{E}[X]| > a) \leq \frac{\text{Var}(X)}{a^2}.$$

Setting $a = \sqrt{n}g(n)$, with $g(n) \gg 1$, then almost surely we have $X \leq (1 + o(1))n$. From the definition, $\|\chi\|^2 = X$. Thus, almost surely

$$\|\chi\|^2 \leq (1 + o(1))n$$

as desired. \square

5 Concluding remarks

In this paper, we examine the spectral relationship between a host graph G and its random subgraph with edge-selection probability p . If G has n vertices with a spectral gap λ and minimum degree d_{\min} , then we prove that a random subgraph of G on n vertices with edge-selection probability p almost surely has a spectral gap of $\lambda - O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)$. The special case of having the host graph as the complete graph on n vertices and a random subgraph H chosen with edge-selection probability p is the Erdős-Rényi graph $G(n, p)$. Since the complete graph K_n has eigenvalue $\lambda_1 = n/(n-1)$, our bound for λ_H is $|1 - \lambda_H| = O(\sqrt{(\log n)/n})$ which is off by a factor of $\sqrt{\log n}$ of the best known spectral bound for $G(n, p)$. Therefore there is room for improvements (e.g., by a factor of $\sqrt{\log n}$) concerning the statements of the main theorem here.

References

- [1] F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.
- [2] F. Chung, L. Lu, and V. Vu, The spectra of random graphs with given expected degrees, *Internet Mathematics* **1** (2004) 257–275.
- [3] F. Chung, and L. Lu, *Complex Graphs and Networks*, AMS Publications, 2006.
- [4] J. Friedman, A proof of Alon’s second eigenvalue conjecture, *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, (2003), 720–724.
- [5] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices, *Combinatorica* **1** (1981), no. 3, 233–241.
- [6] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, Reading, MA, 1994.
- [7] M. Krivelevich and B. Sudakov, Pseudo-random graphs, in: *More Sets, Graphs and Numbers*, Bolyai Society Mathematical Studies 15, Springer, (2006), 199–262.
- [8] V. Vu, Spectral norm of random matrices, *Combinatorica* to appear.
- [9] E. P. Wigner, On the disribution of the roots of certain symmetric matrices, *The Annals of Mathematics*, **67** (1958), 325–327.