Pontifícia Universidade Católica
do Rio de Janeiro

Gabriel Dias do Couto

Two Approaches to Moderate Deviations in Triangle Count in $G(n, m)$ Graphs

Dissertação de Mestrado

Thesis presented to the Programa de Pós–graduação em Matemática, do Departamento de Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática.

Advisor: Prof. Simon Griffiths

Rio de Janeiro
Abril 2022

Pontifícia Universidade Católica
do Rio de Janeiro

Gabriel Dias do Couto

Two Approaches to Moderate Deviations in Triangle Count in $G(n,m)$ Graphs

Thesis presented to the Programa de Pós–graduação em Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática. Approved by the Examination Committee:

Prof. Simon Griffiths
Advisor
Departamento de Matemática – PUC-Rio

Doc. José Diego Alvarado Morales
Departamento de Matemática - USP

Profa. Maria Eulalia Vares
Departamento de Matemática - UFRJ

Prof. Robert David Morris
Departamento de Matemática - IMPA

Rio de Janeiro, Abril 19th, 2022

Gabriel Dias do Couto

Graduado em Licenciatura em Matemática pela Universidade de Brasília.

Bibliographic data

Dias do Couto, Gabriel

Two Approaches to Moderate Deviations in Triangle Count in $G(n, m)$ Graphs / Gabriel Dias do Couto; advisor: Simon Griffiths. – 2022.

75 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, 2022.

Inclui bibliografia

1. Combinatória – Teses. 2. Grafos Aleatórios Erdos-Renyi. 3. Desvios Moderados. 4. Martingais. 5. Teoria Espectral de Grafos. 6. Triângulos. I. Griffiths, Simon. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. III. Título.

CDD: 004

To everyone that is dear to me,
specially those who are not
mathematicians

Acknowledgments

Eu gostaria de agradecer, primeiramente, os meus pais, Cândida e Ladislau, por todo o carinho e suporte que sempre me deram e continuam dando. Sempre ei de valorizar o trabalho desses dois que me são muito queridos. Também aos meus irmãos Marco Túlio, Rafael e Rodolfo, por sempre serem companheiros.

Aos meus amigos de Goiânia, Arley, Bernardo, Carrijo, Gustavo, João Victor e Percival, por estarem comigo desde muito cedo e serem tão cordiais.

To PETMAT-UnB, specially Luciana, for showing me what it is to be a researcher, and much more, and giving me the will to be one.

To the rest of my family that that gave me a very healthy and lovely ambient while I growed up.

To my friends from the rest of Brazil: from DF, MG, TO, SP, RJ and more. You all helped me to discover new worlds.

To all professors that I already had. You all influenced me do the choices that turned me into who I am today.

To the institutions that helped me with my superior studies: UnB and PUC-Rio. Both were very kind and respectful with me and I hope to repay the generosity one day.

To Faperj and CAPES for supporting this master studies.

Finally, to my advisors that I dear the most: Daniele Nantes and Simon Griffiths. Both had a very joyful way of advising me and I'm really glad that I had the opportunity to study with you.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Dias do Couto, Gabriel; Griffiths, Simon (Advisor). **Two Approaches to Moderate Deviations in Triangle Count in $G(n, m)$ Graphs**. Rio de Janeiro, 2022. 75p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

The study of deviations, and in particular large deviations, has a long history in Probability Theory. In recent decades many articles have considered these questions in the context of subgraphs of the random graphs $G(n, p)$ and $G(n, m)$. This dissertation considers the lower tail for the number of triangles in the random graph $G(n, m)$. Two approaches are considered: Martingales, based on the article of Christina Goldschmidt, Simon Griffiths and Alex Scott; and Spectral Graph Theory, based on the article of Joe Neeman, Charles Radin and Lorenzo Sadun. These two approaches manage to find the behavior of the tail in two different regimes. In this dissertation we give an overview of the article of Goldschmidt, Griffiths and Scott, discuss in detail the article of artigo Neeman, Radin and Sadun. In particular, we shall explore the connection between the lower tail of the number of triangles and the behavior of the most negative eigenvalues of the adjacency matrix. We shall see that the triangle count tends to especially depend on the most negative eigenvalue.

Keywords

Erdos-Renyi Random Graphs; Moderate Deviations; Martingales; Spectral Graph Theory; Triangles.

Resumo

Dias do Couto, Gabriel; Griffiths, Simon. **Duas Abordagens em Desvios Moderados para Contagem de Triângulos em Grafos** $G(n,m)$. Rio de Janeiro, 2022. 75p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

O estudo de desvios, e em particular grandes desvios, tem uma história longa na teoria de probabilidade. Nas últimas décadas muitos artigos consideraram essas questões no contexto de subgrafos de grafos aleatórios $G(n,p)$ e $G(n,m)$. Esta dissertação considera a cauda inferior para o número de triângulos no grafo aleatório $G(n,m)$. Duas abordagens estão consideradas: Martingales, a partir artigo de Christina Goldschmidt, Simon Griffiths e Alex Scott; e Teoria Espectral de Grafos, a partir do artigo de Joe Neeman, Charles Radin e Lorenzo Sadun. Essas duas abordagens conseguem encontrar o comportamento da cauda em dois regimes diferentes. Na dissertação discutiremos a visão geral do artigo de Goldschmidt, Griffiths e Scott, e discutiremos em detalhes o artigo de Neeman, Radin e Sadun. Em particular, exploraremos a conexão entre a cauda inferior do número de triângulos e o comportamento dos autovalores mais negativos da matriz de adjacência. Veremos que a contagem tende a depender, essencialmente, do autovalor mais negativo.

Palavras-chave

Grafos Aleatórios Erdos-Renyi; Desvios Moderados; Martingais; Teoria Espectral de Grafos; Triângulos.

Table of contents

Beauty is the first test: there is no permanent place in this world for ugly mathematics.

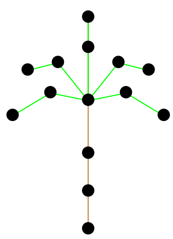
G. H. Hardy, *A Mathematician's Apology.*

1
Introduction

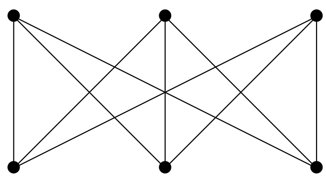
Combinatorics is the most beautiful branch of mathematics that I'm aware of. An outsider of the math world would say that it is the area of mathematics which considers combinations of things. This crude view contains a grain of truth.

This area has a lot of different definitions. For me, Combinatorics is the study of finite objects and their combinations. Not limited to computing the possible combination of them, this branch also studies if we can find or construct some structures with these objects. The array of questions that we can ask in Combinatorics is tremendous. The most intriguing aspect of this branch for me is its complex simplicity. There is a great number of questions that can be understood by most people, but some of them are insanely difficult to answer.

Many of the most intriguing problems in Combinatorics come from Graph Theory. We consider only finite simple graphs. In other words, for us, a **Graph** is a finite set of vertices and a finite set of edges connecting those vertices two-by-two. You can see two examples below.



A beautiful tree



A Bipartite Graph

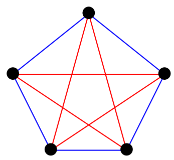
As you can see, there are a lot of simple questions that we can make. For example, I gave names to the graphs. When can I call a graph by the name of **"tree"** or **"bipartite"**? These questions are already answered: a **tree**, by definition, doesn't contain cycles and a **bipartite** graph doesn't contain odd cycles.

The idea of a bipartite graph is that it can be partitioned into two sets of vertices A and B where each edges connects vertices from different sets. What about "tripartite" or "quadripartite" graphs? Is there a equivalent condition for them as bipartite graphs have? As I said, there are simple questions in Combinatorics that are very hard to answer. Today, only the bipartition has a known criteria.

The last two paragraph were to show the relation of Graph Theory with my definition of Combinatorics. Let's continue with another simple question:

Given a complete graph, i.e. a graph where all the vertices are connected with the others, colour the edges with red and blue. What is the smallest number of vertices , n, such that every colouring of the complete graph with n vertices contains a monocromatic triangle? The answer is 6!

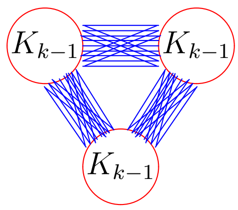
The proof is very simple: since the graph has 6 vertices, each vertex sends at last 3 edges with the same colour, say blue. Either the receiving vertices forms a red triangle or two of them forms a blue triangle with the first vertex. For 5 vertices we have the following counter example



A 2-coloring of a complete graph of five vertices

We, then, can go further: when is a 2-coloured complete graph guarenteed to contain a monocromatic complete graph on 4 vertices? We know the answer to be 18. However, if we ask it for 5 vertices, we are already doomed. Or at least, beyond the knowledge of all the mathematicians of the world. For that range we only know lower and upper bounds.

Let's construct a lower bound. If we are looking for a K_k, a complete graph of k vertices, we can partitionate the graph into $k-1$ sets of $k-1$ vertices and get something like



Case k=3

which gives us a lower bound of $(k-1)^2$, since this constructed structure doesn't contain a monochromatic K_k. This is far from the truth and the best constructive lower bound that I'm aware of, which is due to Frankl and Wilson [7] and is something close to $k^{\sqrt{\ln k}}$.

The best lower bound, up to a constant factor, of this problem come from Erdős [6], from 1947, where he firstly applied a (counting) probabilistic approach to it. The resulting bound is $\sqrt{2}^k$ and is yet to be beaten significantly. In order to prove it, he showed that the probability there exists a colouring with no monochromatic K_k is bigger than zero, and so there must exist such a colouring. This proof opens a brand new idea: we don't need to explicitly *show* a structure to answer some questions, we only need to show that such a structure exists!

After that insight, if we can use probability in order to show that some graph exists, why couldn't we define some graphs probabilistically? And so the random graphs were born. The two most famous random graphs are the $G(n, p)$ and $G(n, m)$. The first one, $G(n, p)$, is defined by tossing a coin for each edge with probability p of that edge existing independently of all other edges. The second random graph, $G(n, m)$, is chosen uniformly in the set of all graphs of n vertices and m edges. Both graphs are known as Erdős-Rényi random graphs, but the $G(n, p)$ is due to Gilbert [10] and the second one to Erdős and Rényi [5].

A very common question to ask for random graphs is when does there exists a certain structure inside it. Of course one can't guarantee that a $G(n, p)$ contains a triangle, since it may happens that the graph has no edges whatsoever. But, one may show that if we let n tend to infinite, the probability of $G(n, p)$ having a triangle goes to 1 if $p \gg \frac{1}{n}$ or goes to 0 if $p \ll \frac{1}{n}$.

The $G(n, m)$ random graph has a much more rigid structure. The set of possible outcomes of $G(n, p)$ is the set of all graphs of n vertices (given that $p \in (0, 1)$). Some events in $G(n, p)$ become more or less likely depending on the number of edges. On the other hand, $G(n, m)$ has only the set of graphs with m edges and n vertices as possible outcomes. Furthermore, we cannot remove or add any edges from it because of m being fixed. One may think that setting $m = p\binom{n}{2}$ would make $G(n, m)$ and $G(n, p)$ be alike. It is true in some sense, but not entirely true (Bollobás [3] showed that some properties hold for both random graphs with this setting).

Looking at triangles, for p constant, and $m = p\binom{n}{2}$ we have the same order for the expected number of triangles in $G(n, m)$ and $G(n, p)$, n^3, but the standard deviation for $G(n, p)$ is of order n^2 and for $G(n, m)$ it is of order $n^{3/2}$. This points out that the easiest way for $G(n, p)$ to have more triangles than expected is to have more edges than expected. As this is an "uninteresting" way to have extra triangles, it is actually more natural to study this problem in $G(n, m)$

Now we are finally reaching the scope of this dissertation: moderate deviations to triangle count in $G(n, m)$. Deviation is a part of probability that asks what is the probability of some certain not average event to happen. There are three kinds of deviations: large deviations (order of the mean), small deviations (order of the standard deviation) and moderate deviations (something in the middle).

Deviations for $G(n, p)$ have been extensively studied. For large deviations, it is known that if the positive deviation is too big or too small, then the result behaves like a $G(n,r)$, for a corresponding r, in the sense that the edges are *uniformly distributed*. But if the deviation is between these two thresholds, then a certain structure is typically the cause of the deviation. As the complement graph has distribution $G(n,q)$, $q = 1-p$, the same result applies for the lower tail. On the other hand, the uniform part of this result do not work for $G(n,m)$ because of its rigid nature. However, the deficits and surplus of deviations are often caused by some specific structure. Indeed, this concept will be used in chapter 3. The reader can see [4] for a nice survey about large deviation in $G(n,p)$.

For small deviations, we can quote Ruciński [19] who stablished that for all p such that $np^{e(H)}$ and $(1-p)n^2$ tend to infinity, the number of copies of H in $G(n,p)$ is asymptotically normally distributed. This result is a "central limit theorem" for the H subgraph count in $G(n,p)$. Janson [15] proved the corresponding result in $G(n,m)$.

Focusing on $G(n,m)$, we will quote two papers for moderate deviations: Goldschmidt, Griffiths, and Scott [11] and Neeman, Sadin, and Radun [18]. These two works are going to be the core of this dissertation. They show that deviations behave differently depending on how large the deviation is. Essentially, Goldschmidt, Griffiths, and Scott [11] show that the central limit theorem behaviour extends to some range of moderate deviations, i.e. for $n^{-3/2} \ll t \ll n^{-1}$ and letting $\tau(G)$ denote the triangle density of $G(n,m)$ with $m = p\binom{n}{2}$ and p a constant

$$\mathbb{P}\left(\Big|\tau(G) - \mathbb{E}[\tau(G)]\Big| \geq t\right) \;=\; \exp\Big(-\Theta(n^3t^2)\Big).$$

This result has some extensions: they gave the value of the constant in the leading-order term of the exponent; and we should point out as well that their work shows that you can let p be any sequence in $(0,1)$ bounded away from 1, however the range of deviations considered, i.e., the range of t, will depend on your choice of p.

On the other hand, Neeman, Sadin, and Radun [18] focused exclusively on the lower tail. That is, they considered the question of how likely it is that a random graph has many *fewer* triangles than expected. They found a regime in which the behaviour is similar to that of large deviations, i.e. for $t \in n^{-3/4} \ll t \ll 1$ and p a constant

$$\mathbb{P}\left(\tau(G) \leq \mathbb{E}[\tau(G)] - t\right) \;=\; \exp\Big(-\Theta(n^2t^{2/3})\Big). \tag{1-1}$$

They also obtained the constant for the leading-order term in the exponent when $p \in [1/2, 1)$, see Theorem 1.1 However, they did not consider the sparse case, in which p tends to 0, which remains open.

We now state their main result, which we shall prove in detail in this dissertation.

Theorem 1.1 *Let $G(n,m)$ be a random graph, with $m = p\binom{n}{2}$, and A its adjacency matrix. Let $\tau(G)$ be the triangle density of $G(n,m)$ and let (λ_i) be the decreasing sequence of eigenvalues of A. If $\frac{1}{2} \leq p < 1$ and $n^{-3/4} \ll t \ll 1$ then*

$$\mathbb{P}(\tau(G) \leq p^3 - t) = \exp\left(-\frac{\ln \frac{p}{1-p}}{2(2p-1)} t^{2/3} n^2 + o(t^{2/3}n^2)\right),$$

letting $\ln \frac{1-p}{p} = 2$ if $p = 1/2$. Moreover, conditioning on $\tau(G) \leq p^3 - t$, whp we have

$$\lambda_n^3 = -tn^3(1-o(1)) \text{ and } \lambda_{n-1}^3 \geq -o(tn^3).$$

Their approach is based on expressing the triangle count in terms of the eigenvalues of the adjacency matrix of the graph. For this reason, a key result on the path towards proving Theorem 1.1 is a deviation result for the eigenvalues and singular values of random matrices, see Theorem 3.17.

We shall also give an overview of the approach of Goldschmidt et al. [11], which is based on a martingale representation of the triangle count deviation. We will not prove the strongest version of their theorem, see [11] for details. We shall prove the following theorem.

Theorem 1.2 *Let $D_\triangle(G_m)$ be the deviation of the number of triangles in $G(n,m)$. There is a non-negative constant $c = c(H)$ such that for all $m \leq \frac{N}{2}$, and all $\alpha, n \geq c^{-1}$, we have*

$$\mathbb{P}\Big(|D_\triangle(G_m)| > \alpha n^{3/2}\Big) \leq \exp\Big(-c\alpha \min\{\alpha, n^{1/2}\}\Big).$$

The layout of the dissertation is the following: Chapter 2 gives the basic definitions and results of Linear Algebra, Spectral Graph Theory and Large Deviation Principle. All of that will support the reader into understanding Chapter 3, where we study thoroughly the paper from Neeman et al. [18]. In particular, in this Chapter we prove a Large Deviation Principal for a constant number of eigenvalues of $G(n,m)$'s adjacency matrix.

In Chapter 4 we give basic definitions and results about matingales. In the same spirit of Chapter 2, Chapter 4 helps the reader into understanding Chapter 5, where we give an overview of the paper from Goldschmidt et al. [11].

Finally, in Chapter 6 we give some final remarks about the problems considered in the dissertation.

2
Basics for the Spectral Approach

For our first approach, we are going to use Spectral Graph Theory coupled with the Large Deviation Principle to control the triangle-count with only a constant number of eigenvalues of the centered adjacency matrix of $G(n, m)$. We will now start to explain each of those terms. Many of these results are considered standard in their respective areas, and we shall give references to their proofs.

2.1
Basics of Spectral Graph Theory

Recalling the basics of Spectral Theory, a vector ϕ is called a eigenvector of M with eigenvalue λ if $M\phi = \lambda\phi$. For our application, we will be interested in symmetric matrices. In this case all eigenvalues are real and there exists an orthonormal basis of eigenvectors.

Theorem 2.1 (The Spectral Theorem [9]) *Let M be a n-by-n real symmetric matrix. There exist real numbers $\lambda_1, \ldots, \lambda_n$, not necessarily distinct, and n mutually orthogonal unit vectors $\phi_1, \ldots, \phi_n$ such that ϕ_i is a eigenvector of M with eigenvalue λ_i for each i.*

A useful tool of combinatorial significance for characterizing eigenvectors of a symmetric matrix is called the Rayleigh quotient.

Definition 2.2 *The **Rayleigh quotient** of a vector v with respect to a matrix M is the quotient*

$$\frac{v^T M v}{v^T v}.$$

Note that if ϕ is a eigenvector of M, then its Rayleigh quotient with respect to M is its eigenvalue λ.

Theorem 2.3 ([9]) *Let M be a real symmetric matrix and v a non-zero vector that maximizes the Rayleigh quotient with respect to M. Then v is an eigenvector of M with eigenvalue equals to its Rayleigh quotient. Moreover, its eigenvalue is the largest one.*

For completeness we will define what are the singular values of a matrix and state the Singular Decomposition of any matrix.

Definition 2.4 *Given a matrix A, its **Singular Values** σ_m are the (non-negative) square roots of the eigenvalues of $A^T A$.*

Note that the Singular Values are always non-negative.

Theorem 2.5 (Singular Value Decomposition [9]) *For any matrix m-by-n with real entries, we have*

$$A = U\Sigma V^T$$

where

- *U is a m-by-m orthogonal matrix whose columns are the eigenvectors $\mathbf{u}_i$ of AA^T;*
- *V is a n-by-n orthogonal matrix whose columns are the eigenvectors $\mathbf{v}_i$ of A^TA and;*
- *Σ is a n-by-m matrix with all but the first $k = rank(A)$ diagonal entries equal to zero. The diagonal entries σ_i are the Singular Values of A and work such that $T(\mathbf{v}_i) = \sigma_i\mathbf{u}$.*

A few extras about Singular Values:

- If we write A as a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, then $T(\mathbf{v}_i) = \sigma_i(A)\mathbf{u}_i$ for $1 \leq i \leq \min(m, n)$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are as in the theorem, and if $i > \min(n, m)$, then $T(\mathbf{v}_i) = 0$
- When A is a symmetric matrix, then $\sigma_i(A) = \sqrt{\lambda_i(A^TA)} = \sqrt{\lambda_i(A^2)} = |\lambda_i(A)|$ (here not the eigenvalues aren't necessarily in decreasing order).

Spectral Graph Theory consists of studying graphs via their associated matrices. This allows us to bring tools of linear algebra into the study of graphs. The most commonly known matrix that is linked to a graph G is the Adjacency Matrix, which is symmetric. As usual, let $V(G) = [n]$.

Definition 2.6 *The **adjacency matrix**, A, of G is a n-by-n matrix whose entries a_{ij} are defined as*

$$a_{ij} = \begin{cases} 1 & \textit{if } (i,j) \in E(G) \\ 0 & \textit{otherwise.} \end{cases}$$

Obs: For the rest of this session, we will denote the eigenvalues of the adjacency matrix by $\lambda_1, \ldots, \lambda_n$ and $\phi_1, \ldots, \phi_n$ as their respective eigenvectors.

In spite of being the simplest matrix to associate to a graph, this is going to be the only matrix we are going to need in order to progress. Now let's try to get some intuition on how we are going to handle triangles using matrices.

Firstly, how can we count the number of triangles in a graph using only its adjacency matrix? We have the following result for counting the number of walks of any length from a vertex to another:

Theorem 2.7 *The number of walks of length ℓ from i to j is the entry ij of the matrix A^ℓ.*

Proof. We are going to prove by induction:

- Base case: $\ell = 0$ gives $A^0 = I$ which works, but (if you are not happy) it also works for $\ell = 1$ by the definition of A.
- Induction Hypothesis: Suppose that the theorem is true for $\ell = L$.
- Induction Step: The set of walks of length $L+1$ from i to j has the same cardinality as the set of walks of length L from i to a neighbour of j. So, we can use the induction hypothesis and get

$$\sum_{(h,j)\in E(G)} (A^L)_{ih} = \sum_{h=1}^{n} (A^L)_{ih} a_{hj} = (A^{L+1})_{ij}.$$

■

So if we take $i = j$ and $\ell = 3$, we have that the trace of A^3, $tr(A^3)$, is six times the number of triangles from G. Now let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of A. We know that they all exist as real numbers because A is symmetric. Moreover, we know that $\sum_{i=1}^{n} \lambda_i^3$ gives six times the number of triangles from G. Can we have a little bit of more control over the triangles looking at fewer eigenvalues? This question is going to be answered in a moment far from now, but let's construct some possible intuition on why it might be true.

Theorem 2.8 (Perron-Frobenius [20]) *Let G be a connected graph. Then*

- $\lambda_1 \geq -\lambda_n$*;*
- *The eigenvalue λ_1 has a strictly positive eigenvector.*

Using this result, one may give a characterization for the bipartite graphs using only Spectral Graph Theory.

Theorem 2.9 *If G is a connected graph, then $\lambda_1 = -\lambda_n$ if, and only if, G is bipartite.*

Proof. For the proof, note that for any vector $\psi \in \mathbb{R}^n$ and $u \in V(G)$,

$$(A\psi)(u) = \sum_{(u,v)\in E(G)} \psi(v).$$

($\Rightarrow$) Let $\lambda_1 = -\lambda_n$. Without the hypothesis, we have the following inequality

$$|\lambda_n| = \left|\phi_n^T A \phi_n\right| = \left|\sum_{u,v} A(u,v)\phi_n(u)\phi_n(v)\right| \leq \sum_{u,v} A(u,v)|\phi_n(u)||\phi_n(v)| \leq \lambda_1$$

where the last inequality is justified by λ_1 being the maximum Rayleigh quotient.

Since we have equality and A is the adjacency matrix, it means that for every pair (u, v) that is an edge of G, we have $\mathrm{sgn}\big(\phi_n(u)\big) = -\mathrm{sgn}\big(\phi_n(v)\big)$. Therefore, we have a bipartition defined by the sign of u in the eigenvector ϕ_n.

($\Leftarrow$) Let G be a bipartite graph with $G = C \cup B$ its bipartition. Define

$$x(u) = \begin{cases} \phi_1(u), & u \in C \\ -\phi_1(u), & u \in B. \end{cases}$$

Then for $u \in C$

$$(Ax)(u) = \sum_{(u,v)\in E(G)} x(v) = -\sum_{(u,v)\in E(G)} \phi_1(v) = -\lambda_1\phi_1(u) = -\lambda_1 x(u),$$

and for $u \in D$

$$(Ax)(u) = \sum_{(u,v)\in E(G)} x(v) = \sum_{(u,v)\in E(G)} \phi_1(v) = \lambda_1\phi_1(u) = -\lambda_1 x(u).$$

Therefore, x is an eigenvector of eigenvalue $-\lambda_1$ which is greater than λ_n, as λ_n is the smallest eigenvalue, and since $-\lambda_1 \leq \lambda_n$, by Perron-Frobenius, we have the equality as λ_n is the smallest eigenvalue of A.

■

Remember that a bipartite graph is a graph without odd cycles and triangles are odd cycles. So we could conjecture that the smallest eigenvalue, or a constant amount of the smallest eigenvalues, is/are responsible for destroying or keeping the triangles existence.

While I would like to continue this discussion of Spectral Graph Theory, this small presentation is enough for what follows. If the reader is interested to learn more about Spectral Graph Theory, including other quantitative results about qualitative properties of graphs, see [2, 20].

2.2 Basics of the Large-Deviation Principle

Large deviation theory is an intriguing branch of Probability Theory with a long history and many applications. However, it is also quite technical and overloaded with notations. Because of that, we are going to introduce its results as we develop our approach. But, at least, let's state what it is and consider one application.

Definition 2.10 *Let $\xi_1, \xi_2, \dots$ be a sequence of random variables in some probability space P with Borel σ-field $\mathcal{B}$. We say that the sequence (ξ_n) satisfies the **large-deviation principle** (LDP) with **rate function** $I : P \to [0, \infty]$, if for any $B \in \mathcal{B}$ we have*

$$\begin{aligned} -\inf_{x \in B^\circ} I(x) &\leq \liminf_{n\to\infty} n^{-1} \ln \mathbb{P}(\xi_n \in B) \\ &\leq \limsup_{n\to\infty} n^{-1} \ln \mathbb{P}(\xi_n \in B) \leq -\inf_{x\in\overline{B}} I(x) \end{aligned} \tag{2-1}$$

Roughly speaking, the LDP gives the exact logarithmic rate for which the probability of a certain deviation goes to zero. As a first example, let us consider the binomial distribution, which may be expresses as a sum of Bernoulli trials.

Let (X_i) be a sequence of i.i.d. Bernoulli random variables with probability $p \in (0,1)$ and denote $S_n = X_1 + \dots + X_n$. Let $x \in (0,1)$ and we want to compute

$$\mathbb{P}\Big(S_n = \lfloor xn \rfloor\Big) = \binom{n}{\lfloor xn \rfloor} p^{\lfloor xn \rfloor}(1-p)^{n-\lfloor xn \rfloor}.$$

For simplicity, let $l = \lfloor xn \rfloor$. Using Stirling's approximation formula, we get

$$\begin{aligned} \mathbb{P}\Big(S_n = l\Big) &= \frac{n!}{l!(n-l)!} p^l (1-p)^{n-l} \\ &= \sqrt{\frac{n}{2\pi l(n-l)}} n^n l^{-l} (n-l)^{-n+l} p^l (1-p)^{n-l} \\ &= \sqrt{\frac{n}{2\pi l(n-l)}} e^{n\ln n - l\ln l - (n-l)\ln(n-l) + l\ln p + (n-l)\ln(1-p)} \\ &= \sqrt{\frac{n}{2\pi l(n-l)}} e^{n[\frac{l}{n}\ln p + (1-\frac{l}{n})\ln(1-p) - \frac{l}{n}\ln\frac{l}{n} - (1-\frac{l}{n})\ln(1-\frac{l}{n})]} \\ &= \frac{1}{\sqrt{2\pi n x(1-x)}} e^{-nI(x)+O(\ln n)} = e^{-nI(x)+O(\ln n)} \end{aligned} \tag{2-2}$$

where $I(x) = x\ln\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p}$ if $x \in [0,1]$, and ∞ otherwise.

With some calculations, one can check that $I(x)$ is convex, $I(p) = I'(p) = 0$ and $I''(p) = (p(1-p))^{-1}$. Therefore $I(x)$ takes its minimum only at the point p, this means that the most probable result of S_n is $\lfloor pn \rfloor$, which is consistent with $n^{-1}S_n \to p$. Also, by 2-2,

$$\lim n^{-1} \ln \mathbb{P}(S_n = l) = -I(x)$$

and, therefore, the sequence $(i^{-1}S_i)$ satisfies the LDP with $I(x)$ as its rate function, i.e., if $x \neq p$, then $-I(x)$ gives the exact exponential rate of convergence to zero of the probability of the event $\{S_n = \lfloor xn \rfloor\}$.

We won't actually use the Bernoulli random variable X, but we will use its centralized form $X - \mathbb{E}[X]$. However, the calculations are very similar and we get that its rate function is

$$I(x) = (p+x)\ln\frac{p+x}{p} + (1-p-x)\ln\frac{1-p-x}{1-p}$$

if $x \in [-p, 1-p]$, and ∞ otherwise

Now, we will define some objects that we are going to use a lot throughout this approach of the problem. We begin with a well known definition.

Definition 2.11 *Let X be a convex subset of a real vector space. A function $f : X \to \mathbb{R}$ is called **convex** if for all $t \in [0,1]$ and $x_1, x_2 \in X$*

$$f(tx_1 - (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

Now for a not so famous concept.

Definition 2.12 *Let ξ be a random variable. Then its **cumulant-generating function** is defined as*

$$\Lambda_\xi(s) := \ln \mathbb{E}[e^{s\xi}]$$

whenever it exists. If $\ln \mathbb{E}[e^{s\xi}]$ doesn't exist, put $\Lambda_\xi(s) = \infty$

I find this definition very similar, in conception, to the standard deviation where we take the square root of the expectation of something squared. This function has some properties that we will use soon.

Theorem 2.13 *Suppose that Λ_ξ exists. Then Λ_ξ has the following properties:*

- *is infinitely differentiable;*
- $\Lambda_\xi(0) = 0$*;*
- *is convex;*
- *and* $\Lambda'_\xi(0) = \mathbb{E}[\xi]$

Proof. Note that $\Lambda_\xi(s) = \ln M_\xi(s)$ where $M_\xi(s)$ is the moment generating function. Since both of the functions $\ln x$ and $M_\xi(s)$ exist and are infinitely differentiable, it follows, by standard results of calculus, that $\Lambda_\xi(s)$ is infinitely differentiable.

The second claim is trivial since $M_\xi(0) = 1$.

For the third claim, let $a+b=1$ and $u,v \in Dom(\Lambda_\xi)$. Then

$$\begin{aligned}
\Lambda_\xi(av+bu) &= \ln \mathbb{E}[e^{av+bu}] \\
&= \ln \mathbb{E}[e^{av}e^{bu}] \\
&\leq \ln(\mathbb{E}[e^{v\xi}]^a \mathbb{E}[e^{u\xi}]^b) \\
&= a \ln \mathbb{E}[e^{v\xi}] + b \ln \mathbb{E}[e^{u\xi}] \\
&= a\Lambda_\xi(v) + b\Lambda_\xi(u).
\end{aligned}$$

where, for the inequality, we used Holder's Inequality with functions $f = e^{v\xi}$ and $g = e^{u\xi}$.

Finally, $\Lambda'_\xi(0) = \frac{M'_\xi(0)}{M_\xi(0)} = \mathbb{E}[\xi]$. ■

Note that a random variable X with normal distribution $N(\mu, \sigma^2)$ has $\Lambda_X(s) = \ln\Big(\exp(\mu s + \frac{\sigma^2 s^2}{2})\Big) = \mu s + \frac{\sigma^2 s^2}{2}$. This lead us to the following definition (thinking of $\mu = 0$).

Definition 2.14 *A random variable ξ is called **subgaussian** if there exists a constant C such that $\Lambda_\xi(s) \leq Cs^2$ for all s.*

We are defining this object because our random variables are going to be subgaussian. Let X be a Bernoulli random variable. Then $\mathbb{E}[X] = p$ and $0 \leq X \leq 1$. Hoeffding's Lemma gives us $\mathbb{E}[e^{\alpha X}] \leq \exp(\alpha p + \frac{\alpha^2}{8})$, which implies that $\Lambda_X(s) \leq sp + \frac{s^2}{8}$. Yet, let $\xi = X - \mathbb{E}[X]$ and the same argument gives $\Lambda_\xi(s) \leq \frac{s^2}{8}$ and, therefore, ξ is a subgaussian random variable.

Our last definition is the Legendre Transformation.

Definition 2.15 *Let $I \subset \mathbb{R}$ be an interval and f a convex function. Then the **Legendre transformation** $f^* : I^* \to \mathbb{R}$ of f is defined as*

$$f^*(x^*) = \sup_{x \in I}(x^*x - f(x))$$

for $x^* \in I^* = \{x^* \in \mathbb{R} : \sup_{x \in I}(x^*x - f(x) < \infty\}$.

To simplify our notation, we will write $f^*(x) = \sup_u(ux - f(u))$.

To give some intuition about the Legendre transformation, suppose that f is differentiable. Then $ux - f(u)$ is differentiable and maximizes at $x = f'(u)$, since f is convex. Also suppose that f' is invertible, then $u = (f')^{-1}(x)$ which gives us that $f^*(x) = (f')^{-1}(x)x - f((f')^{-1}(x))$. This function is minus the value of the intersection between the y-axis and the line tangent to f with inclination equals to x. So, if f is strictly convex, then we can reconstruct it from its Legendre transformation.

Some properties are as follows and won't be proved here but will be used sometimes.

Theorem 2.16 *Let f and g be two convex functions. Then:*

1. *f^* is convex;*

2. *$f^{**} = f$;*

3. *If $f \leq g$, then $g^* \leq f^*$;*

4. *If $f(x) = cx^2$, then $f^*(x) = \frac{x^2}{4c}$.*

With these properties, we can prove the following result which will be important later.

Lemma 2.17 *If ξ is a subgaussian random variable, then*

$$4 \sup_{s \in \mathbb{R}} \frac{\Lambda_\xi(s)}{s^2} = \left(\inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(s)}{s^2} \right)^{-1} < \infty.$$

Proof. The first expression is finite since ξ is subgaussian. And so, we just need to prove that the two expressions are equal.

Let $L := \sup_{s \in \mathbb{R}} \frac{\Lambda_\xi(s)}{s^2}$ and $\ell := \inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(s)}{s^2}$, and define $M_L(s) = Ls^2$. Then $\frac{\Lambda_\xi(s)}{s^2} \leq \frac{M_L(s)}{s^2}$ implies $\Lambda_\xi(s) \leq M_L(s)$ for all s. We now apply the Legendre transformation on both sides, which will switch the inequality by property 3, and get $\Lambda_\xi^*(s) \geq M_L^*(s) = \frac{s^2}{4L}$, where we used property 4 for the calculation. Dividing both sides by s^2

$$\frac{\Lambda_\xi^*(s)}{s^2} \geq \frac{L^{-1}}{4} \quad \Rightarrow \quad 4 \sup_{s \in \mathbb{R}} \frac{\Lambda_\xi(s)}{s^2} \geq \left(\inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(s)}{s^2} \right)^{-1}$$

because it holds for all s. Now we need to get the inequality from the other side. We only repeat the argument.

By the definition of ℓ, we have $\Lambda_\xi^*(s) \geq \ell s^2$ for all s. Therefore, applying the Legendre transformation on both sides, using properties 1, 2, 3 and 4, we get

$$\Lambda_\xi^{**}(s) = \Lambda_\xi(s) \leq \frac{s^2}{4\ell}.$$

Dividing both sides by s^2 gives

$$\frac{\Lambda_\xi(s)}{s^2} \leq \frac{\ell^{-1}}{4} \quad \Rightarrow \quad 4 \sup_{s \in \mathbb{R}} \frac{\Lambda_\xi(s)}{s^2} \leq \left(\inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(s)}{s^2} \right)^{-1}$$

which completes the proof. ∎

All of those definitions and results together lead us to Cramér's Theorem.

Theorem 2.18 (Cramér [16]) *Let (ξ_i) be a sequence of i.i.d. random variables with $\mathbb{E}[\xi_i] = m < \infty$. Then, for any $x \geq m$ we have*

$$n^{-1} \ln \mathbb{P}(n^{-1} \sum_{i=1}^{n} \xi_i \geq x) \to -\Lambda_\xi^*(x).$$

With this machinery, I think that we are ready to start our new approach.

2.2.1 Two Extras

The following two definitions are here since we are going to use them, but they won't be motivated. The motivation, in this case, is going to be the results from the next chapter.

The goodness of the rate function is a topological property.

Definition 2.19 *A **good rate function** is a rate function such that for every $r \in (0, \infty)$, we have $I^{-1}([0, r])$ is compact.*

The speed of the LDP comes from changing the normalization by n by anything that goes to infinity.

Definition 2.20 *Let $\xi_1, \xi_2, \ldots$ be a sequence of random variables in some probability space P with Borel σ-field $\mathcal{B}$. We say that the sequence (ξ_n) satisfies the LDP with rate function $I : P \to [0, \infty]$ and **speed** $m_n \to \infty$, if for any $B \in \mathcal{B}$ we have*

$$\begin{aligned} -\inf_{x \in B^\circ} I(x) &\leq \liminf_{n \to \infty} m_n^{-1} \ln \mathbb{P}(\xi_n \in B) \\ &\leq \limsup_{n \to \infty} m_n^{-1} \ln \mathbb{P}(\xi_n \in B) \leq -\inf_{x \in \overline{B}} I(x) \end{aligned}$$

3
The Spectral Approach

For this chapter, we denote the eigenvalues of our matrix as $\lambda_1, \ldots, \lambda_n$ and the singular values by $\sigma_1, \ldots, \sigma_n$ in decreasing order, unless otherwise stated. The vector v_i will denote the eigenvector of λ_i, unless otherwise stated.

Let's give an overview of the following result and observe that this one is only for the lower tail. Also, we are limited to $p \in [1/2, 1)$.

Theorem 3.1 *Let $G(n,m)$ be a random graph and A its adjacency matrix. Let $\tau(G)$ be the random variable corresponding to the triangle density of $G(n,m)$. If $\frac{1}{2} \leq p < 1$ and $n^{-3/4} \ll t \ll 1$ then*

$$\mathbb{P}(\tau(G) \leq p^3 - t) = \exp\left(-\frac{\ln \frac{p}{1-p}}{2(2p-1)} t^{2/3} n^2 + o(t^{2/3} n^2)\right),$$

letting $\ln \frac{1-p}{p} = 2$ if $p = 1/2$. Moreover, conditioning on $\tau(G) \leq p^3 - t$, with high probability we have

$$\lambda_n^3 = -tn^3(1 - o(1)) \text{ and } \lambda_{n-1}^3 \geq -o(tn^3).$$

Firstly, we are going to show that analysing the centralized adjacency matrix, $\tilde{A} = A - \mathbb{E}[A]$, of $G(n,m)$ is sufficient, instead of using the adjacency matrix.

Then, we will show that the largest eigenvalues and singular values of $\tilde{A}$ obey an LDP with some parameters. In other words, we determine the large deviation behavior of the vectors $(\sigma_1, \sigma_2, \ldots, \sigma_k)$ and $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ for both $\tilde{A}$ and $-\tilde{A}$. Note that we include $-\tilde{A}$, as the *largest* eigenvalues of $-\tilde{A}$ correspond to the *smallest* eigenvalues of the original matrix, which are crucial for understanding the triangle count.

Finally, armed with the LDP property and auxiliary results, we show the upper and lower bounds on the probability of the triangle count deviation. In fact, it will turn out that, the triangle count depends especially on the most negative eigenvalue

3.1
The Centralized Adjacency Matrix Suffices

We will settle the first part, showing that the centralized matrix suffices. Let A_n be the adjacency matrix of $G(n,m)$. If we put $p = m/\binom{n}{2}$, then the centralized matrix $\tilde{A}_n = A_n - \mathbb{E}[A_n] = A_n - p\mathbf{1} + pI$, where $\mathbf{1}$ is the n-by-n

matrix with all entries equals to 1. Also, we denote by $tr[A]$ as the trace of the matrix A.

Lemma 3.2 *Let G be a graph with n vertices and d_i the degree of vertex i or each $i \in [n]$. Then for any $p \in [0,1]$, if A is the adjacency matrix of G, then we have*

$$tr[\tilde{A}_n^3] = tr[A_n^3] - p^3n^3 + p^3n + 6mp(np - 2p + 1) + 3p^3n(n-1) - 3p\sum_i d_i^2.$$

Moreover, if $p = m/\binom{n}{2}$, then

$$tr[\tilde{A}_n^3] \leq tr[A_n^3] - p^3n^3 + p^3n + 6mp$$

Proof. The first claim is achieved by brute force and using some identities:

$$
\begin{aligned}
&tr[(A_n - p\mathbf{1} + pI)^3] = \\
&= tr[A_n^3 - 3pA_n^2\mathbf{1} + 3p^2A_n\mathbf{1}^2 - p^3\mathbf{1}^3 + 3p^3\mathbf{1}^2 - 3p^3\mathbf{1} + p^3I + 3p^2A_n + 3pA_n^2 - 6p^2A_n\mathbf{1}] \\
&= tr[A_n^3] - 3p\sum_i d_i^2 + 6nmp^2 - p^3n^3 + 3p^3n^2 - 3p^3n + p^3n + 3p^2tr[A_n] + 6mp - 12mp^2 \\
&= tr[A_n^3] - p^3n^3 + p^3n + 6mp(n - 2p + 1) + 3p^3n(n-1) - 3p\sum_i d_i^2.
\end{aligned}
$$

where we used that $tr[A_n\mathbf{1}] = tr[A_n^2] = 2m$, since $A_n(G)$ is the adjacency matrix of G. The sum term came from

$$tr[A_n^2\mathbf{1}] = \sum_{(i,j)\in[n]^2} (A_n^2)_{ij} = \sum_{(i,j,k)\in[n]^3} (A_n)_{ik}(A_n)_{kj} = \sum_{k=1}^{n} d_k^2.$$

For the inequality, we apply Cauchy-Schwarz inequality to the sum and get

$$\sum_{k=1}^{n} d_k^2 \geq \frac{1}{n}\left(\sum_{k=1}^{n} d_k\right)^2 = \frac{(2m)^2}{n} = 2mp(n-1).$$

Together with that, note that

$$3p^3n(n-1) = 3p^3\frac{2n(n-1)}{2} = 6p^3\binom{n}{2} = 6p^2m$$

and we have our result. ■

With these expressions we can now bound the deviation on triangle count of the adjacency matrix using only the centralized adjacency matrix.

Corollary 3.3 *For any $t \geq 0$,*

$$\mathbb{P}\Big(tr[A_n^3] \leq \mathbb{E}\Big[tr[A_n^3]\Big] - t\Big) \leq \mathbb{P}\Big(tr[\tilde{A_n}^3] \leq O(n^2) - t\Big).$$

The above corollary follows immediately using that $\mathbb{E}[tr[A_n^3]] = n^3p^3 + O(n^2)$, since it is six times the expectation of the number of triangles in $G(n, m)$.

For a lower bound we are going to do completely different tricks only explained in the end.

In order to proceed to the LDP part, we need to translate things to "Frobenius language" first. Remember, we are trying to prove the LDP for the vectors $(\sigma_1, \sigma_2, \ldots, \sigma_k)$ and $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ which will be normalized by a sequence m_n, where $\sqrt{n} \ll m_n \ll n$. More specifically,

Theorem 3.4 (The general subgaussian LDP) *Let A_n be a n-by-n symmetric random matrix having i.i.d. upper diagonal entries being equal in distribution to ξ, a subgaussian random variable, and zero diagonal entries. For every integer $k \geq 1$ and sequence m_n such that $\sqrt{n} \ll m_n \ll n$, we have that the sequence of vectors*

$$X = (\sigma_1, \sigma_2, \ldots, \sigma_k)/m_n$$

satisfies the LDP with speed m_n^2 and good rate function $I : \mathbb{R}_+^k \to [0, \infty)$ with

$$I(x) = \frac{|x|^2}{2} \inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(s)}{s^2}.$$

If, in addition, the s that achieves the above infimum is finite and non-negative, then the sequence

$$Y = (\lambda_1, \lambda_2, \ldots, \lambda_k)/m_n$$

has the same LDP properties.

As a motivation for the proofs to come, we will now present the start of the proof of Theorem 3.4 and see the results we will need to produce.

Remember that, in order to prove a LDP, we need to prove a lower bound for open sets and a upper bound for closed sets. Let's start with the upper bound.

Let $E \subset \mathbb{R}^k$ be a closed set. Since the singular values are always positive real numbers, we can restrict our attention to $\mathbb{R}_+^k$. Let $t = \inf_{z \in E} ||z||$ using the Euclidean norm. Then, the event $X \in E$ is inside the event $||X|| > t$ and therefore

$$\ln \mathbb{P}\Big(X \in E\Big) \leq \ln \mathbb{P}\Big(||X|| > t\Big) = \ln \mathbb{P}\left(\sum_{i=1}^{k} \sigma_i^2 > m_n^2 t^2\right). \qquad (3\text{-}1)$$

Okay, so we are already stuck on the upper bound of the LDP and need to know how to bound the probability of the event $\sum_{i=1}^{k} \sigma_i^2 > m_n^2 t^2$ for any k, t and $\sqrt{n} \ll m_n \ll n$. Let's check the lower bound.

Let $E \subset \mathbb{R}^k$ be a open set and, again, we focus on the positive quadrant. Since E is open, every point of it is an interior point so we have some room for X to be close to $w \in E$ and still be in E. Instead of the typical "we have a ball around w" argument, we will ask for a stronger property that we will be able to bound. Since $w \in E$ there exists a $\varepsilon > 0$ such that for all $z \in \mathbb{R}_+^k$ if $||z||^2 \leq ||w||^2 + \varepsilon$ and $\langle z, w\rangle \geq ||w||^2$, then $z \in E$. Therefore,

$$\ln \mathbb{P}\Big(X \in E\Big) \;\geq\; \ln \mathbb{P}\Big(||X||^2 \leq ||w||^2 + \varepsilon \text{ and } \langle X, w\rangle \geq ||w||^2\Big).$$

We are able to bound both parts of the event separately. The first one is the complement of the same inequality from the closed side, so we can bound its complement by

$$\ln \mathbb{P}\Big(||X||^2 > ||w||^2 + \varepsilon\Big) = \ln \mathbb{P}\left(\sum_{i=1}^{k} \sigma_i^2 > m_n^2(||w||^2 + \varepsilon)\right).$$

The second one will need a different approach as we will need to bound both sides in terms of w

$$\ln \mathbb{P}\Big(\langle X, w\rangle \geq ||w||^2\Big) = \ln \mathbb{P}\left(\sum_{i=1}^{k} \sigma_i w_i \geq m_n^2 \sum_{i=1}^{k} w_i^2\right).$$

Now that we can see what need to be done, let's build up our machinery.

3.2 Simplifying Our Study of Singular Values and Eigenvalues

Let A and B be two matrices of size n-by-m, for any $n, m \in \mathbb{N}$, and with real entries.

Definition 3.5 *The **Frobenius Inner Product** of A and B is defined as*

$$\langle A, B\rangle_F = \sum_{(i,j)\in[n]\times[m]} a_{ij} b_{ij} = tr(A^T B).$$

*The **Frobenius Norm** of A is, then, defined as*

$$||A||_F = \sqrt{\sum_j \sum_i |a_{ij}|^2}.$$

Some properties are as follows:

Theorem 3.6 *Let A and B be n-by-m real matrices.*

1. $||A||_F = \sqrt{\sum_{i=1}^{\min\{n,m\}} \sigma_i^2(A)}$;

2. $\langle A, B\rangle_F \leq \sum_{i=1}^{k} \sigma_i(A)\sigma_i(B)$;

3. *If A and B are positive semidefinite, then* $\langle A, B\rangle_F \geq 0$;

4. *If U is a unitary matrix, then* $||UA||_F = ||AU||_F = ||A||_F$.

As we are dealing with matrices with real entries, this inner product is similar to the dot product on $\mathbb{R}^{nm}$ as just being the sum of the product of each entry.

Now let's let $\mathcal{M}_k$ be the set of n-by-n matrices of Frobenius norm at most 1 and rank at most k. And let $\mathcal{M}_k^+$ be the set of n-by-n matrices of Frobenius norm at most 1, rank at most k, symmetric and positive semidefinite. We are going to translate the sum of the squares of eigenvalues and singular values as being the supremum over the inner product with the matrices of these sets.

Lemma 3.7 *Let A be a n-by-n symmetric matrix. Then,*

$$\sqrt{\sum_{i=1}^{k} \max\{\lambda_i(A), 0\}^2} = \sup_{M \in \mathcal{M}_k^+} \langle A, M\rangle_F.$$

Moreover, let A be any n-by-n matrix. Then,

$$\sqrt{\sum_{i=1}^{k} \sigma_i^2(A)} = \sup_{M \in \mathcal{M}_k} \langle A, M\rangle_F.$$

Proof. The proof is very straight forward getting a double inequality. Since A is symmetric, let $A = Q^T \Lambda Q$ be its eigendecomposition. Put Λ in decreasing order of the eigenvalues and define Λ_k to be equals to $\max\{\lambda_i(A), 0\}$ only on the k first diagonal entries and 0 on the rest. Then choosing M to be

$$M = \frac{Q^T \Lambda_k Q}{||\Lambda_k||_F}$$

we have

$$\begin{aligned}
\langle A, M\rangle_F &= tr(A^T M) \\
&= tr(Q^T \Lambda Q Q^T \Lambda_k Q)||\Lambda_k||_F^{-1} \\
&= tr(\Lambda Q Q^T \Lambda_k Q Q^T)||\Lambda_k||_F^{-1} \\
&= tr(\Lambda \Lambda_k)||\Lambda_k||_F^{-1} \\
&= ||\Lambda_k||_F = \sqrt{\sum_{i=1}^{k} \max\{\lambda_i(A), 0\}^2},
\end{aligned}$$

where we used the cyclic invariance of the trace. Therefore, $\sqrt{\sum_{i=1}^{k} \max\{\lambda_i(A), 0\}^2} \leq \sup_{M \in \mathcal{M}_k^+} \langle A, M \rangle_F$.

For the other side of the inequality, we will just use $A = A_+ - A_-$, where A_+ and A_- are positive semidefinite matrices (the matrices with positive and negative eigenvalues of A). Let $M \in \mathcal{M}_k^+$ be any matrix. Then, using the properties from the Frobenius inner product

$$
\begin{aligned}
\langle A, M \rangle_F &= \langle A_+, M \rangle_F - \langle A_-, M \rangle_F \\
&\leq \langle A_+, M \rangle_F \\
&\leq \sum_{i=1}^{n} \sigma_i(A_+)\sigma_i(M) \\
&= \sum_{i=1}^{k} \max\{\lambda_i(A), 0\}\sigma_i(M) \\
&= \langle A_{k,\max}, \sigma \rangle \\
&\leq ||A_{k,\max}||.||\sigma|| \\
&\leq ||A_{k,max}|| = \sqrt{\sum_{i=1}^{k} \max\{\lambda_i(A), 0\}^2}.
\end{aligned}
$$

where $A_{k,\max}$ is the vector $(\max\{\lambda_1(A), 0\}, \ldots, \max\{\lambda_k(A), 0\})$ and $\sigma = (\sigma_1(M), \ldots, \sigma_k(M))$, both from $\mathbb{R}^k$. Also, for the forth line, we used that M has rank at most k, so only the first k singular values have the possibility of being bigger than zero; and used the fact that A is symmetric to get $\sigma_i(A) = |\lambda_i(A)|$. As M is an arbitrary matrix, we get that $\sqrt{\sum_{i=1}^{k} \max\{\lambda_i(A), 0\}^2} \geq \sup_{M \in \mathcal{M}_k^+} \langle A, M \rangle_F$. and the result holds.

For the "Moreover" statement, about the singular values, we proceed using the same approach, but we use instead the singular decomposition and get the same results, proving the lemma. ■

With this result from Lemma 3.7, it is enough to look at the $\sup_{M \in \mathcal{M}_k} \langle A, M \rangle$ to solve Theorem 3.4, but we will need to, in order to use a union bound, turn this supremum into a maximum. More explicitly, we will need to find a ε-net in $\mathcal{M}_k$.

3.3 Finding Our Net

Let's first define what a ε-net is.

Definition 3.8 *Let (X, d) be a metric space. A subset $\mathcal{N} \subset X$ is called a* ***ε-net*** *if for every $x \in X$ there is a $n \in N$ such that $d(x, n) \leq \varepsilon$, i.e. $\mathcal{N}$ is ε-close to any point of X.*

The first thing that we need, then, is to find a suitable ε-net. As we will later use a union bound argument, we would like to find an ε-net which is not too large.

Observe that if we take any space $\mathcal{S}$ of matrices, say n-by-m matrices, with the Frobenius norm at most 1, then $\mathcal{S}$ is isometric to the Euclidean unit ball of $\mathbb{R}^{nm}$ with the Euclidean norm. Therefore we can use the following famous result for ε-nets of the Euclidean ball.

Lemma 3.9 ([21]) *Let $B \subset \mathbb{R}^d$ be the unit ball and let $||\cdot||$ be a norm. Then it has a ε-net $\mathcal{N}$ with respect to the norm $||\cdot||$ with*

$$|\mathcal{N}| \leq \left(\frac{3}{\varepsilon}\right)^d.$$

As a corollary, we get

Corollary 3.10 *There is a universal constant C such that for all $\varepsilon \in (0, 1)$, there exists a ε-net for $\mathcal{M}_k$ of size $|\mathcal{N}| \leq (C/\varepsilon)^{2kn}$.*

The idea of the proof is simple. We can use Lemma 3.9 to find a ε-net for the unit ball, but it won't give immediately the net we want because we have the nuance of needing its elements to have rank at most k when translated to matrices. The solution is to make a $\varepsilon/3$-net for each element of the Singular Value Decomposition $M = U\Sigma V^T$ so that together they define the net we want. It will be possible to define the net for Σ in such a way that it has to have rank at most k, which resolves the problem we mentioned (using that $rank(AB) \leq \min\{rank(A), rank(B)\}$).

Proof. We will use Lemma 3.9 two times.

First, let D be the set of diagonal matrices with non-negative entries, Frobenius norm at most 1 and rank at most k. Then Lemma 3.9 grants us a $\frac{\varepsilon}{3}$-net of size

$$|\mathcal{D}| \leq \binom{n}{k}\left(\frac{3}{\varepsilon/3}\right)^k = \binom{n}{k}\left(\frac{9}{\varepsilon}\right)^k$$

since only at most k entries are non-zero and we take a net for each combination of k entries from the diagonal.

Now, let O be the set of orthogonal n-by-k matrices. Then O is a subset of the unit ball of $\mathbb{R}^{nk}$. For this part we change the norms. We will use the $1, 2$ norm, i.e.

$$||S||_{1,2} = \max_i ||S_i||$$

where S_i is the column i of S. As Lemma 3.9 works for any norm, we can define the $\frac{\varepsilon}{3}$-net $\mathcal{O}_{1,2}$ of O such that

$$|\mathcal{O}_{1,2}| \leq \left(\frac{9}{\varepsilon}\right)^{nk}.$$

Finally, we define our net as $\mathcal{N} = \{\bar{U}\bar{\Sigma}\bar{V}^T : \bar{U}, \bar{V} \in \mathcal{O}_{1,2}; \bar{\Sigma} \in \mathcal{D}\}$. Therefore,

$$|\mathcal{N}| \leq |\mathcal{O}_{1,2}|^2|\mathcal{D}| \leq \binom{n}{k}\left(\frac{9}{\varepsilon}\right)^k\left(\frac{9}{\varepsilon}\right)^{2nk} \approx \left(\frac{C}{\varepsilon}\right)^{2kn}.$$

It only remains to prove that $\mathcal{N}$ is, indeed, a ε-net. Let $M \in \mathcal{M}_k$ and its singular decomposition be $M = U\Sigma V^T$. Then there exist an $\bar{M} = \bar{U}\bar{\Sigma}\bar{V}^T$, with $\bar{U}, \bar{V} \in \mathcal{O}_{1,2}$ and $\bar{\Sigma} \in \mathcal{D}$, such that $||U - \bar{U}||_{1,2} \leq \varepsilon/3$, $||V - \bar{V}||_{1,2} \leq \varepsilon/3$ and $||\Sigma - \bar{\Sigma}||_F \leq \varepsilon/3$. Using the triangle inequality over the Frobenius norm

$$\begin{aligned}||M - \bar{M}||_F &= ||U\Sigma V^T - \bar{U}\bar{\Sigma}\bar{V}^T||_F \\ &\leq ||(U - \bar{U})\Sigma V^T||_F + ||\bar{U}(\Sigma - \bar{\Sigma})V^T||_F + ||\bar{U}\bar{\Sigma}(V^T - \bar{V}^T)||_F\end{aligned}$$

We need to bound each of the three parts. As U and V are analogous, we will only prove for one of them. For Σ, we just need to note that $||\overline{U}(\Sigma - \overline{\Sigma})V^T||_F = ||\Sigma - \overline{\Sigma}||_F \leq \varepsilon/3$, since U and V are unitary. Now, since V is a unitary matrix, we get that

$$\begin{aligned}||(U - \overline{U})\Sigma V^T||_F &= ||(U - \overline{U})\Sigma||_F \\ &= \sqrt{\sum_{i=1}^{k} \Sigma_{i,i}^2 ||(\mathbf{u}_i - \overline{\mathbf{u}_i})||^2} \\ &\leq \sqrt{||\Sigma||_F||(U - \overline{U})||_{1,2}^2} \\ &\leq \sqrt{1.\frac{\varepsilon^2}{9}} = \frac{\varepsilon}{3}.\end{aligned}$$

and since $\frac{3\varepsilon}{3} = \varepsilon$, the proof is complete. ∎

Okay, now that we know that an ε-net exists, it makes sense to start to transform our inequalities. The first one that we are going to handle is $\mathcal{M}_k$.

Lemma 3.11 *Take an ε-net $\mathcal{N} \subset \mathcal{M}_k$ with $\varepsilon < 1/2$. Then for any n-by-n symmetric matrix A,*

$$\sup_{M \in \mathcal{M}_k} \langle M, A\rangle \leq \frac{1}{1 - 2\varepsilon} \max_{N \in \mathcal{N}} \langle N, A\rangle.$$

The idea of this proof is very similar to Taylor Series. We are going to approximate a arbitrary matrix $M \in \mathcal{M}_k$ by the sum of *decreasing* elements of the net plus some matrices that goes to zero on Frobenius norm on each step of the approximation.

Proof. For a fixed $M \in \mathcal{M}_k$ we can find an $N \in \mathcal{N}$ such that $||M - N||_F \leq \varepsilon$. Note that $M - N$ has rank at most $2k$ since M and N both have rank at most k.

In order to use the identity $||A||_F = \sqrt{\sum_{i=1}^{rank(A)} \sigma_i^2(A)}$, let $U\Sigma V^T$ be the singular decomposition of $M - N$ and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be its corresponding linear transformation such that $T(\mathbf{v}_i) = \sigma_i(M - N)\mathbf{u}_i$. Decompose T as the sum $T_1 + T_2$ with

$$T_j(\mathbf{v}_i) = \begin{cases} \sigma_i(M-N)\mathbf{u}_i, & 1 + (j-1)k \leq i \leq k + (j-1)k \\ 0 & \text{otherwise.} \end{cases}$$

Then, we can write $M - N = M_1 + M_2$ with M_j being the corresponding matrix of T_j. Moreover, $M_j \in \mathcal{M}_k$ since each one has rank at most k and

$$||M_j||_F = \sqrt{\sum_{i=1}^{rank(M_j)} \sigma_i^2(M_j)} \leq \sqrt{\sum_{i=1}^{rank(M-N)} \sigma_i^2(M-N)} = ||M-N||_F \leq \varepsilon < 1.$$

With this result, we can rewrite M as the sum of a element of the net and two small multiples of elements from $\mathcal{M}_k$

$$M = N + \varepsilon M_1' + \varepsilon M_2'$$

where we normalized the M_j's by ε, for calculation purposes, maintaining the Frobenius norm at most 1. We can repeat this approximation again for each M_j and find $M_j = N_j + \varepsilon M_{j1} + \varepsilon M_{j2}$ so that on step k we have

$$M = \sum_{i=0}^{k-1} \sum_{s \in S_i} \varepsilon^{|s|} N_s + \varepsilon^k \sum_{s \in S_k} M_s$$

where S_i is the set of strings of 1's and 2's of length i. Note that the right sum goes to zero on Frobenius norm since $||M_s||_F \leq 1$ for all $s \in S_k$, $|S_k| = 2^k$ and $\varepsilon < 1/2$. Therefore, the first sum converges to M and

$$M = \sum_{i=0}^{\infty} \sum_{s \in S_i} \varepsilon^{|s|} N_s.$$

Now that we are able to approximate any $M \in \mathcal{M}_k$ by a series of elements of the net (times some decreasing constants), let's do the inner product with

A and get

$$\langle M, A\rangle = \sum_{i=0}^{\infty}\sum_{s\in S_i}\varepsilon^{|s|}\langle N_s, A\rangle \le \max_{n\in\mathcal{N}}\langle N, A\rangle \sum_{i=0}^{\infty}\sum_{s\in S_i}\varepsilon^{|s|} = \max_{n\in\mathcal{N}}\langle N, A\rangle \sum_{i=0}^{\infty} 2^i\varepsilon^i$$
$$= \frac{1}{1-2\varepsilon}\max_{n\in\mathcal{N}}\langle N, A\rangle$$

which follows for any M and we have

$$\sup_{M\in\mathcal{M}_k}\langle M, A\rangle \le \frac{1}{1-2\varepsilon}\max_{N\in\mathcal{N}}\langle N, A\rangle.$$

■

Now we can use a union bound and get the following proposition, which will help us with the closed side.

Proposition 3.12 *Let A be a n-by-n symmetric random matrix having i.i.d. upper diagonal entries. Then for any $k \ge 1$, $0 < \varepsilon < 1/2$, and $t > 0$*

$$\ln \mathbb{P}\left(\sqrt{\sum_{i=1}^{k}\sigma_i^2(A)} > t\right) \le \sup_{M\in\mathcal{M}_k} \ln \mathbb{P}\Big(\langle M, A\rangle > (1-2\varepsilon)t\Big) + O\left(nk\ln\frac{1}{\varepsilon}\right)$$

Proof. Let $\mathcal{N} \subset \mathcal{M}_k$ be an ε-net of size at most $(C/\varepsilon)^{2kn}$. It exists by Corollary 3.10. By Lemma 3.7 and, after, Lemma 3.11

$$\mathbb{P}\left(\sqrt{\sum_{i=1}^{k}\sigma_i^2(A)} > t\right) = \mathbb{P}\Big(\sup_{M\in\mathcal{M}_k}\langle A, M\rangle_F > t\Big) \le \mathbb{P}\Big(\max_{N\in\mathcal{N}}\langle A, N\rangle_F > (1-2\varepsilon)t\Big).$$

Using a union bound over $\mathcal{N}$,

$$\mathbb{P}\Big(\max_{N\in\mathcal{N}}\langle A, N\rangle_F > (1-2\varepsilon)t\Big) \le \sum_{N\in\mathcal{N}}\mathbb{P}\Big(\langle N, A\rangle > (1-2\varepsilon)t\Big)$$
$$\le (C/\varepsilon)^{2kn}\sup_{M\in\mathcal{M}_k}\mathbb{P}\Big(\langle M, A\rangle > (1-2\varepsilon)t\Big).$$

And after applying the ln to both sides, we get the result. Note that we went back to $\mathcal{M}_k$ in the end (a trivial inclusion), but used the net to switch the places of $\mathbb{P}$ and sup. ■

We have just transformed our problem into another! On the next section we are going to find a good upper bound for $\sup_{M\in\mathcal{M}_k}\mathbb{P}\Big(\langle M, A\rangle > (1-2\varepsilon)t\Big)$ using a technique similar to the one used on Hoeffding-Azuma inequality and everything else needed for the LDP.

3.4 All the Tools for the LDP

Let's start with what we promised: the Hoeffding-Azuma argument.

Proposition 3.13 *Let A be a n-by-n symmetric random matrix, with zero diagonal, having i.i.d. upper diagonal entries with the same distribution as ξ, a subgaussian random variable which has a globally finite moment-generating function and a cumulant-generating function $\Lambda_\xi(s) = \ln \mathbb{E}[e^{s\xi}]$. Define $L := \sup\limits_{s>0} \frac{\Lambda_\xi(s)}{s^2}$ and $\ell := \inf\limits_{s>0} \frac{\Lambda_\xi^*(s)}{s^2}$. Then*

$$\sup_{M:||M||_F\leq 1} \mathbb{P}\Big(\langle M, A\rangle > t\Big) \leq \exp\left(-\frac{t^2}{8L}\right) = \exp\left(-\frac{t^2\ell}{2}\right)$$

Proof. It suffices to look at only symmetric matrices M, since $\langle M, A\rangle = \langle \frac{M+M^T}{2}, A\rangle$ and $||\frac{M+M^T}{2}||_F \leq ||M||_F$. So, if we let $m = \binom{n}{2}$ and enumerate the upper diagonal entries of A as ξ_i and of M as a_i, we can see that, since both of them are symmetric, $\langle M, A\rangle = 2\sum\limits_{i=1}^{m} \xi_i a_i$. Therefore, for $s > 0$

$$\begin{aligned}
\mathbb{P}\Big(\langle M, A\rangle > t\Big) &= \mathbb{P}\left(\sum_{i=1}^{m} \xi_i a_i > t/2\right) \\
&= \mathbb{P}\left(\exp\left(s\sum_{i=1}^{m} \xi_i a_i\right) > \exp\Big(st/2\Big)\right) \\
&\leq e^{-st/2}\mathbb{E}\left[\exp\left(s\sum_{i=1}^{m} \xi_i a_i\right)\right] \\
&= \exp\left(\sum_{i=1}^{m} \Lambda_\xi(sa_i) - st/2\right)
\end{aligned}$$

where we used Markov's inequality for the inequality.

We now use the "multiply by one" trick to bound $\sum\limits_{i=1}^{m} \Lambda_\xi(sa_i)$

$$\sum_{i=1}^{m} \Lambda_\xi(sa_i) = \sum_{i=1}^{m} \Lambda_\xi(sa_i)\frac{(sa_i)^2}{(sa_i)^2} \leq s^2\sum_{i=1}^{m} a_i^2 L \leq \frac{s^2 L}{2}$$

where the last inequality came from $||M||_F \leq 1$, and now choose an optimal $s = \frac{t}{2L}$ to get

$$\mathbb{P}\Big(\langle M, A\rangle > t\Big) \leq \exp\left(\frac{t^2}{8L} - \frac{t^2}{4L}\right) = \exp\left(-\frac{t^2}{8L}\right) = \exp\left(-\frac{t^2\ell}{2}\right)$$

where the last equality comes from Lemma 2.17. ■

With propositions 3.12 and 3.13 we are able to build the result that we need for the upper bound part of the LDP.

Corollary 3.14 *With the same setting as Proposition 3.13, for any k and $t > 0$, if $\frac{nk}{t^2\ell} < \frac{1}{2}$, then we have*

$$\ln \mathbb{P}\left(\sqrt{\sum_{i=1}^{k} \sigma_i^2(A)} > t\right) \leq -\frac{t^2\ell}{2} + O\left(nk \ln \frac{t^2\ell}{nk}\right).$$

Proof. Proposition 3.12 gives us that

$$\ln \mathbb{P}\left(\sqrt{\sum_{i=1}^{k} \sigma_i^2(A)} > t\right) \leq \sup_{M \in \mathcal{M}_k} \ln \mathbb{P}\Big(\langle M, A\rangle > (1-2\varepsilon)t\Big) + O\left(nk \ln \frac{1}{\varepsilon}\right).$$

We choose $\varepsilon = \frac{nk}{t^2\ell} < \frac{1}{2}$ and apply Proposition 3.13 to get

$$\begin{aligned}
\ln \mathbb{P}\left(\sqrt{\sum_{i=1}^{k} \sigma_i^2(A)} > t\right) &\leq -\frac{(t-2\varepsilon t)^2\ell}{2} + O\left(nk \ln \frac{1}{\varepsilon}\right) \\
&= -\frac{t^2\ell}{2} + 2nk\left(1 - \frac{1}{t}\right) + O\left(nk \ln \frac{t^2\ell}{nk}\right) \\
&= -\frac{t^2\ell}{2} + O\left(nk \ln \frac{t^2\ell}{nk}\right)
\end{aligned}$$

■

Now we handle the lower bound part of the LDP. As we saw, we need to control the probability of the event $\langle X, w\rangle \geq ||w||^2$, for w being some positive vector. To handle that we are going to control the event $\langle X, w\rangle \geq ||w||\sqrt{t}$ (as if we switched one $||w||$ for $\sqrt{t}$). This transformation helps us as changing the norm of w, in this new setting, does not change the inequality. This can be seen using the equality $\langle X, w\rangle = ||X||.||w||cos\theta$, where θ is the angle between X and w. This property is called *homogeinity in w.*

Using that idea, let's continue.

Proposition 3.15 *With the same setting as Proposition 3.13, suppose that the function $s \mapsto \frac{\Lambda_\xi^*(s)}{s^2}$ minimizes at some finite s. Then, for any $1 \ll t \ll n^2$ and any $w_1, w_2, \ldots, w_k > 0$, we have that*

$$\ln \mathbb{P}\left(\sum_{i=1}^{k} w_i \sigma_i(A) > ||w||\sqrt{t}\right) \geq -\frac{t\ell}{2} - o(t).$$

If the s that minimizes $\frac{\Lambda_\xi^*(s)}{s^2}$ is non-negative, then it works for the eigenvalues as well:

$$\ln \mathbb{P}\left(\sum_{i=1}^{k} w_i \lambda_i(A) > ||w||\sqrt{t}\right) \geq -\frac{t\ell}{2} - o(t).$$

For this proof, let any sequence $(a_i)_1^k$ be represented by a vector $a = (a_1, \dots, a_k)$
Proof. Let s_* be the real number that minimizes $\frac{\Lambda_\xi^*(s)}{s^2}$. Suppose that $s_* \neq 0$ (the case for 0 is done by a continuity argument). Using the homogeneity in w, let $||w|| = \sqrt{t}$ transforming our event into

$$\langle w, \sigma(A)\rangle > t.$$

We now define a sequence $(b_i)_1^k$ such that b_i is the smaller integer such that $b_i - 1 \geq \frac{w_i}{s_*}$. Since s_* is a constant, we have that b_i is comparable to w_i which implies that $||b||$ is comparable to $\sqrt{t}$, as

$$||b||^2 = \sum_{i=1}^{k} b_i^2 \geq \sum_{i=1}^{k} \frac{w_i^2}{s_*^2} = \frac{t}{s_*^2}.$$

The sequence defined above was made in order to build the following n-by-n block matrix M. We are going to use k matrices of size b_i-by-b_i, $1 \leq i \leq k$, each one with all entries being equal to s_*. We fit each block, in order, diagonally in M and put all the leftover entries as being 0. Our choice of the b_i's guarantees that all the matrices will fit inside M as $\sum\limits_{i=1}^{k} b_i \leq \sqrt{k}||b|| \ll n$, since $||b|| \approx \sqrt{t}$ and $t \ll n^2$ from hypotheses.

By the construction of M we have that $rank(M) \leq k$ and $\sigma_i(M) = s_* b_i$, not necessarily in decreasing order. By the definition of b, we have

$$w_i \leq \sigma_i(M) - s_* \leq w_i + s_*^2 c$$

where c is the smallest natural number such that $s_*^2 c \geq 1$. Define $T := \sum\limits_{i=1}^{k} \binom{b_i}{2}$. This is the number of non-zero upper diagonal elements of M. Since they are all equal to s_* we have that the distribution of $\langle A, M\rangle_F$ is equals to $2s_* \sum_{i=1}^{T} \xi_i$, where the ξ_i are the respective entries of A which weren't multiplied by a null entry of M. Hence, we have

$$\mathbb{P}\Big(\langle A, M\rangle_F > t\Big) = \mathbb{P}\left(\sum_{i=1}^{T} \xi_i > \frac{t}{2s_*}\right).$$

In order to use Cramér's Theorem, Theorem 2.16, we need to normalize the left hand size of above event by T. We can do that as $T = \frac{||b||^2}{2} - \frac{\sum_{i=1}^{k} b_i}{2}$ and

$$\frac{t}{2s_*^2} = \frac{\sum_{i=1}^{k} w_i^2}{2s_*^2} \leq 2^{-1} \sum_{i=1}^{k} (b_i - 1)^2 = T - \frac{\sum_{i=1}^{k} b_i}{2} + k.$$

Since $\sum_{i=1}^{k} b_i$ goes to infinite, we can guarantee that $\frac{t}{2s_*^2} \leq T$ for n big enough.

Now we can change the inequality with a normalization by T and get

$$\begin{aligned}
\ln \mathbb{P}\Big(\langle A, M\rangle_F > t\Big) &= \ln \mathbb{P}\left(\operatorname{sgn}(s_*)\sum_{i=1}^{T}\xi_i > \frac{t}{2|s_*|}\right) \\
&\geq \ln \mathbb{P}\left(T^{-1}\operatorname{sgn}(s_*)\sum_{i=1}^{T}\xi_i > |s_*|\right) \\
&= -T\Lambda_\xi^*(s_*) + o(T) \\
&= -\frac{t\Lambda_\xi^*(s_*)}{2s_*^2} - o(t) = -\frac{t\ell}{2} - o(t).
\end{aligned}$$

Finally, we can apply property 2 of 3.6 and Cauchy-Schwarz inequality to get

$$\begin{aligned}
\langle A, M\rangle_F \leq \sum_{i=1}^{k}\sigma_i(A)\sigma_i(M) \leq \sum_{i=1}^{k}\sigma_i(A)(w_i + s_*^2 c) &\leq \sum_{i=1}^{k}\sigma_i(A)w_i + s_*^2 c\sum_{i=1}^{k}\sigma_i(A) \\
&\leq \sum_{i=1}^{k}\sigma_i(A)w_i + s_*^2 c\sqrt{k}\sqrt{\sum_{i=1}^{k}\sigma_i^2(A)}
\end{aligned}$$

and we can write

$$\begin{aligned}
\mathbb{P}(\langle A, M\rangle_F > t) &\leq \mathbb{P}\left(\sum_{i=1}^{k}\sigma_i(A)w_i + s_*^2 c\sqrt{k}\sqrt{\sum_{i=1}^{k}\sigma_i^2(A)} > t\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{k}\sigma_i(A)w_i > t - t^{2/3}\right) + \mathbb{P}\left(s_*^2 c\sqrt{k}\sqrt{\sum_{i=1}^{k}\sigma_i^2(A)} > t^{2/3}\right)
\end{aligned}$$

We, now, can apply Corollary 3.14 on the second term of the left hand size of the above equation to get $\mathbb{P}\left(\sum_{i=1}^{k}\sigma_i^2(A) > \frac{t^{4/3}}{s_*^4 c^2 k}\right) = \Theta\Big(\exp(-t^{4/3})\Big)$. Our final step is bounding $\ln \mathbb{P}\left(\sum_{i=1}^{k}\sigma_i(A)w_i > t - t^{2/3}\right)$ from below. We will use that $\ln(x-y) = \ln x \ln(1 - \frac{y}{x})$ to get

$$\begin{aligned}
\ln \mathbb{P}\left(\sum_{i=1}^{k}\sigma_i(A)w_i > t - t^{2/3}\right) &\geq \ln\left[\mathbb{P}\Big(\langle A, M\rangle_F > t\Big) - \mathbb{P}\left(s_*^2 c\sqrt{k}\sqrt{\sum_{i=1}^{k}\sigma_i^2(A)} > t^{2/3}\right)\right] \\
&= \ln \mathbb{P}\Big(\langle A, M\rangle_F > t\Big) + \ln\left(1 + \frac{\mathbb{P}\left(s_*^2 c\sqrt{k}\sqrt{\sum_{i=1}^{k}\sigma_i^2(A)} > t^{2/3}\right)}{\mathbb{P}(\langle A, M\rangle_F > t)}\right) \\
&\geq -\frac{t\ell}{2} - o(t) + \ln\left(1 - \Theta\left(\frac{\exp(-t^{4/3})}{\exp(-t)}\right)\right) \\
&\geq -\frac{t\ell}{2} - o(t) - o(1) \\
&= -\frac{t\ell}{2} - o(t)
\end{aligned}$$

and changing $t - t^{2/3}$ by $|w|\sqrt{t}$, which is equals to t, only hurts us by a small error factor that is absorbed by $o(t)$.

For the eigenvalue part, remember the proof of Lemma 3.7 and write $A = A^+ - A^-$. Write $\lambda_i^+(A) = \max\{0, \lambda(A_i)\}$ as the eigenvalues of A^+. Also note that A is symmetric, then $|\lambda_i(A)| = \sigma_i(A)$, and M is symmetric and positive semi-definite, since $s_* > 0$, then $\lambda_i(M) = \sigma_i(M)$. Therefore, using properties 2 and 3 of Theorem 3.6

$$\begin{aligned}\langle A, M\rangle_F \leq \langle A_+, M\rangle_F \leq \sum_{i=1}^k \lambda_i^+(A)\lambda_i(M) &\leq \sum_{i=1}^k \lambda_i^+(A)(w_i + s_*^2 c)\\ &\leq \sum_{i=1}^k \lambda^+(A)w_i + s_*^2 c\sum_{i=1}^k \lambda^+(A)\\ &\leq \sum_{i=1}^k \lambda^+(A)w_i + s_*^2 c\sqrt{k}\sqrt{\sum_{i=1}^k \lambda_i^+(A)^2}\end{aligned}$$

the rest of the proof is the same as in the singular value case. ■

Now we can finish the general subgaussian LDP.

3.5
Proof of general LDP and the LDP for G(n,m)

Using the results from the last section, we are now able to prove the general LDP Theorem 3.4. As this theorem does not includes the $G(n, m)$ case, we will need another LDP result for it. Fortunately, the same bound holds.

Proof.[of Theorem 3.4] We start with the upper bound side, E being a closed set, and use Corollary 3.14 to get

$$\begin{aligned}\ln \mathbb{P}\Big(X \in E\Big) &\leq \ln \mathbb{P}\left(\sqrt{\sum_{i=1}^k \sigma_i^2} > m_n t\right)\\ &\leq -\frac{m_n^2 t^2 \ell}{2} + O\left(nk \ln \frac{t^2 m_n^2 \ell}{nk}\right)\\ &= -\frac{m_n^2 t^2 \ell}{2} + O\left(n \ln \frac{m_n^2}{n}\right)\\ &= -\frac{m_n^2 t^2 \ell}{2} + o(m_n^2).\end{aligned}$$

where the first inequality came from our first discussion 3-1 about this theorem, and we have used that $\frac{k}{t^2\ell}$ is smaller than any function that goes to infinite, since $\frac{k}{t^2\ell} \leq \frac{m_n^2}{2n}$ and $\sqrt{n} \ll m_n \ll n$. It only lasts to prove the lower bound part, E being an open set.

Remember that we have shown that, if $w \in E$, then

$$\ln \mathbb{P}\Big(X \in E\Big) \geq \ln \mathbb{P}\Big(||X||^2 \leq ||w||^2 + \varepsilon \text{ and } \langle X, w\rangle \geq ||w||^2\Big)$$

and that we need to bound the probability of each event separately. For the first one we can use, again, Corollary 3.14 to get

$$\begin{aligned} \ln \mathbb{P}\Big(||X||^2 > ||w||^2 + \varepsilon\Big) &= \ln \mathbb{P}\left(\sum_{i=1}^{k} \sigma_i^2 > m_n^2(||w||^2 + \varepsilon)\right) \\ &\leq -\frac{m_n^2(||w||^2+\varepsilon)\ell}{2} + O\left(nk \ln \frac{m_n^4(||w||^2+\varepsilon)^2\ell}{nk}\right) \\ &= -\frac{m_n^2(||w||^2+\varepsilon)\ell}{2} + O\left(n \ln \frac{m_n^4}{n^2}\right) \\ &= -\frac{m_n^2(||w||^2+\varepsilon)\ell}{2} + o\left(m_n^2\right), \end{aligned} \tag{3-2}$$

where we use that $\frac{k}{(||w||^2+\varepsilon)^2\ell} \leq \frac{m_n^4}{n}$ and since $\sqrt{n} \ll m_n \ll n$ and m_n is equals to $\sqrt{n}f(n)$, where $\sqrt{n} \gg f(n) \gg 1$.

For the second event, we will use Proposition 3.15 to get

$$\begin{aligned} \ln \mathbb{P}\Big(\langle X, w\rangle \geq ||w||^2\Big) &= \ln \mathbb{P}\left(\sum_{i=1}^{k} \sigma_i w_i \geq \Big(m_n ||w||\Big)||w||\right) \\ &\geq -\frac{m_n^2||w||^2\ell}{2} - o(m_n^2) \end{aligned}$$

where we only needed to set $t = m_n||w||$.

With those results, we see that the event $\langle X, w\rangle \geq ||w||^2$ has a bigger probability than the event $||X||^2 > ||w||^2 + \varepsilon$. Therefore, the event $\langle X, w\rangle \geq ||w||^2$ dominates the probability and we have

$$\begin{aligned} \ln \mathbb{P}\Big(X \in E\Big) &\geq \ln \mathbb{P}\left(||X||^2 \leq ||w||^2 + \varepsilon \text{ and } \langle X, w\rangle \geq ||w||^2\right) \\ &\geq -\frac{m_n^2||w||^2\ell}{2} - o(m_n^2) \end{aligned}$$

as we wanted and this finish the proof for the Singular Values.

For the second part, the vector $Y = (\lambda_1, \ldots, \lambda_k)/m_n$, observe that we have the extra hypotheses that the s that achieves ℓ is non-negative. Therefore, we can use the second part of Proposition 3.15 on equation 3-2, changing X by Y. The rest of the proof is the same and the result for Eigenvalues is established as well.

■

This theorem works perfectly for the centralized matrix of $G(n, p)$, however we are dealing with $G(n,m)$ instead and need to ensure that it has exactly m positive upper diagonal entries. What we are going to do is to show that the same bounds proved above holds for the $G(n,m)$ as well and compute the value of ℓ for a complete result.

Before starting, we make a quick remark. Since we will work with triangle count with the most negative eigenvalues, we need to switch the order of them because our result uses only the first k of them. For that, we change p by $q = 1-p$ and work with the complement of $G(n,p)$ and $G(n,m)$, which are the same in distribution as $G(n,q)$ and $G(n, q\binom{n}{2})$, respectively. It will switch the order because its the same thing as working with $\mathbb{E}[A_n] - A_n$ instead of $A_n - \mathbb{E}[A_n]$.

The first thing that we have to find for the LDP is the value of ℓ. After that we can state and prove it.

Lemma 3.16 *Let ξ be a random variable such that $\xi = -q$ with probability $1-q$ and $\xi = 1-q$ with probability q. The*

$$\ell = \inf_s \frac{\Lambda_\xi^*(s)}{s^2} = \frac{\ln \frac{1-q}{q}}{1-2q}$$

achieving its minimum at $s = 1-2q$.

Proof. We know that

$$\Lambda_\xi^*(s) = (q+s)\ln\frac{q+s}{q} + (1-q-s)\ln\frac{1-q-s}{1-q}$$

and we can change parameters $r = s+q$ so that we have to minimize

$$M(r) = \frac{\Lambda_\xi^*(r-q)}{(r-q)^2}$$

where $M(q)$ is defined by continuity.

Now we compute some derivatives:

$$M'(r) = -\frac{(q+r)\ln\frac{r}{q} + (2-q-r)\ln\frac{1-r}{1-q}}{(r-q)^3}$$

and we define $N(r)$ to be its numerator. We, then, search for the roots of $N(r)$ in order to find the roots of $M'(r)$ and its minimums.

$$N'(r) = \ln\frac{r}{q} - \ln\frac{1-r}{1-q} + \frac{q}{r} - \frac{1-q}{q-r}$$

and

$$N''(r) = (r-q)\left(\frac{1}{r^2} - \frac{1}{(1-r)^2}\right).$$

Therefore, the second derivative $N''(r)$ has two roots, being $1/2$ and q. This implies that $N(r)$ has at most 4 roots.

We can observe that $N(r)$ vanishes at q and $1-q$ and that $N'(r)$ and $N''(r)$ vanishes at q as well. So $N(r)$ has a triple root at q and a single root at $1-q$. These are all the roots.

As $M'(q) \neq 0$ by definition, we have that the only root of $M'(r)$ is $1-q$. Since the function is defined on $[0,1]$ we have the possibility that the minimum is achieved by either of these values: 0, $1-q$ and 1.

We, finally, only need to show that

$$M'(1-q) = \frac{\ln\frac{1-q}{2}}{1-2q}$$

is the minimum value. Since the other two values are

$$\frac{1}{q^2}\ln\frac{1}{1-q} \text{ and } \frac{1}{(1-q)^2}\ln\frac{1}{q}$$

and q and $1-q$ are symmetric in $[0,1]$, we only need to prove that one of them is bigger then our desired value for all q.

Define

$$f(q) := q^2(1-2q)\left(\frac{\ln\frac{1-q}{2}}{1-2q} - \frac{1}{q^2}\ln\frac{1}{1-q}\right) = (1-q)^2\ln(1-q) - q^2\ln q.$$

Since $f(q) = -f(1-q)$, we only need to prove that $f(q) < 0$ when $0 < q < 1/2$. This is true because $f(0) = f(1/2) = 0$ and $f''(q) = -2\ln q + 2\ln(1-q) > 0$. ■

Finally, the LDP. We will quickly revisit our notation. Here we will prove a LDP for the vectors $(\sigma_1, \sigma_2, \ldots, \sigma_k)/m_n$ and $(\lambda_1, \lambda_2, \ldots, \lambda_k)/m_n$, where k is a constant and $\sqrt{n} \ll m_n \ll n$. So, now, we are proving a LDP for a constant number of Singular Values and Eigenvalues of $G(n,m)$'s centralized adjacency matrix.

Theorem 3.17 (LDP for $G(n,m)$) *Let A_n be a the centered adjacency matrix of $G(n,m)$ and fix $q \in (0,1)$ such that $|m - q\binom{n}{2}| = O(1)$. For every integer $k \geq 1$ and sequence m_n such that $\sqrt{n} \ll m_n \ll n$, we have that the sequence of vectors*

$$X = (\sigma_1, \sigma_2, \ldots, \sigma_k)/m_n$$

satisfies the LDP with speed m_n^2 and good rate function $I : \mathbb{R}_+^k \to [0,\infty)$ with

$$I(x) = \begin{cases} \frac{|x|^2}{2} \frac{\ln \frac{1-q}{q}}{1-2q}, & \text{if } q \neq \frac{1}{2} \\ |x|^2, & \text{if } q = \frac{1}{2}. \end{cases}$$

If, in addition, $q \leq \frac{1}{2}$, then the sequence

$$Y = (\lambda_1, \lambda_2, \ldots, \lambda_k)/m_n$$

has the same LDP properties.

Proof. Let A_q be the centered adjacency matrix of $G(n,q)$.

For the lower bound side, observe that since $|m - q\binom{n}{2}| = O(1)$ we can say that for a given event F, $\mathbb{P}(A_n \in F) = \mathbb{P}(A_q \in F|M)$, where M is the event that $G(n,q)$ has exactly m edges. This gives us that

$$\mathbb{P}\Big(A_n \in F\Big) = \mathbb{P}\Big(A_q \in F|M\Big) = \frac{\mathbb{P}(A_q \in F \text{ and } M)}{\mathbb{P}(M)} \leq \frac{\mathbb{P}(A_q \in F)}{\mathbb{P}(M)}$$

and calculating $\mathbb{P}(M)$ using Stirling's Approximation

$$\begin{aligned}
\mathbb{P}(M) &= \binom{N}{m} p^m (1-p)^{N-m} \\
&\approx \frac{\sqrt{2\pi N}\left(\frac{N}{e}\right)^N}{\sqrt{2\pi(N-m)}\left(\frac{N-m}{e}\right)^{N-m}\sqrt{2\pi m}\left(\frac{m}{e}\right)^m} q^m (1-q)^{N-m} \\
&\approx \frac{\sqrt{2\pi N}\left(\frac{N}{e}\right)^N}{\sqrt{2\pi(1-q)N}\left(\frac{(1-q)N}{e}\right)^{(1-q)N}\sqrt{2\pi qN}\left(\frac{qN}{e}\right)^{qN}} q^{qN} (1-q)^{(1-q)N} \\
&= \frac{1}{\sqrt{2\pi q(1-q)N}} \\
&\approx \frac{1}{n}.
\end{aligned}$$

and then

$$\ln \mathbb{P}\Big(A_n \in F\Big) \leq \ln \mathbb{P}\Big(A_q \in F\Big) - O(\ln n)$$

and the bound of the LDP holds. It only remains to show the upper bound side.

For this side we used Proposition 3.15 and it was done using the distribution of $\langle A_q, M\rangle$, where A_q is the adjacency matrix of $G(n,q)$, which is equal to $2s_* \sum_{i=1}^{T} \xi_i$. The point is: this is the binomial distribution, with some rescaling and translation, with T trials. For the adjacency matrix of $G(n,m)$, it would be a hypergeometric distribution, with some rescaling and translation, with m trials and T possible successes. Moreover, we know that, by our construction, the size of the blocks implies that $T = o(n^2)$. Let $H_{N,T,m}$ denote the hypergeometric random variable with population size N, T trials and m possible successes; and $B_{q,T}$ be the binomial distribution with probability q and T trials. Using one more extensive Stirling's Approximation calculation

$$\Big|\ln \mathbb{P}(H_{N,T,m} = s) - \ln \mathbb{P}(B_{m/N,T} = s)\Big| = O\left(\frac{T^2}{N}\right)$$

and in Proposition 3.15 we have that this $\frac{T^2}{N} = O\left(\frac{t^2}{n^2}\right) = o(t)$, since $t \ll n^2$. Therefore the error term is sufficiently small and we have

$$\ln \mathbb{P}\Big(\langle A_n, M\rangle > t\Big) \geq \ln \mathbb{P}\Big(\langle A_q, M\rangle > t\Big) - o(t).$$

Therefore the LDP works for $G(n, m)$ as well.

Finally, the constant comes from the LDP for $G(n, q)$ and is $\frac{\ln \frac{1-q}{q}}{1-2q}$ provided by Lemma 3.16. ■

3.6 A Very Long Proof of Theorem 3.1

We can finally start to prove the principal theorem of this chapter! Note that all the theorems, propositions, lemmas and corollaries that led to the LDP's theorems for only a constant number of extreme eigenvalues. Therefore, we need to show that the other part, which we will call "the bulk", of them are non-important for the triangle count. The bulk is going to be the eigenvalues that are greater than $-\Omega(\sqrt{Kn})$ for some big K. After controlling them, we need to show that there is only a constant number of eigenvalues that influence the triangle count, and that they are the eigenvalues lesser than $-\Omega(\sqrt{Kn})$. We will do it transforming Corollary 3.14 into a bound on the eigenvalues' values instead of a constant number of them.

3.6.1 Preliminaries

Here we denote $f(A)$ with f a function and A a matrix as $f(A) := Uf(D)V^T$ where UDV^T is the eigendecomposition of A and $f(D)$ is f applied on each diagonal entry.

For our next result, we will need the following theorem, which we won't prove but we give a reference for it after its number. It essentially says that the trace of $f(n^{-1/2}A_n)$, where f is Lipschitz, is concentrated.

Theorem 3.18 ([12]) *Let $||\xi||_\infty < \infty$, let $f : \mathbb{R} \to \mathbb{R}$ be a K-Lipschitz, convex function and let A_n an n-by-n random symmetric matrix with iid upper diagonal entries with distribution ξ and diagonal entries equal to zero.*

If $X_n = n^{-1}\sum_{i=1}^{n} f(n^{-1/2}\lambda_i(A_n))$, then there exists a universal constant C such that

$$\mathbb{P}\Big(|X_n - \mathbb{E}[X_n]| \geq \delta\Big) \leq C \exp\left(-\frac{n^2\delta^2}{C||\xi||_\infty K^2}\right)$$

if $\delta \gg Kn^{-1}$.

With this result we can compute the contribution of the bulk of the eigenvalues.

Lemma 3.19 *Let A_n be the centered adjacency matrix of $G(n,m)$. Then, if K is sufficiently large and $s \gg 3Kn^{-1}$,*

$$\mathbb{P}\left(\sum_{\lambda_i(A_n)\geq -\sqrt{Kn}} \lambda_i^3(A_n) < -s - O(n^2)\right) \leq \exp\left(-\frac{s^2}{n^3K^2}\right).$$

Proof. Again, it will be more convenient to work in $G(n,p)$ rather than $G(n,m)$. Fortunately, the argument given in the proof of Theorem 3.17 again allows the $G(n,p)$ result to transfer to the $G(n,m)$ setting. Therefore, we work with A_n being the centered adjacency matrix of a random graph distributed as $G(n,p)$.

In order to use Theorem 3.18, we need to define a convex, Lipschitz function that is convenient for our application. We wish to consider the cubes of the eigenvalues. However, as the function $x \to x^3$ is not Lipschitz we cannot apply Theorem 3.18 directly with this function. Instead we define *two* convex, Lipschitz functions f_1 and $-f_2$ such that $(f_1 + f_2)(x) = x^3$ in a large interval $[-\sqrt{K}, \sqrt{K}]$. Set:

$$f_1(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^3 & \text{if } 0 < x \leq \sqrt{K} \\ 3Kx - 2K^{3/2} & \text{if } \sqrt{K} < x. \end{cases}$$

The third part was defined to maintain the convexity and the Lipschitz condition as it is just the line which is tangent to x^3 on point $\sqrt{K}$.

Defining $f_2(x) = -f_1(-x)$, we have that $f_2(x)$ is concave, but $-f_2(x)$ is convex. We can still use Theorem 3.18 on the difference of two convex functions, since we can apply it to each of the functions separately and then apply an union bound, increasing the value of C. Therefore we can use the function $f_1 + f_2$.

Note that

$$(f_1+f_2)(x) \le \begin{cases} 0 & \text{if } x \le -\sqrt{K} \\ x^3 & \text{if } -\sqrt{K} \le x \end{cases}$$

and then we get a bound on the sum

$$\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) \;\le\; n^{-3/2} \sum_{\lambda_i(A_n)\ge -\sqrt{Kn}} \lambda_i^3(A_n). \tag{3-3}$$

Now, we use Theorem 3.18 with f_1+f_2 and $s \gg 3Kn^{-1}$ to get

$$\mathbb{P}\left(n^{-1}\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) \le n^{-1}\mathbb{E}\left[\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n))\right]-s\right) \le \exp\left(-\Omega\left(\frac{n^2s^2}{K^2}\right)\right).$$

But if we use 3-3, we get

$$\mathbb{P}\left(\sum_{\lambda_i(A_n)\ge -\sqrt{Kn}} \lambda_i^3(A_n) \le n^{3/2}\mathbb{E}\left[\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n))\right]-s\right) \le$$

$$\mathbb{P}\left(n^{3/2}\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) \le n^{3/2}\mathbb{E}\left[\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n))\right]-s\right) =$$

$$\mathbb{P}\left(n^{-1}\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) \le n^{-1}\mathbb{E}\left[\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n))\right]-sn^{-5/2}\right) \le$$

$$\exp\left(-\Omega\left(\frac{s^2}{n^3K^2}\right)\right).$$

Summarizing,

$$\mathbb{P}\left(\sum_{\lambda_i(A_n)\ge -\sqrt{Kn}} \lambda_i^3(A_n) \le n^{3/2}\mathbb{E}\left[\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n))\right]-s\right) \le \exp\left(-\Omega\left(\frac{s^2}{n^3K^2}\right)\right).$$

It only remains to control $n^{3/2}\mathbb{E}[\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n))]$. If we show that $\mathbb{E}[\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n))] = O(\sqrt{n})$, then we are done.

Another useful inequality that we obtain from the definition of f_1+f_2 is

$$|(f_1+f_2)(x)-x^3| \le \begin{cases} 0 & \text{if } |x| \le \sqrt{K} \\ x^3 & \text{if } \sqrt{K} \le |x| \end{cases}$$

Hence,

$$
\begin{aligned}
\left|\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) - n^{-3/2}\sum_{i=1}^{n}\lambda_i^3(A_n)\right| &= \left|\sum_{i=1}^{n}(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) - n^{-3/2}\lambda_i^3(A_n)\right| \\
&= \left|\sum_{|\lambda_i|>\sqrt{nK}}(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) - n^{-3/2}\lambda_i^3(A_n)\right| \\
&\leq \sum_{|\lambda_i|>\sqrt{nK}}\left|(f_1+f_2)(n^{-1/2}\lambda_i(A_n)) - n^{-3/2}\lambda_i^3(A_n)\right| \\
&\leq \sum_{|\lambda_i|>\sqrt{nK}}\left|n^{-3/2}\lambda_i^3(A_n)\right| \\
&\leq n^{-1/2}\sigma_1^3(A_n)\mathbb{I}_{\sigma_1(A_n)\geq\sqrt{Kn}}
\end{aligned}
$$

where we used that all $\lambda_i(A_n)$ may be bigger than $\sqrt{Kn}$ and that $\sigma_1(A_n) \geq |\lambda_i(A_n)|$ for all i.

We now take expectations and bound from above

$$\mathbb{E}\Big[\sigma_1^3(A_n)\mathbb{I}_{\sigma_1(A_n)\geq\sqrt{Kn}}\Big] \;\leq\; n^3\mathbb{P}\Big(\sigma_1(A_n) \geq \sqrt{Kn}\Big).$$

Using Corollary 3.14 with $k = 1$ and $t = K\sqrt{n}$, if K is sufficiently large, then we get the bound

$$\mathbb{E}\Big[\sigma_1^3(A_n)\mathbb{I}_{\sigma_1(A_n)\geq\sqrt{Kn}}\Big] \;\leq\; n^3\mathbb{P}\Big(\sigma_1(A_n) \geq \sqrt{Kn}\Big) \;\leq\; \exp(-\Omega(n)).$$

Therefore, we finally get

$$\mathbb{P}\left(\sum_{\lambda_i(A_n)\geq-\sqrt{Kn}}\lambda_i^3(A_n) < \mathbb{E}\Big[tr[A_n^3]\Big] - s + \exp(-\Omega(n))\right) \leq \exp\left(\frac{s^2}{n^3K^2}\right)$$

which is what we wanted since $\mathbb{E}\Big[tr[A_n^3]\Big] = O(n^2)$. ∎

Now we need to we take care of the quantity of singular values above a certain bound transforming Corollary 3.14 into

Corollary 3.20 *Using the notation of* 3.14, *then*

$$\ln\mathbb{P}\left(\sqrt{\sum_{\sigma_i(A)>\sqrt{Kn}}\sigma_i^2(A)} \geq t\right) \leq -\frac{t^2\ell}{2} + O\left(\frac{t^2}{K}\ln K\right).$$

The result still holds for A_n being the adjacency matrix of $G(n,m)$.

Proof. The proof is the following: If we suppose that each of the first $k = \left\lceil\frac{t^2}{Kn}\right\rceil$

singular values are bigger than $\sqrt{Kn}$, then

$$\sum_{i=1}^{k} \sigma_i^2(A) \geq k\sqrt{Kn} \geq t^2,$$

Therefore, we have two possibilities:

- either $\sum\limits_{\sigma_i(A)>\sqrt{Kn}} \sigma_i^2(A) \leq \sum_{i=1}^{k} \sigma_i^2(A)$, i.e. there are less than k singular values bigger than $\sqrt{Kn}$;
- or $\sum_{i=1}^{k} \sigma_i^2(A) \geq t^2$.

Hence, we get that

$$\ln \mathbb{P}\left(\sqrt{\sum_{\sigma_i(A)>\sqrt{Kn}} \sigma_i^2(A)} \geq t\right) \leq \ln \mathbb{P}\left(\sqrt{\sum_{i=1}^{k} \sigma_i^2(A)} \geq t\right) \leq -\frac{t^2\ell}{2} + O\left(\frac{t^2}{K}\ln K\right)$$

by Corollary 3.14.

Finally, using again that if we condition $G(n,q)$ on having $q\binom{n}{2}$ edges, then we only lose a small summand of $\ln(n)$, we get our result. ■

We are ready to start the proof now.

3.6.2 The Proof

In order to prove Theorem 3.1, we will need to prove an upper bound on the desired probability and a matching lower bound. Then we will need to handle the last statement showing that the most negative eigenvalue is, indeed, much more negative than all the others.

3.6.2.1 The Upper Bound

Remember that Corollary 3.3 gave us a upper bound on the probability of a certain triangle count deviation using the corresponding bound for the trace of the cube of centered adjacency matrix.

Let A_n be the adjacency matrix of $G(n,m)$ and $\tilde{A}_n$ be the centered one. We use Corollary 3.3 and that $\tau(G) = \frac{tr[A_n^3]}{\binom{n}{3}} = \frac{tr[A_n^3]}{n^3} + O(1/n)$ to get

$$\mathbb{P}\Big(\tau(G) \leq p^3 - t\Big) = \mathbb{P}\Big(tr[A_n^3] \leq n^3p^3 - tn^3 + O(n^2)\Big) \leq \mathbb{P}\Big(tr[\tilde{A}_n^3] \leq -tn^3 + O(n^2)\Big).$$

In order to find this probability, we need to bound the sum of the cubes of the centered adjacency matrix's eigenvalues, since $tr[\tilde{A}_n^3] = \sum_{i=1}^{n} \lambda_i^3(\tilde{A}_n)$.

Let's start with the extremal eigenvalues (letting $\varepsilon \to 0$ and $K \to \infty$ with orders chosen afterwards):

$$\mathbb{P}\left(\sum_{\lambda_i(\tilde{A}_n)<-\sqrt{Kn}} \lambda_i^3(\tilde{A}_n) < -(1-\varepsilon)tn^3\right) \leq \mathbb{P}\left(\sqrt{\sum_{\sigma_i(\tilde{A}_n)>\sqrt{Kn}} \sigma_i^2(\tilde{A}_n)} > (1-\varepsilon)^{1/3}t^{1/3}n\right)$$
$$\leq \exp\left(-\frac{\ell}{2}t^{2/3}n^2 + o(t^{2/3}n^2)\right)$$

where we used Corollary 3.20 in the last inequality. The first inequality came from Jensen's Inequality on

$$\left|\sum_{\lambda_i(\tilde{A}_n)<-\sqrt{Kn}} \lambda_i^3(\tilde{A}_n)\right| \leq \left(\sum_{\lambda_i(\tilde{A}_n)<-\sqrt{Kn}} \lambda_i^2(\tilde{A}_n)\right)^{3/2} \leq \left(\sqrt{\sum_{\sigma_i(\tilde{A}_n)>\sqrt{Kn}} \sigma_i^2(\tilde{A}_n)}\right)^3 .$$

Now we control the bulk with Lemma 3.19:

$$\mathbb{P}\left(\sum_{\lambda_i(\tilde{A}_n)\geq-\sqrt{Kn}} \lambda_i^3(\tilde{A}_n) < -\varepsilon tn^3\right) \leq \exp\left(-\Omega\left(\frac{\varepsilon^2t^2n^3}{K^2}\right)\right) = \exp\Big(-\omega(t^{2/3}n^2)\Big)$$

setting $\frac{\varepsilon}{K} = \omega(n^{-1/2}t^{-2/3})$. This choice is made so that the probability of the bulk doesn't affect our result significantly.

Combining both results yields

$$\ln \mathbb{P}\Big(tr[\tilde{A}_n^3] \leq -tn^3\Big) \leq -\frac{\ell t^{2/3}n^2}{2}(1+o(1))$$

and therefore

$$\ln \mathbb{P}\Big(\tau(G) \leq p^3 - t\Big) \leq -\frac{\ell t^{2/3}n^2}{2}(1+o(1))$$

proving the upper bound.

3.6.2.2 The Lower Bound

For the lower bound we will work with minus the centered adjacency matrix $\mathbb{E}[A_n] - A$ where $q = 1-p$.

We don't have a result like Corollary 3.3 for the lower bound, so our approach is going to be different. We are going to use the following idea: let A and B be events of the probability space. Then $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$. Letting A be the event that $G(n,m)$ has few triangles and B some specific event, then if $\mathbb{P}(A|B)$ is non-negligible compared to $\mathbb{P}(B)$, we can use that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(B)\mathbb{P}(A|B) = \mathbb{P}(A \cap B) \leq \mathbb{P}(A) \tag{3-4}$$

and a lower bound on $\mathbb{P}(B)$ will give a lower bound to $\mathbb{P}(A)$ up to a small order factor.

Our event B is going to be about the behavior of $\tilde{A}_n$ over some blocks of itself. Translating to graph language, this event is about an specific non-balanced bipartition where the smaller part has a lower edge density. With some smart choosing of the parameters for each block, we will get the desired bound and the conditional result. Let's start:

For clarity, remember that $\ell = \inf_s \frac{\Lambda_\xi^*(s)}{s^2}$, that $s^* = 1 - 2q \geq 0$ is the minimizing s from ℓ and let $\xi_1, \ldots, \xi_{\binom{n}{2}}$ be some ordering of the upper diagonal entries of $\tilde{A}_n$. We start by choosing the r-by-r block B_1 which is the top-left corner of $\tilde{A}_n$. Let I_1 be the set of indices of the ξ_i that are in B_1. We complete the covering of $\tilde{A}_n$ with the r-by-$(n-r)$ block B_2, that is on B_1's right, with I_2 analogously defined and B_3 being the $(n-r)$-by-$(n-r)$ block below B_2.

We want to get the parameters so that B_1 has a higher concentration of edges (non-edges of $G(n,m)$) compared to B_2. For such, let $r = z^{1/3}n$ with $r = \frac{nt^{1/3}}{s_*}$ (the closest integer, actually) and we want

$$\sum_{i \in I_1} \xi_i = B_*^1 \tag{3-5}$$

$$\sum_{i \in I_2} \xi_i = B_*^2 \tag{3-6}$$

where $B_*^1 = \left\lfloor s_* \binom{r}{2} \right\rfloor$ and $B_*^2 = \left\lceil s^* z^{1/3} r(n-r) \right\rceil$. Now we are ready to go, since

$$B_*^1 \approx \frac{n^2 t^{2/3}}{2s^*}$$

and

$$B_*^2 \approx \frac{n^2 t^{2/3}}{s_*}.$$

Lemma 3.21 *Let B be the event in which 3-5 and 3-6 happen. Then*

$$\ln \mathbb{P}(B) \geq -\frac{n^2 t^{2/3} \ell}{2}(1 + o(1)).$$

Proof. We are going to use the 3-4 idea. Let B_α be the event such that 3-5 holds and B_β the event where 3-6 holds. Since the ξ_i's are centralized Bernoulli trials, we can use the hypergeometric distribution on $\xi_i + q$. For B_α we have a population of $\binom{n}{2}$, m possible successes, $\binom{r}{2}$ trials and $B_*^1 + q\binom{r}{2}$ successes, since $\sum_{i \in I_1} (\xi_i + q) = q\binom{r}{2} + \sum_{i \in I_1} \xi_i$.

Let H be a hypergeometric random variable with t trials, T population size and δT possible successes, where $\delta \in (0,1)$. Using a very long Stirling approximation, for a integer h

$$\frac{1}{t}\ln \mathbb{P}(H=h) = -D\left(\frac{h}{t},\delta\right) - \frac{1-\frac{t}{T}}{\frac{t}{T}} D\left(\frac{q-\frac{h}{T}}{1-\frac{t}{T}},\delta\right) + O\left(\frac{\ln T}{t}\right)$$

where $D(q+s,q) = \Lambda^*_{\xi_i}(s)$ is the Legendre Transformation of centralized Bernoulli's cumulant-generating function (which is the rate function we've found in the basics).

Translating this result to B_α, we have $t = \binom{r}{2}$, $T = \binom{n}{2}$, $\delta = q$, $h = (s_* + q)\binom{r}{2}$ and $H = \sum\limits_{i\in I_1} \xi_i + q$. Therefore,

$$\begin{aligned}\frac{1}{\binom{r}{2}} \ln \mathbb{P}\left(\sum_{i\in I_1}\xi_i = s_*\binom{r}{2}\right) &= -D\left(s_*+q,q\right) - \frac{1-\binom{r}{2}/\binom{n}{2}}{\binom{r}{2}/\binom{n}{2}} D\left(\frac{q-(s_*+q)\binom{r}{2}/\binom{n}{2}}{1-\binom{r}{2}/\binom{n}{2}},q\right) \\ &\qquad + O\left(\frac{\ln\binom{n}{2}}{\binom{r}{2}}\right) \\ &\to -D\left(s_*+q,q\right) = -\ell s_*^2\end{aligned}$$

when $n\to\infty$, since $\ln n \ll r \ll n$ from choice, $D(q,q)=0$ and s_* being the minimizing value of $\frac{\Lambda^*_\xi(s)}{s^2}$.

Finally,

$$\begin{aligned}\ln \mathbb{P}\left(\sum_{i\in I_1}\xi_i = s_*\binom{r}{2}\right) &= -(1+o(1))\ell s_*^2\binom{r}{2} \\ &= -(1+o(1))\frac{t^{2/3}n^2\ell}{2}.\end{aligned}$$

Proceeding with our strategy, let's compute $\mathbb{P}(B_\beta|B_\alpha)$. Conditioned on B_α, B_β is, if added to q, a hypergeometric distribution with parameters: $r(n-r)\approx nr = n^2 z^{1/3}$ trials, $\binom{n}{2}-\binom{r}{2}\approx n^2$ population size and $m - S_* \approx \binom{n}{2}\left(q-\frac{t^{2/3}}{2s_*}\right)$ possible successes. We use the same Stirling formula from above, which won't fit on the page, and get that

$$\frac{1}{r(n-r)}\ln \mathbb{P}\left(\sum_{i\in I_2}\xi_i = B_*^2\middle| B_\alpha\right) = -\Theta(z^{2/3}) - o(z^{2/3})$$

$$\ln \mathbb{P}\left(\sum_{i\in I_2}\xi_i = B_*^2\middle| B_\alpha\right) = -o(n^2z^{2/3}) = -o(n^2t^{2/3})$$

since $\frac{nt^{1/3}}{s_*} = r = z^{1/3}n$.

Finally,

$$\mathbb{P}(B_\beta|B_\alpha) = \frac{\mathbb{P}(B)}{\mathbb{P}(B_\alpha)}$$

$$\mathbb{P}(B) = \mathbb{P}(B_\beta|B_\alpha)\mathbb{P}(B_\alpha) = \exp\left(-\frac{\ell t^{2/3}n^2}{2} - o(t^{2/3}n^2)\right)$$

as wanted. ■

The next and final step to prove the lower bound is to show that, conditioned on Ω (we change B by Ω from here onward since we are going to use B for a block matrix), $G(n,m)$ has fewer triangles. Let's try to describe the distribution of $G(n,m)$ given Ω. We have tree distributions to do (now over $G(n,m)$, not its complement):

- Block B_1 receives uniformly over its entries $(p - s_*)\binom{r}{2}$ upper diagonal positive entries;
- Block B_2 receives uniformly over its entries $(p + \frac{z^{1/3}}{1-z^{1/3}}s_*)r(n-r) + O(1)$ positive entries;
- Block B_3 receives uniformly over its entries

$$p\binom{n}{2} - (p - s_*)\binom{r}{2} - \left(p + \frac{z^{1/3}}{1 - z^{1/3}}s_*\right) r(n-r) + O(1)$$

$$\approx p\frac{n^2}{2} - p\frac{r^2}{2} + s_*\frac{r^2}{2} - prn - \frac{z^{1/3}}{1 - z^{1/3}}s_*rn + \frac{z^{1/3}}{1 - z^{1/3}}s_*r^2 + pr^2 + O(1)$$

$$= \frac{p}{2}\left(n^2 + -2rn + r^2\right) + s_*\left(\frac{z^{1/3}}{1 - z^{1/3}}r^2 + \frac{r^2}{2} - \frac{z^{1/3}}{1 - z^{1/3}}rn\right) + O(1)$$

$$\approx p\binom{n-r}{2} + s_*\left(\frac{z^{1/3}}{1 - z^{1/3}}r^2 - \frac{r^2}{2} - \frac{z^{1/3}}{1 - z^{1/3}}rn\right) + O(1)$$

$$\approx \left(p - s_*\frac{z^{2/3}}{(1 - z^{1/3})^2}\right)\binom{n-r}{2} + O(1)$$

upper diagonal positive entries (the last part was made comparing $-\frac{z^{1/3}}{1-z^{1/3}}rn$ to n^2 and getting a ratio of $\approx -\frac{z^{2/3}}{1-z^{1/3}}$, since $r = z^{1/3}n$).

The error terms are there only so that every term is an integer and compensation for the $1 - z^{1/3}$ normalization. We are not using s_*, actually, but some constant very close to it such that $(p - s_*)\binom{r}{2}$ is an integer.

Lemma 3.22 *Conditioned on Ω, we have that*

$$\mathbb{E}[\tau(G)] = p^3 - s_*^3 z + o(z)$$

and

$$Var(\tau(G)) = O(n^{-2}).$$

Proof. To compute the number of triangles, we need to know the eigenvalues of A_n. We build a block matrix B with four blocks: an r-by-r block with entries $p - s_*$ for the top left, an $n-r$-by-r and an r-by-$n-r$ block with entries $p + \frac{z^{1/3}}{1-z^{1/3}} s_* + O(n^{-2}z^{-1/3})$ for the sides of the first block, and a $n-r$-by-$n-r$ block with entries $p - \frac{z^{2/3}}{1-z^{2/3}} s_* + O(n^{-2})$ for the last part of B. Note that B has rank 2 and, therefore, has only two non-zero eigenvalues. This matrix agrees with $\mathbb{E}[A_n]$ in every entry that is not in the diagonal.

Now we approximate B by the matrix $\mathbf{1}p - s_* vv^T$, where v is the vector that is equal to 1 on the first r entries and $-\frac{z^{1/3}}{1-z^{1/3}}$ on the rest. These matrix are close to each other as seem using the Frobenius Norm

$$||B - (\mathbf{1}p - s_* vv^T)||_F = \sqrt{0 + 2O(n^{-2}z^{-1}) + O(n^{-2})} = o(1).$$

Now, since the vectors $\bar{\mathbf{1}}$ (all-ones vector) and v are orthogonal ($v^T\bar{\mathbf{1}} = \bar{\mathbf{1}}^T v = 0$) and the matrix $\mathbf{1}$ is equal to $\bar{\mathbf{1}}\bar{\mathbf{1}}^T$, then the matrix $\mathbf{1}p - s_* vv^T$ has eigenvalues pn and $-s_*||v||^2$ since

$$(\bar{\mathbf{1}}\bar{\mathbf{1}}^T p - vv^T s_*)\bar{\mathbf{1}} = \bar{\mathbf{1}}\bar{\mathbf{1}}^T\bar{\mathbf{1}}p - s_* vv^T\bar{\mathbf{1}} = pn\bar{\mathbf{1}}$$

and

$$(\bar{\mathbf{1}}\bar{\mathbf{1}}^T p - s_* vv^T)v = -s_* vv^T v = -s_*\langle v, v\rangle v.$$

Since these matrix are close to each other, Weyl's eigenvalue inequality gives us that the eigenvalues of B are $pn + o(1)$ and $-s||v||^2 + o(1)$. Finally, we find that $tr[B^3] = p^3n^3 - s_*^3||v||^3 = p^3n^3 - s_*^3 n^3 z + O(n^3z^2)$.

Now we find the eigenvalues of $\mathbb{E}[A_n]$. Since B and $\mathbb{E}[A_n]$ disagree only over the entries of the diagonal and only by at most a constant amount, we have that

$$||B - \mathbb{E}[A_n]||_{op} = O(1)$$

where $||\cdot||_{op}$ denotes the operator norm. Therefore, again by Weyl's eigenvalue inequality, the eigenvalues of $\mathbb{E}[A_n]$ are $pn+O(1)$, $-s_*||v||^2+O(1)$ and the other eigenvalues are bounded by those values. Hence, $tr[\mathbb{E}[A_n]^3] = p^3n^3 - s_*||v||^3 = p^3n^3 - s_*^3n^3z + O(n^3z^2)$.

Now we need to compare $tr[\mathbb{E}[A_n]^3]$ and $\mathbb{E}[\tau(G)]$.

Let's expand $tr[\mathbb{E}[A_n]^3]$ in terms of closed walks of length 3 (triangles) as in Theorem 2.7. Let Γ_3 be the set of closed walks of length 3 in K_n. Then

$$tr\Big(\mathbb{E}[A_n]^3\Big) = \sum_{(v_i,v_{i+1},v_{i+2})\in\Gamma_3} \prod_{j=1}^{3} \Big(\mathbb{E}[A_n]\Big)_{v_j,v_{j+1}}.$$

For $\mathbb{E}[\tau(G)]$ its simpler. We already have that

$$\binom{n}{3}\mathbb{E}[\tau(G)] = \sum_{(v_i,v_{i+1},v_{i+2})} \mathbb{P}\Big(\{v_i,v_{i+1}\},\{v_{i+1},v_{i+2}\},\{v_{i+2},v_i\} \in E(G)\Big).$$

Now we only need to show that $\binom{n}{3}\mathbb{E}[\tau(G)] \leq tr(\mathbb{E}[A_n]^3)$. We have that $\mathbb{P}(\{v_i,v_{i+1}\} \in E(G)) = (\mathbb{E}[A_n])_{v_j,v_{j+1}}$ for any i, but as it is a $G(n,m)$ random graph, those probabilities are not independent. But conditioned on the existence of previous edges we have that the probability of existence of an specific edge decreases. Therefore,

$$\mathbb{P}\Big(\{v_i,v_{i+1}\} \in E(G)|\{v_1,v_2\},\ldots,\{v_{i-1},v_i\} \in E(G)\Big) \leq \Big(\mathbb{E}[A_n]\Big)_{v_j,v_{j+1}}$$

and we have what we wanted:

$$\begin{aligned}\binom{n}{3}\mathbb{E}[\tau(G)] &\leq \sum_{(v_i,v_{i+1},v_{i+2})} \mathbb{P}(\{v_i,v_{i+1}\},\{v_{i+1},v_{i+2}\},\{v_{i+2},v_i\} \in E(G)) = tr(\mathbb{E}[A_n]^3)\\ &= p^3n^3 - s_*^3n^3z + O(n^3z^2)\\ \mathbb{E}[\tau(G)] &\approx p^3 - s_*^3z^3 + o(z)\end{aligned}$$

For the variance part, we only need to note that if $T(G)$ is the random variable that denotes the number of triangles in $G(n,m)$, then

$$T(G) = \sum_{S\in\binom{[n]}{3}} \mathbb{I}_{S\in G(n,m)},$$

where S is a triangle in K_n. Then,

$$Var(T(G)) = \sum_{S\in\binom{[n]}{3}} Var(\mathbb{I}_{S\in G(n,m)}) + 2\sum_{i<j} Cov(S_i,S_j) = \sum_{i,j} Cov(S_i,S_j)$$

If two triangles S_1 and S_2 don't have an edge in common, then they are negatively correlated, as we are in $G(n,m)$. Hence,

$$Var(T(G)) \leq \sum_{|S_i\cap S_j|\geq 2} Cov(S_i,S_j).$$

As there are at most $\binom{n}{4}$ such pairs, the covariances are at most 1 and $T(G) = \binom{n}{3}\tau(G)$, we have

$$Var(T(G)) \leq O(n^4)$$

$$Var(\tau(G)) = O(n^{-2})$$

as wanted. ■

Finally, since $\frac{nt^{1/3}}{s_*} = z^{1/3}n$, then $s_*^3 z = t$ and both lemmas imply our lower bound after using Paley-Zygmund inequality, replacing t by $(1 - o(1))t$.

3.6.2.3
The Most Negative Eigenvalue

Remember that for the upper bound we used the inequality $\left|\sum_{\lambda_i(\tilde{A}_n)<-\sqrt{Kn}} \lambda_i^3(\tilde{A}_n)\right| \leq \left(\sum_{\lambda_i(\tilde{A}_n)<-\sqrt{Kn}} \lambda_i^2(\tilde{A}_n)\right)^{3/2}$. In order to get an *equality* in this inequality we need that all but one of the eigenvalues to be zero. As we are working with asymptotics, it should be that one of the eigenvalues is of higher order than the others.

We need the following lemma to show that the eigenvalues of $\tilde{A}_n$ satisfies the claim of Theorem 3.1:

Lemma 3.23 ([18]) *Let $(a_i)_{i\in\mathbb{N}}$ be a sequence of non-negative real numbers in non-increasing order. If $\varepsilon > 0$ and $\sum_{i\geq 2} a_i^3 \geq \varepsilon a_1^3$, then*

$$||a||_2^2 \geq (1+\varepsilon)^{1/3}||a||_3^2$$

For our problem, we will use this lemma on the sequence of extreme eigenvalues taking $a_1 = \lambda_n$. We are also using the notation that $||a||_k^k = \sum_i |a_i|^k$. Now we can prove the claim for $\tilde{A}_n$.

Corollary 3.24 *In the setting of 3.1, with the condition of $\tau(G) \leq p^3 - t$ we have*

$$\lambda_n^3(\tilde{A}_n) < -(1-\varepsilon)tn^3 \text{ and } \lambda_{n-1}^3(\tilde{A}_n) \geq -\varepsilon tn^3$$

with high probability.

Proof. Set N to be the set of indices of eigenvalues smaller then $-\Omega(\sqrt{n})$ (the extreme negative ones). Our proof of Theorem 3.1's upper bound says that for any $\delta > 0$, if $\tau(G) \leq p^3 - t$, then $\sum_{i\in N} \lambda_i^3(\tilde{A}_n) \leq -(1-\delta)tn^3$ with high probability, since, essentially, only the extreme eigenvalues contribute for the triangle deviation.

In order to use Lemma 3.23, we will use the inverse order of the indices in N of $|\lambda_i(\tilde{A}_n)|$ so that it is non-increasing. Over the above event $\left(\sum_{i\in N} \lambda_i^3(\tilde{A}_n) \leq -(1-\delta)tn^3 \text{ conditioned on } \tau(G) \leq p^3 - t\right)$, we have that either

$$\lambda_n^3(\tilde{A}_n) \leq -(1-\delta-\varepsilon)tn^3 \text{ or } \sum_{i\in N\setminus\{n\}} \lambda_i^3(\tilde{A}_n) \leq -\varepsilon tn^3.$$

We now compute the probability that, on this event, we have the second possibility happening, i.e. $\lambda_n^3(\tilde{A}_n) > -(1-\delta-\varepsilon)tn^3$ and $\sum_{i\in N\setminus n} \lambda_i^3(\tilde{A}_n) \leq$

$-\varepsilon tn^3$. Hence,

$$\begin{aligned}||\lambda||_2^2 = \sum_{i\in N}\lambda_i^2(\tilde{A}_n) &\geq (1+\varepsilon)^{1/3}\left(\sum_{i\in N}|\lambda_i(\tilde{A}_n)|^3\right)^{2/3}\\ &\geq (1+\varepsilon)^{1/3}(1-\delta)^{2/3}t^{2/3}n^3\\ &= (1+\Omega(\varepsilon))t^{2/3}n^3\end{aligned}$$

since $\sum_{i\in N\setminus\{n\}}\lambda_i^3(\tilde{A}_n) \leq -\varepsilon tn^3$ implies

$$\varepsilon tn^3 \leq |\sum_{i\in N\setminus\{n\}}\lambda_i^3(\tilde{A}_n)| \leq \sum_{i\in N\setminus\{n\}}|\lambda_i^3(\tilde{A}_n)|,$$

and we chose $\delta = \Omega(\varepsilon)$.

Now we use Corollary 3.20 and get

$$\ln\mathbb{P}\left(||\lambda||_2^2 \geq (1+\Omega(\varepsilon))t^{2/3}n^3\right) \leq -(1+\Omega(\varepsilon))(1-o(1))\frac{\ell t^{2/3}n^2}{2}$$

which is much smaller then our lower bound from Theorem 3.1. Therefore, with high probability, in the desired condition,

$$\lambda_n^3(\tilde{A}_n) < -(1-\varepsilon)tn^3 \text{ and } \lambda_{n-1}^3(\tilde{A}_n) \geq -\varepsilon tn^3$$

proving the corollary. ■

The final thing is to show that the bounds work for A_n as well. Here we use that $A_n = \tilde{A}_n + p\mathbf{1} - pI$.

For the $\lambda_{n-1}(A_n)$ bound we use that since $p\mathbf{1}$ is a positive semidefinite matrix, then $\lambda_{n-1}(A_n) \geq \lambda_{n-1}(\tilde{A}_n - pI)$. By Weyl's eigenvalue inequality, we have that

$$\lambda_{n-1}(A_n) \geq \lambda_{n-1}(\tilde{A}_n - pI) \geq \lambda_{n-1}(\tilde{A}_n) - p = -o(tn^3)$$

with high conditional probability.

For $\lambda_n(A_n)$ we need a result about the vertex degree of the graph, which says that the degrees of this $G(n,m)$ don't deviate too much from the mean. Before that, let's just say that since pI is positive semidefinite, then

$$\lambda_n(\tilde{A}_n + p\mathbf{1}) = \lambda_n(\tilde{A}_n + p\mathbf{1} + (pI - pI)) = \lambda_n(A_n + pI) \geq \lambda_n(A_n)$$

and we need to find an upper bound for $\lambda_n(\tilde{A}_n + p\mathbf{1})$.

Corollary 3.25 *Conditioned on $\tau(G) \le p^3 - t$, let $(d_i)_1^n$ be the degrees of the graph. Then, with high probability*

$$\sum_{i=1}^{n}(d_i - pn)^2 = o(tn^3)$$

Proof. Using that the sum of the degrees is $2m = n^2p - O(n)$ we have

$$\begin{aligned}\sum_{i=1}^{n}(d_i - pn)^2 = \sum_{i=1}^{n} d_i^2 - 2pn\sum_{i=1}^{n} d_i - p^2n^3 &= \sum_{i=1}^{n} d_i^2 - 2p^2n^3 + O(n^2) - p^2n^3 \\ &= \sum_{i=1}^{n} d_i^2 - p^2n^3 + O(n^2).\end{aligned}$$

From Lemma 3.2 we get that

$$\begin{aligned}tr[\tilde{A}_n^3] &= tr[A_n^3] - p^3n^3 + p^3n + 6mp(np - 2p + 1) + 3p^3n(n-1) - 3p\sum_i d_i^2 \\ &= tr[A_n^3] - p^3n^3 + p^3n - 3p\left(\sum_{i=1}^{n} d_i^2 - np\sum_{i=1}^{n} d_i + O(n^2)\right) \\ &\le tr[A_n^3] - p^3n^3 - 3p\sum_{i=1}^{n}(d_i - pn)^2 + O(n^2)\end{aligned}$$

since $3p\sum_{i=1}^{n} d_i^2 - 6p^2n\sum_{i=1}^{n} d_i + 3p^3n^3 \ge 3p\sum_{i=1}^{n} d_i^2 - 3pn\sum_{i=1}^{n} d_i + O(n^2)$.

Using that upper bound, it follows that

$$\mathbb{P}(\tau(G) \le p^3 - t) \le \mathbb{P}\left(tr[\tilde{A}_n^3] \le -tn^3 - 3p\sum_{i=1}^{n}(d_i - pn)^2 + O(n^2)\right).$$

Since $tn^3 \gg n^{-3/4}n^3 \gg n^2$, we can compute the probability that $\sum_{i=1}^{n}(d_i - pn)^2 \ge \varepsilon tn^3$, conditioned on $\tau(G) \le p^3 - t$, getting

$$\begin{aligned}\mathbb{P}\left(\tau(G) \le p^3 - t \text{ and } \sum_{i=1}^{n}(d_i - pn)^2 \ge \varepsilon tn^3\right) &\le \mathbb{P}\Big(tr[\tilde{A}_n^3] \le -tn^3(1 + 3p\varepsilon - O(tn^{-1}))\Big) \\ &= \mathbb{P}\Big(tr[\tilde{A}_n^3] \le -tn^3(1 + \Omega(\varepsilon)\Big) \\ &\le \exp\left(-\frac{\ell t^{2/3}n}{2}(1 + \Omega(\varepsilon))\right)\end{aligned}$$

which is much smaller then $\mathbb{P}(\tau(G) \le p^3 - t)$ from our lower bound of Theorem 3.1.

Therefore, conditioned on $\tau(G) \le p^3 - t$, we have $\sum_{i=1}^{n}(d_i - pn)^2 = o(tn^3)$ with high probability. ■

We are almost finished. Let v_n be the unit eigenvector of eigenvalue $\lambda_n(\tilde{A}_n)$ of $\tilde{A}_n$. The above corollary says that, with the condition, we have with high probability that

$$|\tilde{A}_n\bar{\mathbf{1}}|^2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\xi_{ij}\right)^2 = \sum_{i=1}^{n}(d_i - p(n-1))^2 = o(tn^3) + O(n) = o(tn^3).$$

Also, since v_n is an unit eigenvector, we have

$$\begin{aligned}\tilde{A}_n v_n &= \lambda_n(\tilde{A}_n)v_n\\ \langle \tilde{A}_n v_n, \bar{\mathbf{1}}\rangle &= \lambda_n(\tilde{A}_n)\langle v_n, \bar{\mathbf{1}}\rangle\\ \langle \tilde{A}_n \bar{\mathbf{1}}, v_n\rangle &= \lambda_n(\tilde{A}_n)\langle v_n, \bar{\mathbf{1}}\rangle\end{aligned}$$

and $|\lambda_n(\tilde{A}_n)||\langle v_n, \bar{\mathbf{1}}\rangle| \leq |\tilde{A}_n\bar{\mathbf{1}}||v_n| = o(t^{1/2}n^{3/2})$. Also, over our condition, Corollary 3.24 says that with high probability $|\lambda_n(\tilde{A}_n)| > |(1-o(1))tn^3|$. Hence,

$$|\langle v_n, 1\rangle| = o(t^{-1/2}n^{-3/2}) = o(1).$$

Finally,

$$\frac{v_n^T(\tilde{A}_n + p\mathbf{1})v_n}{v_n^T v_n} = \frac{\lambda_n(\tilde{A}_n)v_n^T v_n + p|\langle v_n, 1\rangle|^2}{v_n^T v_n} = \lambda_n(\tilde{A}_n) + o(1)$$

and Rayleigh's criterion gives us

$$\lambda_n(\tilde{A}_n + p\mathbf{1}) \leq \lambda_n(\tilde{A}_n) + o(1) = -(1 - o(1))tn^3$$

with high conditional probability.

The approach is done!

We only make a quick observation that the result is not tight for $p \in (0, 1/2)$ only by a constant factor on the exponential. The proof of this fact can be seen in [18].

Also, one may ask why this approach can't be used on upper tails deviations. The problem is that upper tails probabilities are controlled by perturbations to the largest eigenvalue, as done in Battacharya and Ganguly [1]. However, centering the matrix vanishes λ_1, so our method doesn't help there.

4
Basics for the Martingale Approach

The next two sections are related to the martingale approach to subgraph count deviations [11]. We will not include all of the results of [11]. In particular, we focus on the triangle count, and we prove a weaker bound. We obtain the rate associated with the deviation probability up to a constant, whereas a $(1 + o(1))$ type result is proved in [11]. As we are just giving an overview we do not prove all the results.

In this section we give an introduction to the vital tool: Martingale. It is usually described, playfully, with a casino. You should imagine that you are a gambler and start with X_0 reals/dolars/euros... and that there is a number of fair random games, in the sense that the expected gain in each game is zero, that you can play. Furthermore, this remains true no matter what happened in previous games. Let's formalize it.

Definition 4.1 *An increasing sequence of σ-algebras $(\mathcal{F}_n)$ is called a **filtration**.*

Definition 4.2 *A sequence of random variables (S_n) is said to be **adapted** to a filtration $(\mathcal{F}_n)$ if $S_i \in \mathcal{F}_i$ for all i.*

Definition 4.3 *Let $(\mathcal{F}_n)$ be a filtration. Let (S_n) be a sequence of random variables. If the sequence (S_n) has the following properties*

1. $\mathbb{E}[|S_n|] < \infty$ *for every n;*
2. *(S_n) is adapted to $(\mathcal{F}_n)$;*
3. *and* $\mathbb{E}[S_{n+1}|\mathcal{F}_n] = S_n$,

*then (S_n) is said to be a **martingale** with respect to $(\mathcal{F}_n)$. Its **increments** are defined as $X_i = S_{i+1} - S_i$, for all $i \geq 1$, and note that $\mathbb{E}[X_n|\mathcal{F}_n] = 0$.*

As the Erdős-Rényi random graph process can be described as an increasing sequence of graphs, it is natural to try to use a martigale to study numerical properties of $G(n, m)$'s.

Also, as we are studying deviations, it would be great to have tools to generate bounds of these deviations. There are two very famous inequalities that we are going to use in this overview: Hoeffding-Azuma's Inequality and Freedman's Inequality. We will prove the first (only the positive part), because we did a similar argument in the Spectral approach and it would be nice to know why we called it the "Hoeffding's argument". The second one will be only stated.

Lemma 4.4 (Hoeffding-Azuma Inequality) *Let $(S_m)_0^M$ be a martigale of M steps with increments $(X_i)_1^M$, and let $c_i = ||X_i||_\infty$ for every $1 \leq i \leq M$. Then, for $a > 0$,*

$$\mathbb{P}(S_M - S_0 > a) \leq \exp\left(\frac{-a^2}{2\sum_{i=1}^M c_i^2}\right).$$

In particular, the result is the same for the symmetric statement $S_M - S_0 < -a$.

Proof. Note that $S_M - S_0 = \sum_{i=1}^M S_i - S_{i-1} = \sum_{i=1}^M X_i$. Let $\alpha > 0$. Since α is positive, we can multiply both sides by it and exponentiate to use Markov's inequality and get

$$\begin{aligned}
\mathbb{P}(S_M - S_0 > a) = \mathbb{P}\left(\sum_{i=1}^M X_i > a\right) &= \mathbb{P}\left(\exp\left(\alpha\sum_{i=1}^M X_i\right)\exp\left(-a\alpha\right) > 1\right) \\
&\leq \mathbb{E}\left[\exp\left(\alpha\sum_{i=1}^M X_i\right)\right]\exp(-\alpha a) \\
&= \exp(-\alpha a)\prod_{i=1}^M \mathbb{E}[\exp(\alpha X_i)] \\
&\leq \exp(-\alpha a)\prod_{i=1}^M \cosh(\alpha c_i) \\
&\leq \exp(-\alpha a)\prod_{i=1}^M \exp\left(\frac{\alpha^2 c_i^2}{2}\right) \\
&= \exp\left(\frac{\alpha^2\sum_{i=1}^M c_i^2}{2} - \alpha a\right) \leq \exp\left(\frac{-a^2}{2\sum_{i=1}^M c_i^2}\right)
\end{aligned}$$

where the optimal α is $a/\sum_{i=1}^M c_i^2$. ■

There are a lot of variations for the Hoeffding-Azuma's inequality. The one that we are going to use fit very well with the fact that the increments of the martingale will be attached with the exposition of its corresponding edge. For that, we need to define some ideas.

Let $\mathcal{G}_{n,m}$ be the family of all graphs of order n and size m. We may define that two graphs of this family are adjacent if they differ by only two edges (since they have the same number of edges, they are different in at least two edges) and in this case we say they have **edit distance** 1. In this spirit, we may say that a function $f : \mathcal{G}_{n,m} \to \mathbb{R}$ is C-Lipschitz if $|f(G) - f(G')| \leq C$ for all adjacent pairs $G, G' \in \mathcal{G}_{n,m}$.

Finally, we make this Lipschitz condition particular to each edge.

Definition 4.5 *Let $\psi : E(K_n) \to \mathbb{R}$ be a function. We say that $f : \mathcal{G}_{n,m} \to \mathbb{R}$ is ψ-**Lipschitz** if, for every adjacent pair $G, G' \in \mathcal{G}_{n,m}$, which differs only in e_i and e_j we have*

$$|f(G) - f(G')| \leq \psi(e_i) + \psi(e_j).$$

We are now able to understand the idea of the following corollary

Corollary 4.6 ([11]) *Given $\psi : E(K_n) \to \mathbb{R}^+$ and a ψ-Lipschitz function $f : \mathcal{G}_{n,m} \to \mathbb{R}$, we have*

$$\mathbb{P}(f(G_m) - \mathbb{E}[f(G_m)]| \geq a) \leq \exp\left(\frac{-a^2}{8||\psi||_2^2}\right).$$

In particular, the result is the same for the symmetric statement $f(G_m) - \mathbb{E}[f(G_m)] < -a$.

We now state the Freedman's inequality. The big advantage in using it instead of Hoeffding-Azuma's inequality happens when we have a better control over the increments. If they are usually smaller then their maximum possible value, then Freedman's inequality gives us tighter bounds. Specifically, if $\mathbb{E}[X_i^2|\mathcal{F}_i]$ is smaller than $||X_i||_\infty^2$.

Theorem 4.7 (Freedman's Inequality [8]) *Let $(S_m)_0^M$ be a martigale of M steps with increments $(X_i)_1^M$ with respect to a filtration $(\mathcal{F}_m)_0^M$, let $R \in \mathbb{R}$ be such that $\max_i |X_i| \leq R$ almost surely, and let*

$$V(M) := \sum_{i=1}^{m} \mathbb{E}\left[|X_i|^2|\mathcal{F}_{i-1}\right].$$

Then, for every $\alpha, \beta > 0$, we have

$$\mathbb{P}(S_m - S_0 \geq \alpha \textit{ and } V(m) \leq \beta \textit{ for some } m) \leq \exp\left(\frac{-\alpha^2}{2(\beta + R\alpha)}\right).$$

These are the only results that we need in order to continue.

5
An Overview of the Martingale Approach

Remark: We make here a quick remark in order to understand the parameters used in this approach. Here we will write about the triangle count of a random graph $G(n, m)$ instead of the triangle density. Therefore, the deviation term is $\binom{n}{3}$ times bigger than the one from the previous approach.

We are going to use $\alpha n^{3/2}$ for the deviation term from now on and we note that

$$tn^3 \approx \alpha n^{3/2} \quad \Leftrightarrow \quad t \approx \alpha n^{-3/2}.$$

Then, the result with a Spectral approach is tight for $n^{-3/4} \ll t \ll 1$ and the following approach is going to be tight for $n^{-3/2} \ll t \ll n^{-1}$.

Also, we denote the triangle count deviation by $D_{\triangle}(G_m)$ and $G(n, m)$ by G_m. The main result that we are aiming to get is

Theorem 5.1 *There is a non-negative constant $c = c(H)$ such that for all $m \leq \frac{N}{2}$, and all $\alpha, n \geq c^{-1}$, we have*

$$\mathbb{P}\Big(|D_{\triangle}(G_m)| > \alpha n^{3/2}\Big) \leq \exp\Big(-c\alpha \min\{\alpha, n^{1/2}\}\Big).$$

(For any notational misunderstanding, see the next section 5.1).

In order to reach this, we will do the following: as we are doing a martingale approach, we need a precise martingale expression for $D_{\triangle}(G_m)$. We will then see that the deviation only depends on the deviations of the increments of triangles, $\triangle$, and paths of length two, P_2. Finally, we will manage these deviations by controlling the degrees' behaviour of $G(n, m)$ in order to regulate P_2 deviations, and controlling the codegrees' behaviour in order to regulate $\triangle$ deviations.

5.0.1
About Complements

We won't actually use the following results for our main result proof, but we will quote it on the last section of this chapter when we discuss a generalized version of the main result. We suggest that the reader skip this subsection until then.

The following lemma allow us to handle H-counts using the H-count in the complement of G_m. In order to prove it, you will need to use the inclusion-exclusion principle.

Lemma 5.2 *Let H and G be graphs, then*

$$N_H(G) = \sum_{H' \subset E(H)} (-1)^{e(H')} N_{H'}(G^c) \tag{5-1}$$

where H' is a subgraph of H.

For the corollary, you just need to use the linearity of the expectation and subtract the result from 5-1.

Corollary 5.3 *Let H and G be graphs, then*

$$D_H(G) = \sum_{H' \subset E(H)} (-1)^{e(H')} D_{H'}(e(G^c)).$$

5.1 Notation

We have to settle some notation first, and we have a lot of notation.

Notation 5.4 *Let G and H be graphs.*

- $N_H(G)$ *is the number of isomorphic copies of H in G.*
- $\binom{G}{H}$ *is the number of copies of H in G without repetition.*
- $L_H(m)$ *is the expected number of copies of H in $G(n, m)$.*
- $D_H(G_m) = N_H(G_m) - L_H(m)$ *is the deviation of the H-count in $G(n, m)$.*
- *The model used in this chapter is as follows: let $\{G_{n,m} : m = 0, \ldots, N\}$, $n \geq 1$ be a set of independent copies of the Erdős-Rényi random graph process, and let $(G_{n,r})_{n \geq 1}$ be the sequence of random graphs $(G_{n,m_n})_{n \geq 1}$, where $m_n = \lfloor rN \rfloor$. Sometimes we will write G_m instead of $G_{n,r}$ as in the last item.*

There will be a larger number of needed notation, but they will be done during the others sections when needed.

5.2 The Martingale Expression and Approximation

Let's start by finding the martingale expression for $D_H(G_m)$, where H is any fixed graph. For such, we need to define a sequence of random variables that gives us the deviation's increments on each edge addition.

Fixing n and letting $(G_m : m = 0, \dots, N)$ be a realization of the Erdős-Rényi random graph process, we define

$$A_H(G_m) := N_H(G_m) - N_H(G_{m-1}) \tag{5-2}$$

to be the number of newborn isomorphic copies of H with the m-th's edge addition. As we need information about the increments, we use the centralized form

$$X_H(G_m) := A_H(G_m) - \mathbb{E}\Big[A_H(G_m)|G_{m-1}\Big] \tag{5-3}$$

which gives us the deviation increment of the m-th step.

With these random variables defined, we are ready to state the martingale for the H-count deviation.

Theorem 5.5 *Let H be a graph. Then*

$$D_H(G_m) = \sum_{i=1}^{m} \sum_{F \subset E(H)} \frac{(N-m)_{e(F)}(m-i)_{e(H)-e(F)}}{(N-i)_{e(H)}} X_F(G_i) \tag{5-4}$$

where the sum is over all $2^{e(H)}$ subgraphs F of H.

The first thing that we need to prove is that it is a martingale. But $X_H(G_i)$ is a martingale increment (with respect to the filtration $G_0, \dots, G_N$) and the expression for our increment is a linear combination of these martingale increments, and so, it is a martingale as well.

The second thing is that the equality really holds. It follows from the following lemma and some induction.

Lemma 5.6 *For the Erdős-Rényi random process $(G_m : m = 0, \dots, N)$,*

$$\mathbb{E}\Big[A_H(G_m)|G_{m-1}\Big] = \frac{1}{N-m+1} \sum_{f \in E(H)} \Big(N_{H\setminus f}(G_{m-1}\Big) - N_H(G_{m-1})\Big)$$
$$= \Big(L_H(m) - L_H(m-1)\Big) + \frac{1}{N-m+1} \sum_{f \in E(H)} \Big(D_{H\setminus f}(G_{m-1}) - D_H(G_{m-1})\Big).$$

Proof. In order to prove the second equality, you need two things: you need to note that $N_H(G_n) = L_H(m) + D_H(G_m)$ and

$$L_H(m) - L_H(m-1) = \frac{1}{N-m+1} \sum_{f \in E(H)} \Big(L_{H\setminus f}(m-1) - L_H(m-1)\Big). \tag{5-5}$$

You can see that 5-5 is true if you think that each edge has probability $\frac{1}{N-m+1}$ of being chosen and that the addition of f to $H \setminus f$ completes H.

For the first equality, we only need to prove that the equality holds for the number of embeddings of H created if an specific edge e_m is the image of $f \in E(H)$, that is

$$A_H(G_m) = \sum_{f \in E(H)} A_{H,f}(G_m).$$

where $A_{H,f}(G_m)$ is the number of embeddings created if $e_m = f$.

We now only need to prove that, for each $f \in H$,

$$\mathbb{E}\Big[A_{H,f}(G_m)|G_{m-1}\Big] = \frac{\Big(N_{H\setminus f}(G_{m-1}) - N_H(G_{m-1})\Big)}{N-m+1}.$$

We now do a counting argument. Let $f \in H$ be fixed. A injective map $\phi : V(H) \to V(G)$ is an embedding of H in G_m and not in G_{m-1} if, and only if, ϕ embbeds $H \setminus f$ in G_{m-1}, $\phi(f)$ isn't an edge of G_{m-1} and $e_m = f$.

The set of injective maps that satisfies the first two conditions is of size $N_{H\setminus f}(G_{m-1}) - N_H(G_{m-1})$ and the probability of the last condition happening is $\frac{1}{N-m+1}$. Then the result holds.

■

Here is the proof of Theorem 5.5.

Proof. We will do an induction over the number of edges of H and on $m \in \{0, \dots, N\}$. The base case is $e(H) = 1$ and $m = 0$, which is trivial. For the general case, you use that

$$N_H(G_m) = A_H(G_m) + N_H(G_{m-1})$$

in order to use the induction hypothesis.

Then

$$\begin{aligned} D_H(G_m) &= N_H(G_m) - L_H(m) \\ &= N_H(G_{m-1}) + A_H(G_m) - L_H(m) \\ &= N_H(G_{m-1}) + A_H(G_m) - L_H(m) \pm \Big(L_H(m-1) + \mathbb{E}[A_H(G_m)|G_{m-1}]\Big) \\ &= D_H(G_{m-1}) + X_H(G_m-) + \mathbb{E}[A_H(G_m)|G_{m-1}] - \Big(L_H(m) - L_H(m-1)\Big) \end{aligned}$$

where we only used the definitions of the random variables, and $\pm$ here means "add and subtract this term" (we are summing zero and rearrange terms on the last line of the equation). Now you only need to apply the induction hypothesis and Lemma 5.6. ■

Since we are only interested in triangles, we can compute the martingale expression for $D_\triangle(G_m)$ getting

$$D_\triangle(G_m) = \sum_{i=1}^{m} \left(3\frac{(N-m)_2(m-i)}{(N-i)_3} X_{P_2}(G_i) + \frac{(N-m)_3}{(N-i)_3} X_\triangle(G_i) \right)$$

where P_2 denotes the path of length two. So, in order to study the deviation of of triangles, we need to study the step deviations of both triangles and paths of length two.

To simplify the notation, define

$$\mathbb{X}_\triangle(G_i) := 3\frac{(N-m)_2(m-i)}{(N-i)_3} X_{P_2}(G_i) + \frac{(N-m)_3}{(N-i)_3} X_\triangle(G_i)$$

and then

$$D_\triangle(G_m) = \sum_{i=1}^{m} \mathbb{X}_\triangle(G_i).$$

5.3
Controlling the Degrees and Codegrees Deviations

As we saw on last section, we need to control $X_\triangle(G_i)$ and $X_{P_2}(G_i)$. The way that we found to do it is by controlling the degree's and codegree's deviations, The degrees are going to manage $X_{P_2}(G_i)$ and the codegrees will manage $X_\triangle(G_i)$.

Here we control the deviation of the degrees and codegrees in one way: the sum of squares of deviations. This proof only requires Corollary 4.6 in order to be completed. We prove the one related to degrees and, since the proof is analogous, we only state the one for codegrees.

5.3.1
Degrees' Deviation

Our notation here is $d_u(G)$ for the degree of a vertex u in a graph G. Note that for $G_m \sim G(n,m)$, the expected value of $d_u(G_m)$ is $\frac{2m}{n}$ and, therefore, the deviation of u's degree in G_m is

$$D_u(G_m) := d_u(G_m) - \frac{2m}{2}.$$

We now prove the lemma for degrees.

Lemma 5.7 *There is a constant C such that for all $b \geq 30$, and all $m \leq N$, we have*

$$\mathbb{P}\left(\sum_{u \in V(G_m)} D_u(G_m)^2 > Cbn^2 \right) \leq \exp(-bn).$$

Proof. Let $l = \lfloor \log_2 n \rfloor$ and define a function f_σ for each string $\sigma \in \{0, \pm 1, \pm 2, \ldots, \pm 2^l\}^{V(G_m)}$ as

$$f_\sigma(G_m) := \sum_{v \in V(G_m)} \sigma(v) D_v(G_m).$$

Let σ^* be a string such that it is equal to zero if $D_v(G_m) \leq n^{1/2}$, and otherwise $\sigma^*(v) D_v(G_m) \geq 0$ and $|\sigma(v)|$ is the largest power of two such that $|\sigma^*| n^{1/2} \leq D_v(G_m)$. Then,

$$f_{\sigma^*}(G_m) \geq ||\sigma^*||^2 n^{1/2}.$$

Here is the connection of this and the lemma: if $\sum\limits_{u \in V(G_m)} D_u(G_m)^2 > Cbn^2$ with $C \geq 129$, then

$$||\sigma^*||^2 = \sum_{u \in V(G_m)} \sigma^*(v)^2 \geq \sum_{u \in V(G_m)} \frac{D_u(G_m)^2 - n}{4n} \geq 32bn.$$

Therefore, you only need to show that

$$\sum_{\sigma : ||\sigma|| \geq 32bn} \mathbb{P}\Big(f_\sigma(G_m) > ||\sigma||^2 n^{1/2} \Big) \leq \exp(-bn).$$

The first thing is to bound the probability for each individual σ using Corollary 4.6 using the fact that f_σ is ψ-Lipschitz if $\psi(uv) = |\sigma(u)| + |\sigma(v)|$. Therefore,

$$\sum_{e \in E(K_n)} \psi(e)^2 \leq 2n \sum_{v \in V(G_m)} \sigma_v^2 = 2n ||\sigma||^2$$

and Corollary 4.6 grants, using the fact that $\mathbb{E}[f_\sigma(G_m)] = 0$,

$$\mathbb{P}\Big(f_\sigma(G_m) > ||\sigma||^2 n^{1/2} \Big) \leq \exp\left(\frac{-||\sigma||^2}{16} \right) \leq \exp(-bn) \exp\left(\frac{-||\sigma||^2}{32} \right)$$

since $||\sigma^*||^2 > 32bn$.

With that bound we can translate the problem into proving that

$$\sum_{\sigma : ||\sigma|| \geq 32bn} \exp\left(\frac{-||\sigma||^2}{32} \right) \leq 1.$$

Now you need to split the strings into *types*. We say that σ has *type* $x = (x_{-l-1}, \ldots, x_{l+1})$ if x_0 vertices have $\sigma(u) = 0$, x_i vertices have $\sigma(u) = 2^{i-1}$ and x_{-i} vertices have $\sigma(u) = -2^{i-1}$ for each i. If we set $S_x := \{\sigma :$

σ has type x and $||\sigma||^2 > 32bn\}$, then the new translation is

$$\sum_{\sigma \in S_x} \exp\left(\frac{-||\sigma||^2}{32}\right) \leq \exp(-n)$$

for each x.

As a final translation, note that all strings σ with the same type have the same $||\sigma||^2$ value, which can be written as

$$\varphi(x) := \sum_{j \neq 0} x_j 4^{|j|+1}.$$

Therefore, S_x is empty if $\varphi(x) \leq 32bn$. Fixing a type x such that $\varphi(x) \geq 32bn$, our final translation is proving that

$$|S_x| \leq \exp\left(\frac{\varphi(x)}{32} - n\right).$$

Finally, you only need to compute things now and the proof is done. ■

5.3.2 Codegrees' Deviation

As above, the notation here is $d_{u,w}(G)$ for the codegree of a pair of vertices u and w in a graph G. Note that for $G_m \sim G(n,m)$, the expected value of $d_{u,w}(G_m)$ is $\frac{(n-2)(m_2)}{(N)_2}$ and, therefore, the deviation of u, w codegree in G_m is

$$D_{u,w}(G_m) := d_{u,w}(G_m) - \frac{(n-2)(m_2)}{(N)_2}.$$

The proof for the codegrees lemma is very similar to the corresponding on for degrees and will be omitted. If the reader is curious, the proof can be read in [11] Lemma 4.6.

Lemma 5.8 (Lemma) *There is a constant C such that for all $b \geq 30$, and all $m \leq N$, we have*

$$\mathbb{P}\left(\sum_{u,w} D_{u,w}(G_m)^2 > Cbn^2\right) \leq \exp(-bn).$$

5.4 A Short Proof of Theorem 5.1

We are already only one step from being ready to handle the main result of this chapter. In order to prove Theorem 5.1 we need one more lemma! This lemma shows that we are able to control the squares of $X_{P_2}(G_i)$ and $X_\triangle(G_i)$ by controlling the sum of squares of degree deviations and the sum of squares of codegree deviations.

Lemma 5.9 *There is a constant C such that for all $1 \leq i \leq N/2$, and all $\eta \geq 1$ there is probability at least $1 - \exp(-\eta n^{1/2})$ that*

$$\mathbb{E}\Big[X_{P_2}(G_i)^2 | G_{i-1}\Big] \ \leq \ Cn^{1/2} \max\{\eta, n^{1/2}\}$$

and

$$\mathbb{E}\Big[X_\triangle(G_i)^2 | G_{i-1}\Big] \ \leq \ Cn^{1/2} \max\{\eta, n^{1/2}\}$$

Proof. Let C' be twice the constant obtained by Lemmas 5.7 and 5.8. Define the events E_1

$$\sum_{u \in V(G_m)} D_u(G_m)^2 > C' n^{3/2} \max\{\eta, n^{1/2}\}$$

and E_2

$$\sum_{u,w} D_{u,w}(G_m)^2 > C' n^{5/2} \max\{\eta, n^{1/2}\}.$$

As either one of $\eta \geq n^{1/2}$ and $\eta < n^{1/2}$ can happen, we divide the event E_1 for each case and

$$\mathbb{P}(E_1) \leq \exp(-2n)\mathbb{I}_{\eta \leq n^{1/2}} + \exp(-2\eta n^{1/2})\mathbb{I}_{\eta > n^{1/2}} \leq 2\exp(-2\eta n^{1/2})$$

where we used Lemma 5.7 to bound the probabilities. The same argument, using Lemma 5.8, works for E_2 and a union bound gives

$$\mathbb{P}\Big(E_1 \cup E_2\Big) \leq 4\exp(-2\eta n^{1/2}) \leq \exp(-\eta n^{1/2}).$$

Therefore, we only need to show that the first event of the theorem fails inside event E_1 and the second event fails inside E_2.

Suppose that E_1^c happens. Then,

$$\sum_{u \in V(G_m)} D_u(G_m)^2 \leq C' n^{3/2} \max\{\eta, n^{1/2}\}.$$

Remembering the definition of $X_{P_2}(G_i) = A_{P_2}(G_i) - \mathbb{E}\Big[A_{P_2}(G_i) | G_{i-1}\Big]$, we get that

$$\begin{aligned}\mathbb{E}\Big[X_{P_2}(G_i)^2 | G_{i-1}\Big] &= Var\Big(A_{P_2}(G_i) \Big| G_{i-1}\Big) \\ &\leq \mathbb{E}\left[\left(\frac{8(i-1)}{n} - A_{P_2}(G_i)\right)^2 \middle| G_{i-1}\right].\end{aligned}$$

Since we know that

$$\begin{aligned} A_{P_2}(G_i) &= 2\Big(d_u(G_{i-1}) + d_w(G_{i-1})\Big) \\ &= \frac{8(i-1)}{n} + 2\Big(D_u(G_{i-1}) + D_w(G_{i-1})\Big), \end{aligned}$$

the term inside the expectation becomes $4\Big(D_u(G_{i-1}) + D_w(G_{i-1})\Big)^2$, with u, w being the vertices of the i-th edge. Hence,

$$\begin{aligned} \mathbb{E}\Big[X_{P_2}(G_i)^2|G_{i-1}\Big] &\leq \frac{1}{N-i+1} \sum_{uw \notin E(G_{i-1})} 4\Big(D_u(G_{i-1}) + D_w(G_{i-1})\Big)^2 \\ &\leq \frac{16(n-1)}{N} \sum_u D_u(G_{i-1})^2 \\ &\leq 32C' n^{1/2} \max\{\eta, n^{1/2}\} \end{aligned}$$

since we are in E_1^c. It follows that E_1^c is inside the first event and, therefore, the first event fails in E_1.

We can make the same argument using the equality

$$A_\triangle(G_i) = \frac{6(n-2)(i-1)_2}{(N)_2} + 6D_{u,w}(G_{i-1})$$

for the second event and get

$$\begin{aligned} \mathbb{E}\Big[X_\triangle(G_i)^2|G_{i-1}\Big] &= Var\Big(A_\triangle(G_i)|G_{i-1}\Big) \\ &\leq \mathbb{E}\left[\left(A_\triangle(G_i) - \frac{6(n-2)(i-1)_2}{(N)_2}\right)^2 \middle| G_{i-1}\right] \\ &= \frac{1}{N-i+1} \sum_{uw \notin E(G_{i-1})} \Big(6D_{u,w}(G_{i-1})\Big)^2 \\ &\leq \frac{36}{N-i+1} \sum_{uw} D_{u,w}(G_{i-1})^2 \\ &\leq 36C' n^{1/2} \max\{\eta, n^{1/2}\} \end{aligned}$$

if we are in E_2^c. As before, it follows that E_2^c is inside the second event and, therefore, the second event fails in E_1, proving the lemma. ■

Finally, the proof of our main result 5.1!.

Proof.[of Theorem 5.1] We will finally use the fact that

$$D_\triangle(G_{n,p}) := \sum_{i=1}^{m} \mathbb{X}_\triangle(G_i, p)$$

is a martingale using Freedman's inequality on it. We, then, have to find our parameters α', β and R.

Remembering that

$$\mathbb{X}_H(G_i) := 3\frac{(N-m)_2(m-i)}{(N-i)_3}X_{P_2}(G_i) + \frac{(N-m)_3}{(N-i)_3}X_\triangle(G_i)$$

we note that the coefficients of $X_{P_2}(G_i)$ and $X_\triangle(G_i)$ are at most 3. It implies that

$$\mathbb{E}\Big[\mathbb{X}_\triangle(G_i)^2|G_{i-1}\Big] \leq 18\mathbb{E}\Big[X_{P_2}(G_i)^2|G_{i-1}\Big] + 18\mathbb{E}\Big[X_\triangle(G_i)^2|G_{i-1}\Big]$$

where we used that $(x+y)^2 \leq 2(x^2+y^2)$.

With Lemma 5.9 in mind, we define the event $E_{var}(i-1)$ to be where

$$\mathbb{E}\Big[\mathbb{X}_\triangle(G_i)^2|G_{i-1}\Big] \geq 36C'n^{1/2}\max\{\alpha, n^{1/2}\}$$

where C' comes from the Lemma. By the same Lemma, we know that

$$\mathbb{P}\Big(E_{var}(i-1)\Big) \leq \exp(-\alpha n^{1/2}).$$

If we sum over all i's, we have the event E_{var}

$$\sum_{i=1}^{m}\mathbb{E}\Big[\mathbb{X}_\triangle(G_i)^2|G_{i-1}\Big] \geq 36C'mn^{1/2}\max\{\alpha, n^{1/2}\}$$

that has probability at most $\exp(-\alpha n^{1/2}/2)$, by a union bound. We have found our $\beta = 36C'mn^{1/2}\max\{\alpha, n^{1/2}\}$.

R can be easily found to be $6n$ observing that both $X_{P_2}(G_i)$ and $X_\triangle(G_i)$ have absolute value at most n and their coefficients are at most 3.

Finally, α' comes from the proposition's statement and is $\alpha n^{3/2}$.

We are now able to apply the Freedman's inequality and get

$$\begin{aligned}
\mathbb{P}\Big(D_\triangle(G_m) > \alpha n^{1/2}\Big) &\leq \exp\left(\frac{-(\alpha')^2}{2(\beta + R\alpha')}\right) \\
&= \exp\left(\frac{-\alpha^2 n^3}{72C'mn^{1/2}\max\{\alpha, n^{1/2}\} + 12\alpha n^{5/2}}\right) \\
&\leq \exp\left(\frac{-c\alpha^2 n^3}{n^{5/2}\max\{\alpha, n^{1/2}\} + \alpha n^{5/2}}\right) \\
&= \exp\left(\frac{-c\alpha^2 n^{1/2}}{\max\{\alpha, n^{1/2}\} + \alpha}\right) \\
&\leq \exp\left(\frac{-c'\alpha^2 n^{1/2}}{\max\{\alpha, n^{1/2}\}}\right) = \exp\Big(-c'\alpha\min\{\alpha, n^{1/2}\}\Big)
\end{aligned}$$

where we used that $m \leq \frac{N}{2}$. The proof is complete. ■

5.4.1 Thightness of the Result

Our application for triangles is the following: let $p \in (0, 1/2)$ be a constant and $\alpha = \alpha(n)$ such that $1 \ll \alpha \ll n^{1/2}$. Then, Theorem 5.1 implies that

$$\mathbb{P}\Big(|D_\triangle(G_{n,r})| > \alpha n^{-3/2}\Big) \leq \exp\Big(-c\alpha^2\Big).$$

The approach of martingale is only tight for the interval where $1 \ll \alpha \ll n^{1/2}$. Doing the translation backwards,the Spectral approach was tight for the interval where $n^{3/4} \ll \alpha \ll n^{3/2}$. The middle interval $n^{1/2} \ll \alpha \ll n^{3/4}$ is still open to be solved.

5.5 Going Further

As we were only looking for the triangle result, we didn't explore the entirety of [11]. Indeed, this paper extends this result for any subgraph and any $m = r\binom{n}{2}$, where $r = r(n) \in (0, 1)$ is not necessarily a constant and is bounded away from 1. Furthermore, we only saw the result that obtain the rate associated with the deviation probability up to a constant, while they manage to prove a stronger $(1 + o(1))$ type result.

Our take showed that looking at $X_\triangle(G_i)$ and $X_{P_2}(G_i)$ was all we needed to do in order to find the deviation of the triangle count. Surprisingly, this is the case for any fixed subgraph H counting. Of course, in order to prove it they need a vaster amount of ideas. Let's take a look at some.

In order to prove our main result for any subgraph H with e edges and v vertices, we need to see the following three approximations to our martingale expression

$$D_H(G_m) = \sum_{i=1}^{m} \sum_{F \subset E(H)} \frac{(N-m)_{e(F)}(m-i)_{e(H)-e(F)}}{(N-i)_{e(H)}} X_F(G_i).$$

The first comes from deviations of P_2's and $\triangle$'s:

$$\Lambda_H(G_{n,r}) := n^{v-3}r^{e-2}\left(\binom{H}{P_2} - 3\binom{H}{\triangle}\right) D_{P_2}(G_{n,r}) + n^{v-3}r^{e-3}\binom{H}{\triangle} D_\triangle(G_{n,r}) \tag{5-6}$$

The second is a continuous version:

$$\Lambda_H^*(G_{n,r}) := \sum_{i=1}^{m} \mathbb{X}_H(G_i, r) \tag{5-7}$$

where $s := i/N$ and $\mathbb{X}_H(G_i; p)$ is

$$\mathbb{X}_H(G_i; r) :=$$
$$n^{v-3}p^{e-3}\left(r\binom{H}{P_2}\frac{(1-r)^2}{(1-s)^2}X_{P_2}(G_i) + \binom{H}{\triangle}\frac{(1-r)^3}{(1-s)^3}(X_\triangle(G_i) - 3sX_{P_2}(G_i))\right).$$

And the third approximation

$$\Lambda_H^{**}(G_{n,r}) := \sum_{i=1}^{m}\sum_{F\subset E(H)}\frac{(1-r)^{e(F)}(r-s)^{e-e(F)}}{(1-s)^e}X_F(G_i). \tag{5-8}$$

The idea here is to show that each of these approximations are close to $D_H(G_m)$ deterministically and one approximation is close to the last one probabilistically (with small enough upper bounds). Given that this is true, we "only" need to apply triangle inequality three times to get the main result. Also, the result is done only for $r \in (0, 1/2)$ and extended to $r \in [1/2, 1)$ using complementary graphs with Lemma 5.2 and Corollary 5.3.

As we said, we need to control the degrees and codegrees to manage the results. Each approximation need a different set of controlling results of them, so lemmas 5.7 and 5.8 aren't enough. We also need to deal with the maximum deviation and the sum of forth power of deviations of degrees and codegrees. They are done using Corollary 4.6 and some probability results on hypergeometric distributions, as degrees and codegrees have hypergeometric distributions with some parameters.

Finally, the $(1+o(1))$ type result need to control the conditional variance $X_H(G_i)$ and the conditional covariance of $X_H(G_i)X_{H'}(G_i)$. [11] shows that they are predictable, as they are generally close to a deterministic function that depends only on n, s, H and H'.

With this arsenal, one may finally prove the stronger result.

6
Some Final Words

To wrap up this text, let's return to the Spectral notation.

Remember that, focusing on the lower tail and p constant, the Spectral approach is of right order for the range of $n^{-3/4} \ll t \ll 1$ with result

$$\mathbb{P}\Big(\tau(G) \le \mathbb{E}[\tau(G)] - t\Big) \;=\; \exp\Big(-\Theta(n^2 t^{2/3})\Big)$$

and that the martingale approach is tight for $n^{-3/2} \ll t \ll n^{-1}$ with result

$$\mathbb{P}\Big(\tau(G) \le \mathbb{E}[\tau(G)] - t\Big) \;=\; \exp\Big(-\Theta(n^3 t^2)\Big).$$

Also, since the standard deviation of the triangle count is of order $n^{3/2}$, the standard deviation of for the triangle density is of order $n^{-3/2}$. Hence, if $t \le \Theta\Big(n^{-3/2}\Big)$, then the central limit theorem for $G(n, m)$ gives a bound for that interval, as done by Janson [13].

With these three results, the remaining open case is the range of $n^{-1} \ll t \ll n^{-3/4}$. Neeman et al. [18] conjectured that the Goldschimidt et at. [11] result would extend for that interval as well. It as a natural conjecture as the exponents $n^2 t^{2/3}$ and $n^3 t^2$ cross over at $t = \Theta\Big(n^{-3/4}\Big)$. While natural, it might be wrong. In their newest version of the paper, Neeman et al. showed that their bound, mutatis mutandis, holds for all odd cycles of length of 5 or higher for $n^{-1} \ll t \ll 1$. So it might be the case that at n^{-1} there is a jump for the lower tail bound.

That's it! It was a really nice and engaging experience studying these topics where I had little to no experience whatsoever. We are currently working on the interval that remains open for the triangle count with a third approach! It is remarkable that such a simple structure as a triangle is being this hard to handle and with vastly different approaches. Beginning with the triangles is a standard choice for building up new methods or results, e.g. Janson's Inequality [14] and Mantel's Theorem [17], but they have been really stubborn for our deviation study.

Thanks for reading!

7
Bibliography

[1] B. Bhattacharya and S. Ganguly, *Upper Tails for Edge Eigenvalues of Random Graphs*, SIAM Journal on Discrete Mathematics **53** (2020), 1069–1083.

[2] N. Biggs, *Algebraic graph theory*, 2 ed., Cambridge University Press, 1993.

[3] Béla Bollobás, *Modern graph theory*, 1 ed., Graduate Texts in Mathematics 184, Springer-Verlag New York, 1998.

[4] S. Chatterjee, *An introduction to large deviations for random graphs*, Amer. Math. Soc. **53** (2016), 617–642.

[5] P Erdös and A Rényi, *On random graphs i*, Publicationes Mathematicae Debrecen **6** (1959), 290–297.

[6] P. Erdös, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. 53 (1947), 292–294.

[7] P. Frankl and R. M. Wilson, *Intersection Theorems with Geometric Consequences*, Combinatorica **1** (1981), 292–294.

[8] D.A. Freedman, *On tail probabilities for martingales*, Ann. Probab. **3** (1975), 100–118.

[9] S. R. Garcia and R. H. Horn, *A second course in linear algebra*, 1 ed., Cambridge Mathematical Textbooks, 2017.

[10] E. N. Gilbert, *Random graphs*, Annals of Mathematical Statistics **30** (1959), no. 4, 1141–1144.

[11] C. Goldschimdt, S. Griffiths, and A. Scott, *Moderate deviations of subgraph counts in the Erdös-Rényi random graphs G(n, m) and G(n, p)*, Trans. Amer. Math. Soc. 373 (2020), 5517–5585.

[12] A. Guionnet and O. Zeitouni, *Concentration of the Spectral Measure for Large Matrices*, Electron. Commun. Probab. **5** (2007), 119–136.

[13] F. Janson, *A functional limit theorem for random graphs with applications to subgraph count statistics*, Random Structures Algorithms **1** (1990), 15–37.

[14] ______, *A functional limit theorem for random graphs with applications to subgraph count statistics*, Random Structures Algorithms **1** (1990), 221–229.

[15] S. Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics*, Mem. Amer. Math. Soc. **111** (1994), no. 534.

[16] O. Kallenberg, *Foundation of modern probability*, 2 ed., Springer, 2002.

[17] W. Mantel, *Problem 28 (Solution by H. Gouwentak, W. Mantel, J. Teixeira de Mattes, F. Schuh and W. A. Wythoff),*, Wiskundige Opgaven **10** (1907), 60–61.

[18] J. Neeman, C. Radin, and L. Sadun, *Moderate deviations in triangle count*, arXiv preprint arXiv:2101.08249v2 (2021).

[19] A. Ruciński, *When are small subgraphs of a random graph normally distributed?*, Probab. Theory Related Fields **53** (1988), 79–94.

[20] D. Spielman, *Spectral Graph Theory*, Lecture Notes (2015).

[21] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*, 1 ed., Cambridge University Press, 2018.