# Random walks, effective resistance and neighbourhood statistics in binomial random graphs

by

**John A. Sylvester**

**Thesis**

Submitted for the degree of

**Doctor of Philosophy**

**Mathematics Institute**

**The University of Warwick**

**September 2017**

# Contents

# Acknowledgments

First and foremost I give thanks to my supervisors Agelos Georgakopoulos and David Croydon for their help, support, guidance and many interesting discussions. Without them this work would not have been possible.

I would also like to thank my new line manager Thomas Sauerwald for his support and patience while I was finishing my thesis.

I also owe many thanks to everyone involved in the MASDOC doctoral training centre, supported by EPSRC grant EP/HO23364/1. I also give thanks to the Warwick Maths and Stats departments at Warwick for providing me with an excellent environment to study.

In addition I thank my friends and colleges at Warwick for being a pleasure to be around.

Finally I wish to thank my family and Jess for their love and support.

# Declarations

Most of the material of Chapters 2, 3, 4 and 5 are from the paper [71]. Chapters 6 and 7 contain more recent work and the results of Chapter 6 are currently being prepared as a paper [70].

Whenever I state a result which is not my own I have cited it, this citation typically appears in the title of the Theorem. Known results for which I cannot find an adequate citation are cited as folklore. I have also tried to cite papers which were influential to certain proofs.

I declare that, to the best of my knowledge, the material contained in this thesis is original and my own work, conducted under the supervision of Agelos Georgakopoulos and David Croydon, except where otherwise indicated or cited.

The material in this thesis is submitted to the University of Warwick for the degree of Doctor of Philosophy, and has not been submitted to any other university or for any other degree.

# Abstract

The binomial random graph model $\mathcal{G}(n,p)$, along with its near-twin sibling $\mathcal{G}(n,m)$, were the starting point for the entire study of random graphs and even probabilistic combinatorics as a whole. The key properties of these models are woven into the fabric of the field and their behaviour serves as a benchmark to compare any other model of random structure. In this thesis we contribute to the already rich literature on $\mathcal{G}(n,p)$ in a number of directions.

Firstly, vertex to vertex hitting times of random walks in $\mathcal{G}(n,p)$ are considered via their interpretation as potential differences in an electrical network. In particular we show that in a graph satisfying certain connectivity properties the effective resistance between two vertices is typically determined, up to lower order terms, by the degrees of these vertices. We apply this to obtain the expected values of hitting times and several related indices in $\mathcal{G}(n,p)$, and to prove that these values are concentrated around their mean.

We then study the statistics of the size of the $r$-neighbourhood of a vertex in $\mathcal{G}(n,p)$. We show that the sizes of these neighbourhoods satisfy a central limit theorem. We also bound the probability a vertex in $\mathcal{G}(n,p)$ has an $r$-neighbourhood of size $k$ from above and below by functions of $n, p$ and $k$ which match up to constants.

Finally, in the last chapter the extreme values of the $r$-degree sequence are studied. We prove a novel neighbourhood growth estimate which states that with high probability the size of a vertex's $r$ neighbourhood is determined, up to lower order terms, by the size of its first neighbourhood. We use this growth estimate to bound the number of vertices attaining a smallest $r$-neighbourhood.

# Chapter 1

# Introduction

In this thesis we shall consider various properties and statistics of the binomial random graph model $\mathcal{G}(n,p)$. The binomial random graph is the distribution over $n$-vertex simple labelled graphs generated by including each edge independently with probability $p$. The random graph $\mathcal{G}(n,p)$ has been the subject of intense study over the previous half century and there are several books on the topic [20, 43, 50].

The model is studied for a variety of reasons, one reason is that many graph properties undergo a transition around some threshold value of $p$. One of the most remarkable of these transitions is the "double jump" phase transition for the size of the largest component of the graph, this was first observed and shown by Erdős and Rényi in the landmark paper [38]. The model has served as a test bed for other percolation techniques such as susceptibility [49].

Random walks on graphs have also received a lot of attention throughout the last century [2, 57, 58]. Many graph properties such as expansion are closely related to the hitting and cover times of random walks [23], there is also a close connection to spectral properties [5]. We can view a group as a graph and then amenability can be determined by the rates of escape of a random walk on the group [9]. Many algorithms utilise random walks for example random walks were used to solve the undirected reachability problem in polynomial time and log space [3]. The celebrated Markov chain Monte Carlo algorithms for sampling from a probability distribution can also be seen as random walks on graphs and their success is dependent on the mixing time of this walk [32, 61]. Some applications of random walks on random graphs are surveyed in [30].

Interpreting a graph as an electrical network has yielded many great things, most notable for this thesis is the connection to random walks [24, 33, 72]. Electrical networks and resistance forms play a role in theory of convergence of stochastic processes on discrete metric spaces [53]. Recently the effective resistance has also found a use in graph sparsification [69].

The first part of the thesis is concerned with effective resistance and parameters related to vertex hitting times of random walks on $\mathcal{G}(n,p)$. We calculate the effective resistance $R(i,j)$ between two vertices $i,j$ of $\mathcal{G}(n,p)$ and show that with high probability for a given pair the leading order term for the effective resistance is given by the sum of the reciprocals of the degrees of $i$ and $j$ (Theorem 5.3.1). Exploiting the strong connection between electrical networks and random walks we use our resistance bounds to calculate the expectation of the vertex hitting times, Kirchhoff index and a number of related indices (Theorem 5.1.1). We also show concentration around the mean for these quantities (Theorem 5.2.1). In particular we show that expected time for a random walk to travel between two typical vertices in $\mathcal{G}(n,p)$ is $n \pm o(n)$ w.h.p. (Theorem 5.2.1 and Corollary 2.3.3).

The second half of the thesis is concerned with the distribution of the $r$-neighbourhoods in $\mathcal{G}(n,p)$. As will be discussed in Section 1.2 many problems in random graphs such as determining the diameter of the graph and existence of the giant component can be approached by showing that the neighbourhoods of vertices in the graph grow at a certain rate. This is typically done by comparing the neighbourhoods to generations in a branching process or Galton-Watson tree. It seems natural to study the distribution of the $r$-neighbourhoods sizes and a greater understanding of these may be useful for other problems in $\mathcal{G}(n,p)$. A good way to understand the $r$-neighbourhood distribution when a closed description of the law in terms of known distributions seems elusive is to study the probability that a vertex $u$ has exactly $k$ vertices lying at distance precisely $r$ from it in $\mathcal{G}(n,p)$. We bound this probability from above and below by functions which differ only by a constant factor (Theorem 6.0.1). We also prove a central limit theorem for the size of the $r$-neighbourhood of a vertex in $\mathcal{G}(n,p)$ (Theorem 6.0.2). Indeed, what we show is that the size of the $r^{th}$-neighbourhood of a vertex centred by the mean $(np)^r$ and scaled by standard deviation $(np)^{(2r-1)/2}$ converges in distribution to a standard normal distribution. The order of the mean here is what one would expect from a graph with average degree $np$ and good expansion properties. Why the standard deviation is $(np)^{(2r-1)/2}$ is less clear.

In the last Chapter we prove a growth estimate for the $r$-neighbourhoods (Theorem 7.1.1). This estimate differs from any I have seen in the literature as it states that the $r$-neighbourhood of a vertex is concentrated around $(np)^{r-1}$ times the size of the first neighbourhood. This is used to give a bound on the number of vertices with minimum degree which is smaller than any fixed positive decimal power of $n$, for $\mathcal{G}(n,p)$ above the connectedness threshold (Theorem 7.3.4). In the final section we offer a conjecture regarding the threshold for uniqueness of vertices with a minimum $r$-neighbourhood in $\mathcal{G}(n,p)$ (Conjecture 7.4.1).

We now give a more detailed overview of the results and the structure of the thesis before reviewing the literature relating to the results presented in this thesis.

## 1.1 Overview and results

In Chapter 2 we will introduce binomial random graphs, hitting times and the related random walk indices we study in the first half of the thesis. We shall also formally define what we mean by the effective resistance $R(i,j)$ between two vertices $i,j \in V(G)$, informally this is the potential difference between $i$ and $j$ when one unit of current flows from $i$ to $j$ and each edge has unit resistance. For now we shall try and state the results with as few definitions as possible. The theorems are numbered by the chapter/section where they are proved.

### 1.1.1 Random walks and effective resistance

In Chapter 3 we define the modified-breadth first search algorithm $\mathrm{MBFS}(G,\{i,j\})$. This algorithm takes a graph $G$ and two vertices $i,j$ as input and outputs a sequence of "pruned" neighbourhoods indexed by distance from $\{i,j\}$. We then define a connectivity property for graphs, Definition 3.3.1, which we call the strong $k$-path property. This property roughly says there are many paths of length at most $k$ between the second neighbourhoods of $i$ and $j$. The main theorem of the chapter gives an upper bound for the effective resistance in graphs satisfying the strong $k$-path property. The bound on $R(i,j)$ is in terms of the reciprocals of the pruned neighbourhoods output by the algorithm $\mathrm{MBFS}(G,\{i,j\})$.

In Chapter 4 we show that provided the edge parameter $p$ in the random graph $\mathcal{G}(n,p)$ lies in the range $c\log n \leq np < n^{1/10}$ for some fixed $c > 0$ then the strong $k$-path property is satisfied for a given pair of a vertices $i,j$ with high probability. We also prove a number of results about the distribution of the pruned neighbourhoods output by the MBFS algorithm run on $\mathcal{G}(n,p)$ and provide couplings to the original first and second neighbourhoods of $i,j$ in $V(\mathcal{G})$.

In Chapter 5 we apply the results of Chapter 4 to $\mathcal{G}(n,p)$. Firstly, for $i \in V(G)$ we let $\gamma_r(i)$ denote the number of vertices at distance $r$ from $i$, thus $\gamma_1(i)$ is the degree of $i$. Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ mean $\mathcal{G}$ is distributed according to the law of $\mathcal{G}(n,p)$. By applying the resistance bound from Chapter 4 to $\mathcal{G}(n,p)$ we obtain

**Theorem 5.3.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ and $i,j \in V, i \neq j$.*

*(i) For every $c>0$ if $c\log n \leq np \leq n^{1/10}$, then*

$$\mathbb{P}\left(\left|R(i,j)-\left(\frac{1}{\gamma_1(i)}+\frac{1}{\gamma_1(j)}\right)\right| > \max\left\{\frac{1}{\gamma_1(i)^2}+\frac{1}{\gamma_1(j)^2}, \frac{9(\gamma_1(i)+\gamma_1(j))\log n}{\gamma_1(i)\gamma_1(j)np\log(np)}\right\}\right) \leq 2np^2 + o\left(e^{-np/4}\right).$$

*(ii) For every $c>0$ if $np = c\log n$, then for any fixed $k>0$*

$$\mathbb{P}\left(\left|R(i,j)-\frac{2}{c\log n}\right| > \frac{10}{c^2\log(n)\log\log(n)}\right) \leq \frac{5}{(\log n)^k}.$$

*(iii) If $np=\omega(\log n)$ and $np \leq n^{1/10}$, then*

$$\mathbb{P}\left(\left|R(i,j)-\frac{2}{np}\right| > \frac{7\sqrt{\log n}}{(np)^{3/2}}\right) = o\left(\frac{1}{n^{7/2}}\right).$$

Thus with high probability the main contribution to the effective resistance $R(i,j)$ between two given vertices $i,j\in V$ comes from the flow through edges connecting $i$ and $j$ to their immediate neighbours.

Exploiting the strong connection between random walks and effective resistance, an outline of this connection is given in Sections 2.2 & 2.5, we can apply our bounds on resistance to a number of random walk indices. The first of these are random walk hitting and commute times, denoted $h(i,j)$ and $\kappa(i,j)$ respectively; they are the expected time taken for a random walk from $i\in V$ to first visit $j\in V$, and then also return to $i$ in the case of $\kappa(i,j)$. Another is the Kirchhoff index, $K(\mathcal{G})$, which is the sum of all effective resistances in the graph. The remaining indices are stationary hitting times $H_i(\mathcal{G})$, the mean hitting time $T(\mathcal{G})$, Kemeny's constant $H(\mathcal{G})$, and cover costs $cc_i(\mathcal{G})$, $\overline{cc}(\mathcal{G})$. These are sums of hitting times weighted by combinations of stationary or uniform distributions of vertices. These quantities will all be defined and introduced formally in Chapter 2.

Let $\mathcal{C}:=\mathcal{C}_n$ be the event that $\mathcal{G}\sim_d \mathcal{G}(n,p)$ is connected. Let $a(n),b(n):\mathbb{N}\to\mathbb{R}$, then for ease of presentation we use the notation

$$a(n)\overset{O}{=}b(n) \quad \text{to denote} \quad a(n)=\left(1\pm O\left(\frac{\log n}{np\log(np)}\right)\right)b(n).$$

Theorem 5.1.1 concerns moments of the above graph indices on $\mathcal{G}(n,p)$ conditioned to be connected. This conditioning is to ensure the expectation is bounded.

**Theorem 5.1.1.** *Let $\mathcal{G}\sim_d\mathcal{G}(n,p)$ with $\log n+\log\log\log n \leq np \leq n^{1/10}$. Then for any $i,j\in V(\mathcal{G})$ where $i\neq j$,*

| | | | |
|---|---|---|---|
| *(i)* | $\mathbb{E}\big[R(i,j)\big\vert\mathcal{C}\big] \overset{O}{=} \dfrac{2}{np}$, | $\mathbb{E}\big[h(i,j)\big\vert\mathcal{C}\big] \overset{O}{=} n$, | $\mathbb{E}\big[\kappa(i,j)\big\vert\mathcal{C}\big] \overset{O}{=} 2n$, |
| *(ii)* | $\mathbb{E}\big[K(\mathcal{G})\big\vert\mathcal{C}\big] \overset{O}{=} \dfrac{n}{p}$, | $\mathbb{E}\big[\overline{cc}(\mathcal{G})\,\big\vert\mathcal{C}\big] \overset{O}{=} n$, | $\mathbb{E}\big[cc_i(\mathcal{G})\big\vert\mathcal{C}\big] \overset{O}{=} n$, |
| *(iii)* | $\mathbb{E}\big[K(\mathcal{G})^2\big\vert\mathcal{C}\big] \overset{O}{=} \dfrac{n^2}{p^2}$, | $\mathbb{E}\big[h(i,j)^2\big\vert\mathcal{C}\big] \overset{O}{=} n^2$, | $\mathbb{E}\Big[cc_i(\mathcal{G})^2\,\Big\vert\mathcal{C}\Big] \overset{O}{=} n^2$, |
| *(iv)* | $\mathbb{E}\big[H_i(\mathcal{G})\big\vert\mathcal{C}\big] \overset{O}{=} n$, | $\mathbb{E}\big[H(\mathcal{G})\big\vert\mathcal{C}\big] \overset{O}{=} n$, | $\mathbb{E}\big[T(\mathcal{G})\big\vert\mathcal{C}\big] \overset{O}{=} n$, |
| *(v)* | $\mathbb{E}\big[H_i(\mathcal{G})^2\big\vert\mathcal{C}\big] \overset{O}{=} n^2$, | $\mathbb{E}\big[H(\mathcal{G})^2\big\vert\mathcal{C}\big] \overset{O}{=} n^2$, | $\mathbb{E}\big[T(\mathcal{G})^2\big\vert\mathcal{C}\big] \overset{O}{=} n^2$. |

We also have concentration for these indices resulting from their moments.

**Theorem 5.2.1.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$ *with* $\log n + \log\log\log n \leq np \leq n^{1/10}$, $f(n) : \mathbb{N} \to \mathbb{R}_+$. *Then for* $X \in \{h(i,j),\ \kappa(i,j),\ K(\mathcal{G}),\ H_i(\mathcal{G}),\ H(\mathcal{G}),\ T(\mathcal{G}),\ cc_i(\mathcal{G}),\ \overline{cc}(\mathcal{G})\}$, $i,j \in V$, $i \neq j$,

$$\mathbb{P}\left(\Big|X - \mathbb{E}\big[X\big\vert\mathcal{C}\big]\Big| > \mathbb{E}\big[X\big\vert\mathcal{C}\big]\sqrt{\frac{f(n)\log n}{np\log(np)}}\right) = O\left(\frac{1}{f(n)}\right) + \mathbb{P}(\mathcal{C}^c)\,.$$

Notice that $\mathbb{P}(\mathcal{C}^c) \leq e^{\log(n)-np} \leq 1/\log\log(n)$ by Theorem 2.3.1, so in particular by choosing $f(n) = \log\log(np)$ above we see that these random variables concentrate in a sub-mean interval with high probability. Theorems 5.1.1 and 5.2.1 are valid only for $np \leq n^{1/10}$, however concentration for all of the aforementioned random variables has been determined for $np$ above this range by spectral methods or otherwise. The original contribution of this thesis is determining expectation and concentration close to the connectedness threshold $np = \log n$ where it is hard to obtain good estimates on the relevant spectral statistics of $\mathcal{G}(n,p)$. In particular the results of this thesis extend or complement some or all of the results in the papers [22, 51, 59, 73], this will be outlined in Section 1.2. It is noteworthy that if $np = \omega(\log(n))$ then tighter concentration can be obtained for the above quantities by Theorem 5.3.1 (iii). For example the next corollary, proved in section 2, follows almost directly from Tetali's formula (2.14) and Theorem 5.3.1 (iii).

**Corollary 2.3.3.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$ *where* $np = \omega(\log n)$. *Then*

$$\mathbb{P}\left(\sup_{\{i,j\}\subseteq V} |h(i,j) - n| > 11n\sqrt{\frac{\log n}{np}}\right) = o\left(\frac{1}{n^{3/2}}\right).$$

Notice that Theorem 5.2.1 holds for one pair $i,j \in V$ and that the exceptional probability is too large for a union over all pairs. This seems disappointing however if $np = \Theta(\log(n))$ then the statement of Theorem (5.2.1) does not hold for all pairs of vertices as shown by the following proposition.

**Proposition 3.1.4.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$. If $np = \log(n) + 100 \log\log\log(n)$, then*

$$\mathbb{P}(\textit{there exists } i, j \in V : h(i,j) > n\log(n)/2) = 1 - o(1).$$

*For any $1 < c < \infty$ if $np = c\log(n)\,(1 \pm o(1))$ then there is an $a > 0$ such that*

$$\mathbb{P}(\textit{there exists } i, j \in V : h(i,j) > (1+a)n) = 1 - o(1).$$

Bollobás & Thomason [21, Theorem 1] showed that the threshold for having minimum degree $k(n)$ coincides with the threshold for having at least $k(n)$ vertex-disjoint paths between any two points. Let $paths_2(i,j,l)$ be the maximum number of paths of length at most $l$ between vertices $i$ and $j$ of $\mathcal{G}$ that are vertex disjoint on $V \setminus (B_1(i) \cup B_1(j))$. The strong $k$-path property can be used to prove a related "local first neighbourhood relaxation" of this statement for two vertices.

**Theorem 5.4.2.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where for any $c > 0$, $c\log n \le np \le n^{1/10}$. Let $l := \log n / \log(np) + 9$. Then for $i, j \in V$ where $i \neq j$,*

*(i)* $\mathbb{P}(paths_2(i,j,l) \neq \min\{\gamma_2(i), \gamma_2(j)\}) \le 5n^3p^4 + o\left(e^{-7\min\{np,\log n\}/2}\right)$,

*(ii)* $\mathbb{P}\left(\left|paths_2(i,j,l) - (np)^2\right| > 3(np)^{3/2}\sqrt{\log np}\right) = o\left(1/np\right)$.

### 1.1.2 The statistics of $r$-neighbourhoods in $\mathcal{G}(n,p)$

In Chapter 6 we consider the distribution of the size of the $r$-neighbourhood of a vertex $u$ in the binomial random graph $\mathcal{G}(n,p)$. It is fairly straightforward to see that conditional on the sizes of all the preceding neighbourhoods $\{\gamma_1(u)\}_{i=0}^{r-1}$ we have that $\gamma_r(u)$ is distributed according to $Bin\left(n - \sum_{i=0}^{r-1}\gamma_i(u), 1 - (1-p)^{\gamma_{r-1}}\right)$. We are interested in $\mathbb{P}(\gamma_r(u) = k)$, this is the probability that a vertex attains a neighbourhood size $k$. It is not so clear how this probability should depend on $n, p$ and $k$ from the conditional distribution, we show the following.

**Theorem 6.0.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, where $np \to \infty$, and $u \in V$. Let $r := r(n)$, $r \ge 1$ and $k := k(n)$ be such that $(np)^{2r} = o(n)$ and $k = \Theta((np)^r)$. Let $\alpha = k/(np)^r$. Then there exists $C := C(\alpha) < \infty$ such that*

$$\mathbb{P}(\gamma_r(u) = k) \le C \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{\sqrt{(np)^{2r-1}}}.$$

*If in addition* $\alpha > 1/2\pi$ *then there exists* $c := c(\alpha) > 0$ *such that*

$$\mathbb{P}(\gamma_r(u) = k) \geq c \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{\sqrt{(np)^{2r-1}}}.$$

This theorem applies to any $r(n)$ up to $\approx \log(n)/2\log(np)$ which is close to half the diameter of $\mathcal{G}(n,p)$ [25]. Theorem 6.0.1 is almost a local limit theorem for $\gamma_r(u)$ however, we only know the limit up to a constant. We also prove the following central limit theorem for $\gamma_r(u)$.

**Theorem 6.0.2** (Central Limit Theorem for $\gamma_r(u)$)**.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$*, where* $np \to \infty$*. Let* $r := r(n)$ *be such that* $(np)^{r+1/2} = o(n)$ *and let* $u \in V$*. Then*

$$\left(\frac{\gamma_r(u) - (np)^r}{(np)^{(2r-1)/2}}\right) \xrightarrow{d} \mathcal{N}(0,1).$$

The convergence above is in distribution to a normal random variable with mean zero and variance one.

The final chapter, Chapter 7, is somewhat more speculative. We consider one extreme of the $r$-degree sequence, namely $\delta_r$ which is the smallest $r$-neighbourhood of any vertex in $\mathcal{G}(n,p)$. We prove a technical theorem which shows that the $r$-neighbourhood of vertex in $\mathcal{G}(n,p)$ is essentially determined by its first neighbourhood.

**Theorem 7.1.1.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$*,* $u \in V$ *and* $r := r(n) \geq 1$ *be such that* $(np)^{r+1} = o(n)$*. Let* $\lambda^* := \sqrt{\min\{10\gamma_1(u)\log(np), 2\log(n)\}}$ *and define the event*

$$\mathcal{E}_{u,r} := \bigcap_i^r \left\{ \left|\gamma_i(u) - \gamma_1(u)(np)^{i-1}\right| \leq \lambda^*(np)^{i-1}\sqrt{\frac{\gamma_1(u)}{np}} \right\}.$$

*For any* $c > 0$ *if* $np \geq c\log n$ *then* $\mathbb{P}((\mathcal{E}_{u,r})^c) = o\left(\frac{1}{n}\right) + o(e^{-np})$*.*

This estimate is applied to prove a novel upper bound on the number of vertices attaining an $r$-neighbourhood of minimum size.

**Theorem 7.3.4.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$ *where* $\liminf_{n\to\infty} np - \log(n) > -\infty$*. Let* $r := r(n) \geq 2$ *be such that* $(np)^{r+1} = o(n)$*. Then with high probability the number of vertices attaining an* $r$*-neighbourhood of minimum size is bounded above by* $e^{O\left(\sqrt{\log(n)}\right)}$*.*

I believe that this bound is far from optimal and in the last section I motivate Conjecture 7.4.1 which states that above the connectedness threshold there is a unique vertex with smallest $r$-neighbourhood for $r \geq 2$.

## 1.2 Literature and background

As noted above many results in the literature on random walk indices arise from connections with spectral theory. To discuss these results we must first clarify some definitions. Let $A$ be the adjacency matrix of a graph $G$ and $D$ be the diagonal matrix with $D_{i,j} = \gamma_1(i)$ if $i = j$ and $D_{i,j} = 0$ otherwise. The combinatorial Laplacian $L$ is defined as $L := D - A$. Let $L^\dagger(G)$ denote the Moore-Penrose pseudoinverse of $L(G)$. This is a generalisation of the inverse of a matrix, see [66] for more details.

Boumal & Cheng [22] exploit an expression for the Kirchhoff index $K(G)$ in terms of the trace of $L^\dagger(G)$ to obtain expectation and concentration for $K(\mathcal{G})$ on $\mathcal{G}(n,p)$ with $np = \omega\left((\log n)^6\right)$. We will now outline a related expression for $K(G)$ and explain how this can also be used with spectral statistics to control $K(G)$. Let $\lambda_i$ be the eigenvalues of $L(G)$, where $G$ is a finite connected graph. Then by the matrix tree theorem we have the following expression for the Kirchhoff index [45]

$$K(G) = \sum_{\lambda_i \neq 0} \frac{1}{\lambda_i}. \tag{1.1}$$

A theorem of Coja-Oghlan, [27, Theorem 1.3], states that if $\mathcal{G} \sim_d \mathcal{G}(n,p)$ with $np \geq C_0 \log n$ for sufficiently large $C_0$ then the non-zero eigenvalues of $L(\mathcal{G})$ concentrate around the means. Combining these estimates with (1.1) yields concentration for $K(\mathcal{G})$ and with extra work the leading order term of $\mathbb{E}\left[K(\mathcal{G})\middle|\mathcal{C}\right]$ can be determined when $np \geq C_0 \log n$. It is of note however that Boumal & Cheng obtain second order terms for $\mathbb{E}\left[K(\mathcal{G})\middle|\mathcal{C}\right]$, which is not possible with the latter method. Theorems 5.1.1 and 5.2.1 extend the range of known results giving expectation and concentration for $K(\mathcal{G})$ when $np \geq \log n + \log\log\log n$.

Löwe & Torres [59] obtain concentration results for $H(\mathcal{G}), H_i(\mathcal{G}),\ \kappa(i,j)$ on $\mathcal{G}(n,p)$, defined as Kemeny's constant, stationary hitting time and commute time respectively. Again, the result comes from using expressions for these quantities in terms of the eigenvectors and eigenvalues of the transition matrix of the simple random walk, these expressions can be found in [58]. Löwe & Torres then apply results from Erdős et. al. [35, 36] to bound from above the reciprocal of the spectral gap. Löwe & Torres require $np = \omega\left((\log n)^{C_0}\right)$ for some $C_0 > 0$ sufficiently large as this is needed to apply the results in [35, 36]. Theorems 5.1.1 and 5.2.1 extend these results to the range $np \geq \log n + \log\log\log n$.

Von Luxburg, Radl & Hein [73, Theorem 5] prove bounds on the difference of $h(i,j)/2|E|$ and $\kappa(i,j)/2|E|$ from $1/\gamma_1(i) + 1/\gamma_1(j)$ and $1/\gamma_1(i)$ respectively for non bipartite graphs by the reciprocal of the spectral gap and the minimum degree of $G$. They then apply these to various geometric random graphs. The issue with applying these bounds to Erdős-Rényi graphs is that we have to bound from above the reciprocal of the spectral gap so a lower bound on the spectral gap is required. This appears to be a very hard problem and to the author's knowledge the state of the art in eigenvalue separation for $\mathcal{G}(n,p)$ are the papers [35, 36]. So, as is the case with the Löwe & Torres result, if we wish to apply these to get concentration for $h(i,j)$, $\kappa(i,j)$ in $\mathcal{G}(n,p)$, then we have to make the assumption $np = \omega\left((\log n)^{C_0}\right)$ for some $C_0$ sufficiently large. Theorem 5.2.1 however provides concentration results for $h(i,j)$ and $\kappa(i,j)$ when $\log n + \log\log\log n \leq np \leq n^{1/10}$.

In [51] Jonasson studies the cover time, the expected time to visit all vertices from the worst start vertex, for $\mathcal{G}(n,p)$. He bounds the cover time by showing effective resistances and hitting times on $\mathcal{G}(n,p)$ concentrate in the regimes where $\omega(\log n) = np \leq n^{1/3}$. Jonasson does not use spectral methods and instead achieves an upper bound on the effective resistance by finding a suitable flow. This is the approach we have also taken, however we use a refined analysis and extend Jonasson's results for hitting times to the case where $np \geq \log n + \log\log\log n$ and for effective resistance to the case $np \geq c\log n$, $c > 0$.

It is worth noting that the cover time has since been determined for all connected $\mathcal{G}(n,p)$ by Cooper & Frieze [28] using mixing time estimates. One cannot deduce much about the individual hitting times $h(i,j)$ from this result. The question we address is: "what does a typical hitting time look like?". The mixing time of $\mathcal{G}(n,p)$ has also received much attention [13, 14, 42, 63].

A neighbourhood growth estimate is the statement: *with probability $X$ the random variable $\gamma_r(u)$ lies in the interval $Y$.* We use these frequently, an example of a standard growth estimate is Lemma 4.1.2, whereas Theorem 7.1.1 is a more specialised. One problem ffor which neighbourhood growth estimates are invaluable is estimating the diameter. One way to bound the diameter is to let the neighbourhoods branch from two distinct vertices until they are large enough that they share a common vertex or there is an edge between them with high probability. It turns out that this is quite an effective way to estimate the diameter in $\mathcal{G}(n,p)$ see [16, 25, 67]. We prove the strong $k$-path property in Chapter 3 using a similar idea.

Another set of problems where the distribution of the $r$-neighbourhoods of $\mathcal{G}(n,p)$ have been considered is in the reconstruction and assembly problems [17, 62]. They ask whether given certain local information, such as some neighbourhood statistics, about a (random) graph it is possible to determine the graph in question. These problems are related to the graph isomorphism problem [8]. To my knowledge a central or local limit theorem for the $r$-neighbourhoods has never been proven.

# Chapter 2

# Preliminaries

We must now clarify some notation. Throughout, unless otherwise stated, we shall take $G = (V, E)$ to be an $n$-vertex simple graph. For a graph $G$ let $d(i, j)$ be the graph distance between $i, j \in V$ and define the following

$$\Gamma_k(i) := \{j \in V : d(i, j) = k\}, \qquad \gamma_k(i) := |\Gamma_k(i)|, \qquad B_k(i) := \bigcup_{h=0}^{k} \Gamma_h(i), \quad (2.1)$$

which are the $k^{th}$ neighbourhood of $i$, size of $k^{th}$ neighbourhood and the ball of radius $k$ centred at $i$ respectively. Let $\mathbb{R}_+$ and $\mathbb{Z}_+$ denote the strictly positive real numbers and integers respectively. We use the following asymptotic notation. Let,

- $f(n) = O(g(n))$ if there exists $N \in \mathbb{N}, K \in \mathbb{R}_+$ such that for all $n > N, |f(n)| \leq Kg(n)$,
- $f(n) = o(g(n))$ if for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\forall n > N, |f(n)| < \varepsilon g(n)$,
- $f(n) = \Omega(g(n))$ if there exists $N \in \mathbb{N}, K \in \mathbb{R}_+$ such that for all $n > N, f(n) \geq Kg(n)$,
- $f(n) = \omega(g(n))$ if for any $K \in \mathbb{R}_+$ there exists $N \in \mathbb{N}$ such that for all $n \geq N, f(n) \geq K|g(n)|$,
- $f(n) = \Theta(g(n))$ if $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$,
- $f(n) \sim g(n)$ if for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n > N, |f(n)/g(n) - 1| < \varepsilon$.

Throughout $\log(\cdot)$ denotes the natural logarithm base $e$. We will make frequent use of the following inequalities.

Hölder's inequality: For $k = 1, \ldots, n$ let $X_k$ be random variables, $p_k \in [1, \infty)$ where $\sum_{k=1}^{n} 1/p_k = 1$ and $\mathbb{E}\left[X_k^{p_k}\right]$ exists, then

$$\mathbb{E}[X_1 \cdots X_n] \leq \mathbb{E}[X_1^{p_1}]^{1/p_1} \cdots \mathbb{E}[X_n^{p_n}]^{1/p_n}. \tag{2.2}$$

Bernoulli's inequality: Let $x \geq -1$, then

$$(1+x)^r \leq 1 + rx \ \text{ for } \ 0 \leq r \leq 1 \quad \text{ and } \quad (1+x)^r \geq 1 + rx \ \text{ for } \ r \geq 1. \tag{2.3}$$

Note also that if $r \geq 1$ and $0 \leq x \leq 1$ and $x \to 0$ then by the binomial theorem

$$(1-x)^r \leq 1 - rx + (rx)^2/2. \tag{2.4}$$

## 2.1 Probabilistic notions and tools

For a sequence of events $\mathcal{E} := \{\mathcal{E}_n\}_{n=0}^{\infty}$ we say that the event $\mathcal{E}$ holds with high probability (w.h.p.) if $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1$.

For a random variable $X$ let $X \sim_d Y$ denote $X$ being distributed according to the law of $Y$.

For random variables $A, B$, we say that $B$ dominates $A$ if $\mathbb{P}[A > x] \leq \mathbb{P}[B > x]$ for every $x$ and we use the notation $B \succeq_1 A$, or $A \preceq_1 B$ in this case. If $A \preceq_1 B$ and $A, B \geq 0$ then $\mathbb{E}[A^\alpha] \leq \mathbb{E}[B^\alpha]$ for any $\alpha \geq 1$.

Mill's ratio [40, Chapter VII]: Let $X \sim_d \mathcal{N}(\mu, \sigma^2)$, be normally distributed with mean $\mu$ and variance $\sigma^2$. Then

$$\mathbb{P}(|X - \mu| \geq x) \leq \sqrt{\frac{2\sigma^2}{\pi}} \frac{e^{-x^2/2\sigma^2}}{x}. \tag{2.5}$$

Let $Bin(n, p)$ denote the binomial distribution over $n$ trials each of probability $p$. This is the distribution with probability density given by

$$\mathbb{P}(Bin(n,p) = k) := \binom{n}{k} p^k (1-p)^{n-k}.$$

We will make frequent use of the following binomial tail bounds.

**Lemma 2.1.1** (Chernoff bounds [26, Theorem 2.4]). *If $X \sim_d Bin(n, p)$, then for any $a > 0$*

*(i)* $\mathbb{P}[X < np - a] \leq \exp\left(-\frac{a^2}{2np}\right)$,

*(ii)* $\mathbb{P}[X > np + a] \leq \exp\left(-\frac{a^2}{2(np + a/3)}\right)$.

The following is a basic coupling inequality relating two random variables on a common probability space. I could not find a proof for this so I include one for completeness.

**Lemma 2.1.2** (Folklore)**.** *If $X, Y$ are real valued random variables on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, then for any $B \subset \mathbb{R}$,*

$$\Big| \mathbb{P}(X \in B) - \mathbb{P}(Y \in B) \Big| \leq \mathbb{P}(X \neq Y) .$$

*Proof.* First for the lower bound we have

$$\begin{aligned} \mathbb{P}(X \in B) &\geq \mathbb{P}(\{X \in B\} \cap \{X = Y\}) = \mathbb{P}(\{Y \in B\} \cap \{X = Y\}) \\ &= 1 - \mathbb{P}(\{Y \notin B\} \cup \{X \neq Y\}) \geq \mathbb{P}(Y \in B) - \mathbb{P}(X \neq Y) . \end{aligned}$$

Then for the upper bound:

$$\begin{aligned} \mathbb{P}(X \in B) &= \mathbb{P}(\{X \in B\} \cap \{X = Y\}) + \mathbb{P}(\{X \in B\} \cap \{X \neq Y\}) \\ &\leq \mathbb{P}(Y \in B) + \mathbb{P}(X \neq Y) . \end{aligned}$$

Combining these two inequalities yields the result. □

Let $Y$ be a real random variable and $f : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[f(Y)]$ exists. Let $\mathcal{E}$ be any event such that $\mathbb{P}(\mathcal{E}) \geq 1/2$, then

$$\mathbb{E}\left[ f(Y) \middle| \mathcal{E} \right] \leq \frac{\mathbb{E}\left[f(Y)\right]}{\mathbb{P}\left(\mathcal{E}\right)} \leq \mathbb{E}\left[f(Y)\right] + \frac{\mathbb{E}\left[f(Y)\right] \mathbb{P}(\mathcal{E}^c)}{1 - \mathbb{P}(\mathcal{E}^c)} \leq \mathbb{E}\left[f(Y)\right] (1 + 2\mathbb{P}(\mathcal{E}^c)) .$$

We shall use the second moment method for some arguments in this thesis. This is use of the Chebyshev inequality [6, Theorem 4.1.1]: Let $X$ be a real-valued random variable with expected value $\mu$ and variance $\sigma^2$. Then for any $t > 0$

$$\mathbb{P}(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2}. \tag{2.6}$$

Another common form is known Cantelli second moment inequality:

$$\mathbb{P}(X \leq \mu + t) \leq \frac{\sigma^2}{\sigma^2 + t^2}. \tag{2.7}$$

## 2.2 Random walks on graphs and related observables

Throughout we will be working on a finite simple connected graph $G = (V, E)$ with $|V| = n$ and $|E| =: m$. Let $X := (X_t)_{t \geq 0}$ be the simple random walk on $G$. The hitting time $h(i,j)$ is the expected time for $X$ to hit vertex $j$ when started from vertex $i$. That is, if $\mathbf{P}_i^G$ is the law of $X$ on $G$ started from $i \in V$ then

$$h(i,j) := \mathbf{E}_i^G[\tau_j], \quad \text{where } \tau_j := \inf\{t : X_t = j\}.$$

Let $\pi(u) = \gamma_1(u)/(2m)$ be the mass of $u \in V$ with respect to the stationary distribution of the simple random walk $X$ on $G$. We then define the following two functions for $j \in V$,

$$H_j(G) := \sum_{i \in V} \pi(i)h(i,j), \qquad H(G) := \sum_{j \in V} \pi(j)h(i,j). \tag{2.8}$$

The function $H_j(G)$ is known as the stationary hitting time to $j$, $H(G)$ is known as the random target time or Kemeny's constant, see [2, 57]. Kemeny's constant is independent of the vertex $i$, see [58, Eq. 3.3]. The quantities $H(G), H_i(G)$ arise in the study of random walks and Markov chain mixing [2, 57]. Let

$$T(G) := \sum_{i,j \in V} \pi(i)\pi(j)h(i,j) \tag{2.9}$$

be the mean hitting time of $G$, see [2, 57]. The expected running time of Wilson's algorithm for sampling a uniform spanning tree from a connected graph $G$ is $O(T(G))$, [75]. Let $R(i,j)$ be the effective resistance between two vertices $i,j \in V$ with unit resistances on the edges, this is formally defined in Section 2.5. The following sum of resistances is known as the Kirchhoff index, see [22, 45, 54],

$$K(G) := \sum_{\{i,j\} \subseteq V} R(i,j). \tag{2.10}$$

The Kirchhoff index features heavily in Mathematical Chemistry. The cover cost $cc_i(G)$ of a finite connected graph $G$ from a vertex $i$ was studied in [44, 45]. We also introduce the uniform cover cost $\overline{cc}(G)$. For $i \in V$ we define these terms as

$$cc_i(G) := \frac{1}{n-1} \sum_{j \in V} h(i,j), \qquad \overline{cc}(G) := \frac{1}{n(n-1)} \sum_{i,j \in V} h(i,j). \tag{2.11}$$

The hitting times $h(i,j)$ can be far from symmetric in the example of the lollipop graph, which is a path of length $n$ which shares one of its end vertices with vertex from the clique on $n$ vertices, the orders of hitting times range from 1 to $n^3$ [58]. The commute time $\kappa(i,j)$ is the expected number of steps for a random walk from $i$ to reach $j$ and return back to $i$. The commute time $\kappa(i,j)$ is symmetric and related to hitting times and effective resistances by the commute time formula [72]

$$\kappa(i,j) := h(i,j) + h(j,i) = 2m \cdot R(i,j). \tag{2.12}$$

Using (2.12) we can relate the uniform cover cost to the Kirchhoff index

$$\overline{cc}(G) = \frac{1}{n(n-1)} \sum_{(i,j)\in V^2} h(i,j) = \frac{1}{n(n-1)} \sum_{\{i,j\}\subseteq V} \kappa(i,j) = \frac{2m \cdot K(G)}{n(n-1)}. \tag{2.13}$$

The following relation for hitting times is know as Tetali's formula [58]

$$h(i,j) = mR(i,j) + \sum_{u\in V} \frac{\gamma_1(u)}{2} \left[R(j,u) - R(u,i)\right]. \tag{2.14}$$

Relations (2.12), (2.13) and (2.14) will be useful to us as they allow us to control commute times, cover costs and hitting times by effective resistances.

## 2.3 The binomial random graph model

The binomial random graph model $\mathcal{G}(n,p)$ is often referred to as the Erdős-Rényi random graph however it was originally introduced by Gilbert [46]. The model is a probability distribution over simple $n$ vertex graphs, where an $n$ vertex graph $G = (V,E)$ is sampled with probability

$$\mathbb{P}(\mathcal{G} = G) = p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|}.$$

This $\mathbb{P}$ is the product measure over edges of the complete graph $K_n$ where each edge occurs as an i.i.d. Bernoulli random variable with probability $0 < p := p(n) < 1$. Throughout $\mathbb{E}$ will denote expectation with respect to $\mathbb{P}$. We take $\mathcal{G} \sim_d \mathcal{G}(n,p)$ to mean $\mathcal{G}$ is distributed according to the law of $\mathcal{G}(n,p)$. The binomial random graph $\mathcal{G}(n,p)$ is closely related to another random graph model $\mathcal{G}(n,m)$ introduced by Erdős-Rényi [37] around the same time. In $\mathcal{G}(n,m)$ one starts with the empty graph on $n$ vertices and adds $m$ edges uniformly. The two models are very similar although generally $\mathcal{G}(n,p)$ is easier to analyse as there is more independence between events described by collections of edges.

One feature of $\mathcal{G}(n,p)$ worth mentioning is that for each $u \in V$ the degree of $u$ is binomially distributed $\gamma_1(u) \sim_d Bin(n-1,p)$ and the degrees are not independent. The number of edges $|E| = m$ is also binomially distributed, $m \sim_d Bin\left(\binom{n}{2}, p\right)$ since there are $\binom{n}{2}$ potential edges and each occurs independently with probability $p$.

For more information consult one of the many books on random graphs [20, 43, 50].

In this thesis we will study the graph indices described in Section 2.2 when the graph is drawn from $\mathcal{G}(n,p)$, so each of the graph indices becomes a random variable. For any of these random variables to be well defined and finite we need $\mathcal{G}$ to be connected. Take $\mathcal{C} := \mathcal{C}_n$ to be the event $\mathcal{G}$ is connected; we will drop the subscript $n$ where it is implicit. Let $\mathbb{P}_{\mathcal{C}}(\cdot) := \mathbb{P}(\cdot \mid \mathcal{C})$ and $\mathbb{E}_{\mathcal{C}} := \mathbb{E}[\cdot \mid \mathcal{C}]$ be the expectation with respect to $\mathbb{P}_{\mathcal{C}}$. The following theorem gives a bound on being disconnected above the $np = \log n$ connectivity threshold.

**Theorem 2.3.1** ([19, Theorem 9, Ch. VII.3])**.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, $np = \log n + \omega(n)$ where $\omega(n) \to \infty$. Then*

$$\mathbb{P}(\mathcal{C}^c) \leq 4 \cdot e^{-\omega(n)}. \tag{2.15}$$

Let $X_{i,k}$ be the collection of vertices with an $i^{th}$-neighbourhood of size $k$, more formally

$$X_{i,k} := \{v \in V : \gamma_i(v) = k\}.$$

The degree sequence of a graph is the vector $(|X_{1,0}|, |X_{1,1}|, \ldots, |X_{1,n-1}|)$, likewise for some $r \geq 1$ we define the $r$-degree sequence $(|X_{r,0}|, |X_{r,1}|, \ldots, |X_{r,n-1}|)$. The shape of the degree sequence of $\mathcal{G}(n,p)$ will feature heavily in this thesis and so the following theorem will be invaluable.

**Theorem 2.3.2** (Theorem 1 in [18])**.** *Let $\varepsilon > 0$ be fixed, $\varepsilon n^{-3/2} \leq p := p(n) < 1 - \varepsilon n^{-3/2}$, let $k = k(n)$ be a natural number and set $\lambda_k := \lambda_k(n) = n \cdot \binom{n-1}{k} p^k (1-p)^{n-k-1}$, this is the expected number of vertices of degree $k$. Then the following assertions hold.*

*(i) If $\lim_{n\to\infty} \lambda_k(n) = 0$, then $\lim_{n\to\infty} \mathbb{P}(|X_{1,k}| = 0) = 1$.*

*(ii) If $\lim_{n\to\infty} \lambda_k(n) = \infty$, then $\lim_{n\to\infty} \mathbb{P}(|X_{1,k}| > t) = 1$, for every fixed $t$.*

*(iii) If $0 < \liminf_{n\to\infty} \lambda_k(n) < \limsup_{n\to\infty} \lambda_k(n) < \infty$, then $|X_{1,k}|$ has asymptotically Poisson distribution with mean $\lambda_k$: for every fixed $r$*

$$\mathbb{P}(|X_{1,k}| = r) \sim \frac{e^{\lambda_k} \lambda_k^r}{r!},$$

*for every fixed $r$.*

We now have what we need to prove the following Corollary to Theorem 5.3.1 (iii).

**Corollary 2.3.3.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where $np = \omega(\log n)$. Then*

$$\mathbb{P}\left(\sup_{\{i,j\}\subseteq V} |h(i,j) - n| > 11n\sqrt{\frac{\log n}{np}}\right) = o\left(\frac{1}{n^{3/2}}\right).$$

*Proof.* Recall that $m := |E|$ and let $\mu := \binom{n}{2}p$. Define the the following events

$$\begin{aligned}
\mathcal{E}_1 &:= \bigcap_{u\in V} \left\{|\gamma_1(u) - np| \leq 3\sqrt{\log(n)np}\right\} \\
\mathcal{E}_2 &:= \left\{|m - \mu| \leq 3\sqrt{\log(n)\mu}\right\}. \\
\mathcal{E}_3 &:= \bigcap_{\{i,j\}\in V} \left\{\left|R(i,j) - \frac{2}{np}\right| \leq \frac{7\sqrt{\log n}}{(np)^{3/2}}\right\}.
\end{aligned}$$

Recall $\gamma_1(u) \sim_d Bin\,(n-1,p)$, thus by the Chernoff bound, Lemma 2.1.1, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_1^c) &\leq n\cdot\exp\left(-\frac{\left(3\sqrt{\log(n)np}\right)^2}{2(n-1)p}\right) + n\cdot\exp\left(-\frac{\left(3\sqrt{\log(n)np}\right)^2}{2\left(np+\sqrt{\log(n)np}\right)}\right) \\
&= o\left(\frac{1}{n^2}\right).
\end{aligned}$$

Similarly, since for the number of edges $m$ we have $m \sim_d Bin\left(\binom{n}{2}, p\right)$, the Chernoff bound yields $\mathbb{P}(\mathcal{E}_2^c) = o\left(1/n^3\right)$. Recall Theorem 5.3.1 (iii) and observe that

$$\mathbb{P}(\mathcal{E}_3^c) \leq \binom{n}{2}\cdot o\left(\frac{1}{n^{7/2}}\right) = o\left(\frac{1}{n^{3/2}}\right).$$

Conditional on $\mathcal{E} := \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ we have the following by Tetali's formula (2.14)

$$\begin{aligned}
|h(i,j) - n| &= \Bigg|\left(\binom{n}{2}p \pm 3\sqrt{\log(n)\binom{n}{2}p}\right)\left(\frac{2}{np} \pm \frac{7\sqrt{\log n}}{(np)^{3/2}}\right) \\
&\qquad \pm \sum_{u\in V}\frac{1}{2}\left(np \pm 3\sqrt{\log(n)np}\right)\left(\frac{14\sqrt{\log n}}{(np)^{3/2}}\right) - n\Bigg| \\
&= \left|n\left(1 \pm \frac{7\sqrt{\log n}}{2\sqrt{np}}(1\pm o(1))\right) \pm n\left(1 \pm o(1)\right)\frac{7\sqrt{\log n}}{\sqrt{np}} - n\right| \\
&\leq 11n\sqrt{\frac{\log n}{np}}
\end{aligned}$$

for any $\{i,j\} \subseteq V$. The result follows since $\mathbb{P}(\mathcal{E}^c) \leq \mathbb{P}(\mathcal{E}_1^c) + \mathbb{P}(\mathcal{E}_2^c) + \mathbb{P}(\mathcal{E}_3^c) = o\left(\frac{1}{n^{3/2}}\right)$.

□

## 2.4 Reciprocal moments of binomial random variables

We also have the following closed form for moments of binomial random variables,

**Theorem 2.4.1** ([55, Theorem 4.1])**.** *Let $X \sim_d Bin(n,p)$, $n^{\underline{i}} := n(n-1)\dots(n-i+1)$ and $S(d,i)$ be the Stirling partition number of $d$ items into $i$ subsets. Then for $d \geq 0$,*

$$\mathbb{E}\Big[X^d\Big] = \sum_{i=0}^{d} S(d,i)p^i n^{\underline{i}}, \qquad \textit{where} \qquad S(d,i) := \frac{1}{i!}\sum_{k=0}^{i}(-1)^{k+i}\binom{i}{k}k^d.$$

One practical consequence of the above theorem is the following. Let $X \sim_d Bin(n,p)$, $0 < p := p(n) < 1$ and $d \geq 0$ fixed, then Theorem 2.4.1 above yields

$$\mathbb{E}\Big[X^d\Big] = S(d,d)p^d n^{\underline{d}} \pm O\Big(p^{d-1}n^{\underline{d-1}}\Big) = (np)^d \pm O\Big((np)^{d-1}\Big). \tag{2.16}$$

This next Proposition is useful in combination with the lemma following it.

**Proposition 2.4.2.** *Let $X \sim_d Bin(n,p)$, $Y \sim_d Bin(n-1,p)$, $\alpha \in \mathbb{Z}$, $\alpha \geq 1$. Then*

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{X\geq 1\}}}{X^\alpha}\right] := \sum_{k=1}^{n}\frac{1}{k^\alpha}\binom{n}{k}p^k(1-p)^{n-k} = \mathbb{E}\left[\frac{np}{(Y+1)^{\alpha+1}}\right].$$

*Proof.*

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{1}_{\{X\geq 1\}}}{X^\alpha}\right] &:= \sum_{k=1}^{n}\frac{1}{k^\alpha}\binom{n}{k}p^k(1-p)^{n-k}\\
&= \sum_{k=0}^{n-1}\frac{1}{(k+1)^\alpha}\binom{n}{k+1}p^{k+1}(1-p)^{n-1-k}\\
&= \sum_{k=0}^{n-1}\frac{np}{(k+1)^{\alpha+1}}\binom{n-1}{k}p^k(1-p)^{(n-1)-k}\\
&= \mathbb{E}\left[\frac{np}{(Y+1)^{\alpha+1}}\right].
\end{aligned}$$

□

The lemma below gives an upper bound on the expectation of reciprocal powers of $X \sim_d B(n,p)$ when $p := p(n)$ is allowed to tend to 0. This lemma may be of independent interest since other results in the literature appear to require $p$ bounded away from 0.

**Lemma 2.4.3.** *Let $X_n \sim_d Bin(n,p)$ for $p := p(n)$ with $np \to \infty$, $a \in \mathbb{R}$, $b \in \mathbb{Z}$, $a, b > 0$. Then*

$$\frac{1}{(a+np)^b} \leq \mathbb{E}\left[\frac{1}{(a+X_n)^b}\right] \leq \frac{1}{(a+np)^b} + O\left(\frac{1}{(np)^{(b+1)}}\right).$$

*Proof.* Let $f(x) := f_{a,b}(x) = (a+x)^{-b}$ for any constants $a, b > 0$. The lower bound follows from Jensen's inequality since $f(x)$ is convex for $a, b > 0$ and $x > -a$.

Let $\mu_n = \mathbb{E}[X_n] = np$. When $np \to \infty$ it is possible to find some $r := r(n)$ such that $r = \omega(\sqrt{np\log(np)})$ and $r = o(np)$. The Chernoff bound, Lemma 2.1.1 (i), then yields

$$\mathbb{P}(X_n \leq \mu_n - r) \leq \exp\left(-r^2/2\mu_n\right) = o(1/np).$$

With this $r$ we can achieve the following a priori upper bound for any $b \geq 1$:

$$\begin{aligned}\mathbb{E}[f(X_n)] &\leq \frac{1}{a^b}\mathbb{P}(X_n \leq \mu_n - r) + f(\mu_n - r)\mathbb{P}(X_n > \mu_n - r) \\ &= (1+o(1))f(\mu_n). \end{aligned} \tag{2.17}$$

By Taylor's theorem there is some $\xi_n$ between $X_n$ and $\mu_n$ such that

$$f(X_n) = f(\mu_n) + f'(\mu_n)\left(X_n - \mu_n\right) + f''(\xi_n)\left(X_n - \mu_n\right)^2.$$

Using Hölder's inequality (2.2) and the fact $f(x)$ is decreasing when $x > 0$, we have

$$\begin{aligned}(\mathbb{E}[f(X_n)] - f(\mu_n))^2 &\leq \left(f'(\mu_n)\mathbb{E}[X_n - \mu_n] + \mathbb{E}\left[f''(\xi_n)\left(X_n - \mu_n\right)^2\right]\right)^2 \\ &\leq \mathbb{E}\left[f''(\xi_n)^2\right]\mathbb{E}\left[\left(X_n - \mu_n\right)^4\right] \\ &\leq \mathbb{E}\left[f''(X_n)^2\mathbf{1}_{\{X_n \leq \mu_n\}}\right]\mathbb{E}\left[\left(X_n - \mu_n\right)^4\right] \\ &\quad + \mathbb{E}\left[f''(\mu_n)^2\mathbf{1}_{\{X_n > \mu_n\}}\right]\mathbb{E}\left[\left(X_n - \mu_n\right)^4\right] \\ &\leq (2+o(1))f''(\mu_n)^2\mathbb{E}\left[\left(X_n - \mu_n\right)^4\right]. \end{aligned} \tag{2.18}$$

The last inequality follows by (2.17) since $f''(\mu_n) = b \cdot (b+1) \cdot (a+\mu_n)^{-(b+2)}$. Observe

$$\mathbb{E}\left[\left(X_n - \mu_n\right)^4\right] = np(1-p)\left(3p(n-2) - 3p^2(n+2) + 1\right) = O((np)^2), \tag{2.19}$$

this can be calculated using the binomial moment generating function or by Theorem 2.4.1. Hence by (2.18), (2.19) and $(f_{a,b}(x))'' = b(b+1)f_{a,(b+2)}(x)$, we have

$$\begin{aligned}\mathbb{E}[f(X_n)] &\leq f(\mu_n) + \left(O\Big((a+\mu_n)^{-2(b+2)}\Big)\cdot O((np)^2)\right)^{1/2} \\ &= \frac{1}{(a+np)^b} + O\left(\frac{1}{(np)^{b+1}}\right).\end{aligned}$$

□

## 2.5 Electrical network basics

There is a rich connection between random walks on graphs and electrical networks. Here we will give a brief introduction to this area in order to cover essential notation and definitions used for our results; consult one of the books [33, 60, 65] for more complete introduction to the subject.

An electrical network, $N := (G, C)$, is a graph $G$ and an assignment of conductances $C : E(G) \to \mathbb{R}^+$ to the edges of $G$. The resistance $r(e)$ of an edge $e$ is defined by $r(e) := 1/C(e)$.

The graph $G$ is undirected and we define $\vec{E}(G) := \{\vec{xy} : xy \in E(G)\}$, this is the set of all possible orientations of edges in $G$. For some $i, j \in V(G)$, a flow from $i$ to $j$ is a function $\theta : \vec{E}(G) \to \mathbb{R}$ satisfying $\theta(\vec{xy}) = -\theta(\vec{yx})$ for every $xy \in E(G)$ as well as Kirchhoff's node law for every vertex apart from $i$ and $j$, i.e.

$$\sum_{u\in\Gamma_1(v)} \theta(\vec{uv}) = 0 \qquad \text{for each}\, v \in V,\ v \neq i, j.$$

A unit flow from $i$ and $j$ is a flow with strength one, where by strength one we mean

$$\sum_{u\in\Gamma_1(i)} \theta(\vec{iu}) = 1, \qquad \sum_{u\in\Gamma_1(j)} \theta(\vec{uj}) = 1.$$

For the network $N = (G, C)$ we can then define the effective resistance $R_C(i, j)$ between two vertices $i, j \in V(G)$. First for a flow $\theta$ on $N$ let

$$\mathcal{E}(\theta) = \sum_{e\in\vec{E}} \frac{\theta(e)^2}{2C(e)},$$

be the energy dissipated by $\theta$. Then for $i, j \in V(G)$, $R_C(i, j)$ can be defined as

$$R_C(i, j) := \min\{\mathcal{E}(\theta) : \theta \text{ is a unit flow from } i \text{ to } j\}.$$

This is the energy dissipated by the current of strength 1 from $i$ to $j$ in $N = (G, C)$. We will work with unit conductances so we have $C(e) = 1$ for all $e \in E(G)$, when this is the case we write $R(i, j)$ instead of $R_C(i, j)$. This corresponds to the effective resistance in Equations (2.10), (2.12) and (2.14).

A useful property is Rayleigh's monotonicity law [60, § 2.4 ]: If $C, C' : E(G) \to \mathcal{R}^+$ are conductances on the edge set $E(G)$ of a connected graph $G$ and $C(e) \leq C'(e)$ for all $e \in E(G)$ then for all pairs $\{i, j\} \subset V(G)$, we have $R_{C'}(i, j) \leq R_C(i, j)$.

One can also recover the familiar laws of resistors in series and parallel:

- Series Law. Two edges, with resistances $r_1$ and $r_2$, arranged in series are equivalent to a single edge of resistance $r_1 + r_2$.
- Parallel Law. Two edges, with resistances $r_1$ and $r_2$, arranged in parallel are equivalent to a single edge of resistance $r_3$ where $r_3$ satisfies the formula

$$\frac{1}{r_3} = \frac{1}{r_1} + \frac{1}{r_2}.$$

# Chapter 3

# Bounds on effective resistances in graphs

The aim of this section is to obtain lower and upper bounds on $R(u, v)$ for $u, v \in V(G)$ for a graph $G$ where the main contribution to $R(u, v)$ is from the first neighbourhoods of $u$ and $v$. These bounds will later be applied to the binomial random graph $\mathcal{G}(n, p)$.

## 3.1 Lower bounds in terms of degrees

The following is a very useful lower bound on the effective resistance between two vertices of a graph.

**Theorem 3.1.1** (Nash-Williams inequality [64]). *Let $\{\prod_k\}$ be disjoint edge-cutsets which separate the vertex $i$ from $j$. Then*

$$R(i, j) \geq \sum_k \left( \sum_{e \in \prod_k} c(e) \right)^{-1},$$

*where $c(e)$ is the conductance of $e$.*

As a warm up we shall apply the Nash-Williams inequality to locally tree-like graphs. We call a graph on $n$ vertices *locally tree-like* if for a $1 - o(1)$ proportion of vertices the following property holds: for any fixed radius $R$, the ball of radius $R$ around a vertex does not contain a cycle. We will use the Nash-Williams inequality above to prove Proposition 3.1.2, a lower bound on the commute time in regular tree-like graphs.

**Proposition 3.1.2.** *Let $d \geq 3$ be fixed and $G$ be connected d-regular locally tree-like graph. Then for $(1-o(1))\binom{n}{2}$ pairs of vertices $\{i,j\} \subseteq V$ we have*

$$\kappa(i,j) \geq (2-o(1))\frac{(d-1)}{(d-2)}n.$$

*Proof.* In any $n$ vertex graph of maximum degree $d$ it holds that $(1-o(1))\binom{n}{2}$ pairs of vertices are at distance at least $\Omega(\log n/\log d)$ from each other. To see this let $M = \log n/(2\log d)$, then

$$\begin{aligned}|\{\{i,j\} \subset V : d(i,j) \leq M\}| &\leq \sum_{i \in V} |\{j \in V : d(i,j) \leq M\}| \\ &\leq nd^M \\ &= o\left(n^2\right).\end{aligned}$$

If in addition the graph is locally tree-like then a $(1-o(1))$ proportion of these "well separated" pairs also have finite neighbourhoods that look like d-regular trees. For two such vertices $i,j$ we can take the edges at distance $0 \leq k \leq M$ from $\{i,j\}$ as disjoint edge-cutsets - where distance of an edge from a vertex is measured by the graph distance to the nearest end vertex of the edge.

By the Nash-Williams inequality we have

$$\begin{aligned}R(i,j) &\geq \sum_k \left(\sum_{e \in \prod_k} 1\right)^{-1} \\ &\geq 2\sum_{k=0}^{M} \frac{1}{d}\frac{1}{(d-1)^k} \\ &= \frac{2}{d}\left(\frac{1-\frac{1}{(d-1)^M}}{1-\frac{1}{d-1}}\right) \\ &= (1-o(1))\frac{2}{d}\frac{(d-1)}{(d-2)}.\end{aligned}$$

Then by the commute time formula:

$$\kappa(i,j) = 2mR(i,j) \geq 2\cdot\frac{dn}{2}\cdot(1-o(1))\frac{2}{d}\frac{(d-1)}{(d-2)} = (2-o(1))\frac{(d-1)}{(d-2)}n.$$

□

Note this lower bound from the Nash-Williams inequality gives the correct constant for random regular graphs as Cooper, Freize and Radzik [29] claim that most hitting times in a random $d$-regular graph have the asymptotic $h(i,j) \sim \frac{(d-1)}{(d-2)}n$.

Recall that $\gamma_1(v)$ denotes the size of the first neighbourhood of vertex $v \in V(G)$. Jonasson gives the following lower bound on effective resistance, we shall include a proof for completeness as it follows easily from the Nash-Williams inequality, Theorem 3.1.1

**Lemma 3.1.3** ([51, Lemma 1.4])**.** *For any graph $G = (V, E)$ and $u, v \in V$, $u \neq v$*

$$R(u,v) \geq \frac{1}{\gamma_1(u)+1} + \frac{1}{\gamma_1(v)+1}.$$

*Proof.* There are two cases; either $uv \notin E$ or $uv \in E$. In the first case, since $uv \notin E$, the edges adjacent to $u$ and the edges adjacent to $v$ form two disjoint edge-cutsets. These two edge-cutsets have sizes $\gamma_1(u)$ and $\gamma_1(v)$ respectively. Thus by the Nash-Williams inequality, Theorem 3.1.1, we have

$$R(u,v) \geq \frac{1}{\gamma_1(u)} + \frac{1}{\gamma_1(v)}.$$

In the second case we replace the edge $uv$ with two internally disjoint paths of length two both beginning at $u$ and ending at $v$. This adds two vertices of degree two to the graph which are both attached to $u$ and $v$ alone and by the laws of resistors in series and parallel the resistance of these two paths in parallel is the same as the single edge $uv$. We are now in a situation however where $u$ is no longer connected to $v$ by an edge. Thus we can revert to the first case taking the edges adjacent to $u$ and the edges adjacent to $v$ form two disjoint edge-cutsets with sizes $\gamma_1(u)+1$ and $\gamma_1(v)+1$ respectively. Thus by the Nash-Williams inequality, Theorem 3.1.1, we have

$$R(u,v) \geq \frac{1}{\gamma_1(u)+1} + \frac{1}{\gamma_1(v)+1}.$$

Since in this case the resistance is always lower we take this bound. □

Observe that although the above bound holds for any two distinct vertices it is only really meaningful if they are in the same connected component. This is since otherwise the effective resistance between the two vertices is defined to be infinite.

We are now ready to prove the Proposition 3.1.4, this proposition featured in the introduction as a example to show that close to the connectedness threshold we have some hard to reach vertices. This is proven using Lemma 3.1.3 and some facts about random graphs and random walks from Chapter 2.

**Proposition 3.1.4.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$. If $np = \log(n) + 100 \log\log\log(n)$, then*

$$\mathbb{P}(\textit{there exists } i, j \in V : h(i,j) > n\log(n)/3) = 1 - o(1).$$

*For any $1 < c < \infty$ if $np = c\log(n)\,(1 \pm o(1))$, then there is an $a > 0$ such that*

$$\mathbb{P}(\textit{there exists } i, j \in V : h(i,j) > (1+a)n) = 1 - o(1).$$

*Proof.* Let $X_{1,1} := \{v \in V : \gamma_1(v) = 1\}$ be the number of vertices in $\mathcal{G}$ with degree 1. Let $\lambda_1(n)$ be the expected number of vertices of degree 1, thus

$$\begin{aligned}
\lambda_1(n) &= n \cdot \binom{n-1}{1} p(1-p)^{n-2} \\
&= n^2 p e^{-\log(n) - 100\log\log\log(n) + O\left(\log(n)^2/n\right)} \\
&\geq \log(n) e^{-100\log\log\log(n)}.
\end{aligned}$$

Thus $\lambda_1(n) \to \infty$ as $n \to \infty$ and so by Theorem 2.3.2 for any fixed $t$

$$\lim_{n\to\infty} \mathbb{P}(|X_{1,1}| \geq t) = 1.$$

Thus with high probability there are is at least one pair of vertices $i, j$ both with degree 1. We have the following lower bound on the effective resistance between two vertices of degree 1 by Lemma 3.1.3

$$R(i,j) \geq \frac{1}{\gamma_1(i)+1} + \frac{1}{\gamma_1(j)+1} = \frac{1}{2} + \frac{1}{2} = 1.$$

Let $m := |E|$ be the number of edges in $\mathcal{G} \sim_d \mathcal{G}(n,p)$. We know that $m$ is a binomially distributed random variable with parameters $\binom{n}{2}$ and $p$ since each edge occurs with probability $p$ independently from the others. Therefore by the Chernoff bound, Lemma 2.1.1, we have

$$\mathbb{P}\left(m < \binom{n}{2}p - \sqrt{2\log(n)\binom{n}{2}p}\right) \leq \exp\left(-\frac{\left(\sqrt{2\log(n)\binom{n}{2}p}\right)^2}{2\binom{n}{2}p}\right) \leq e^{-\log(n)}. \quad (3.1)$$

Hence with probability $1 - o(1) - 1/n$ there is a pair of vertices $i, j$ both with degree 1 and $m \geq \binom{n}{2}p - \sqrt{2\log(n)\binom{n}{2}p}$. Thus by the commute time formula 2.12 we have

$$\kappa(i,j) = 2m \cdot R(i,j) \geq 2\left(\binom{n}{2}p - \sqrt{2\log(n)\binom{n}{2}p}\right) \cdot 1 = (1 - o(1))n\log(n).$$

Since the commute time $\kappa(i,j) = h(i,j) + h(j,i)$ at least one of $h(i,j)$ of $h(j,i)$ must be greater than $n\log(n)/3$ with probability $1 - o(1) - 1/n$.

We now turn to the proof for the case $np = (c + o(1))\log(n)$, this follows the rough idea of the proof of [51, Theorem 2.2]. Let $k = (1-\varepsilon)np$ for some $0 < \varepsilon < 1$ and we shall again appeal to Theorem 2.3.2 with the following argument

$$\lambda_k = n\binom{n-1}{k}p^k(1-p)^{n-1-k} \geq \frac{n}{\sqrt{2\pi k}}\left(\frac{e}{(1-\varepsilon)}\right)^{(1-\varepsilon)np} e^{-np}(1-o(1))$$

Observe that $-\log(1-t) = +t - t^2/2 + t^3/3 + t^4/4 \cdots \geq t + t^2/2$ for $t < 1$ we have by the Taylor series. Thus we have the following

$$\lambda_k \geq \frac{n}{3\sqrt{k}} e^{-\varepsilon np - (1-\varepsilon)\log(1-\varepsilon)np} \geq \frac{n}{3\sqrt{k}} e^{-\varepsilon np + (1-\varepsilon)(+\varepsilon+\varepsilon^2/2)np} \geq \frac{ne^{-\frac{\varepsilon^2(1+\varepsilon)np}{2}}}{3\sqrt{k}}$$

Thus for any $0 < \varepsilon < 1$ satisfying $\frac{\varepsilon^2(1+\varepsilon)c}{2} < 1$, a concrete example would be $\varepsilon = \sqrt{1/(c+1)}$, we have that $\lambda_k \to \infty$ and so by Theorem 2.3.2 there are at least two vertices with degree less than $(1-\varepsilon)np$ w.h.p. Thus, as before, by (3.1) and the commute time formula 2.12 we have the following w.h.p. for such a pair of vertices

$$\kappa(i,j) = 2m \cdot R(i,j) \geq 2\left(\binom{n}{2}p - \sqrt{2\log(n)\binom{n}{2}p}\right) \cdot \frac{2}{(1-\varepsilon)np} = \frac{(2-o(1))n}{1-\varepsilon}.$$

Thus one or both of the hitting times $h(i,j)$ or $h(j,i)$ must be greater than $(1+a)n$ for some $a > \frac{\varepsilon}{2(1-\varepsilon)} > 0$ with high probability. □

## 3.2 Upper bounds on the effective resistance

The first upper bound on the effective resistance is of course the trivial bound $R(i,j) \leq d(i,j)$. This follows since if $u$ and $v$ are connected then there is a path of length $d(i,j)$ between them so we can send one unit of current through this path, the resistance of the path is $d(i,j)$ by the formula for resistors in series. To get any better bounds than this on the effective resistance something must be assumed about the graph.

One direction is to bound the resistance in terms of isoperimetric or expansion quantities. Expansion in a graph can be defined in many ways however one definition that seems fairly universal and will be of use to us for stating the next theorem is the following from [23], which first appeared in [4]. Firstly we define the vertex boundary $\partial A$ of set $A$ of vertices to be the vertices of $G\backslash A$ with neighbours in $A$.

An $(n, d, \alpha)$-expander is a graph $G = (V, E)$ on $n$ vertices of maximal degree $d$, such that every subset $X \subseteq V$ satisfying $|X| \leq n/2$ has $|\partial X| \geq \alpha |X|$.

Expander graphs have many application in graph theory, computer science and beyond [47]. Chandra et. al. [24] showed the following bound on the maximum effective resistance in an expander.

**Theorem 3.2.1** ([24, Theorem 5.2.]). *A connected $(n, d', \alpha)$-expander $G$ with minimum degree $d$, has resistance at most* $24/(\alpha^2(d+1))$.

This was later refined by Sauerwald and Stauffer to give the following.

**Corollary 3.2.2** ([68, Corollary 6.1]). *For any graph with vertex expansion $\alpha$,*

$$R(s,t) \leq \frac{24}{\alpha^2 \cdot (\min\{d(s), d(t)\} + 1)}.$$

The binomial random graph $\mathcal{G}(n,p)$ shares many properties with expander graphs and typical sets of vertices have a large vertex boundary. Although $\mathcal{G}(n,p)$ satisfies good expansion properties trying to pin down the specific dependence, on $p$, of the vertex-isoperimetrical constant $\alpha$ is challenging. Benjamini et. al. study the related edge-isoperimetrical constant [10]. Benjamini and Kozma present the following upper bound for the effective resistance [11, 12], where the second citation is an amended version of the same paper.

**Theorem 3.2.3** ([11, Theorem 2.1.]). *Let $G$ be a finite graph. Let $w$ and $u$ be vertices of $G$. Let $R(w,u)$ be the electric resistance between $w$ and $u$. Then*

$$R(w,u) \leq C\left(L_w + L_u\right), \quad L_v := \sum_{i=1}^{\lfloor \log_2 |G| \rfloor} \max_{\substack{v \in A \\ A \text{ connected} \\ |G|2^{-(i-1)} \leq |A| \leq |G|2^{-i}}} \frac{|A|}{|\partial A|^2} + \frac{1}{|\partial A|}$$

If we wish to apply either of the last two bounds to $\mathcal{G}(n,p)$ then, at best, they should give us $R(i,j) \leq O(1/\gamma_1(u) + 1/\gamma_1(u))$, which is what we want up a constant. We seek a bound which allows us to recover the correct constant for the leading order term thus we will need a new upper bound. This is what we present in the remainder of this chapter.

## 3.3 The strong $k$-path property

We now aim to obtain an upper bound where the dominant term looks roughly like the one in Lemma 3.1.3. To achieve this we analyse the following modified breadth-first search (MBFS) algorithm. The inputs to the MBFS algorithm are a graph $G$

and a subset $I_0 = \{u, v\} \subseteq V(G)$, the outputs are sets $I_i, S_i \subseteq V(G)$ and $E_i \subseteq E(G)$ indexed by the iteration of the algorithm. The algorithm is similar to one used in [6, Ch. 11.5] to explore the giant component of an Erdős-Rényi graph. However, the MBFS algorithm differs from other variations on breadth-first search algorithms used in the literature as it starts from two distinct vertices. More importantly it also differs by removing clashes, where a clash is a vertex with more than one parent in the previous generation as exposed by a breadth-first search from two root vertices.

**Modified breadth-first search algorithm,** MBFS$(G, I_0)$**:** To begin set $S_0 := V\backslash I_0$, and $S_i = I_i = E_i = \emptyset$ for all $i \geq 1$. Then generate the sets $S_i, I_i$ and $E_i$ for $i \geq 1$ iteratively by the following procedure:

Step 1: For each $w \in S_{i-1}$ check all pairs $\{w, w'\}$ where $w' \in I_{i-1}$ and do the following:

- If, for every $w' \in I_{i-1}$, $ww' \notin E(G)$ then add $w$ to $S_i$.
- If there is one unique $w' \in I_{i-1}$ such that $ww' \in E(G)$ then add $w$ to $I_i$ and add $ww'$ to $E_i$.

Step 2: If $S_i \neq S_{i-1}$ then advance $i$ to $i + 1$ and return to Step 1. Otherwise end.

The set $I_i$ contains the "active" vertices in the $i^{th}$ iteration and $S_i$ is the set of vertices that have not been used in the first $i$ iterations and $E_i$ is the set $\{xy \in E(G) : x \in I_{i-1}, y \in I_i\}$ of edges "accepted" by the algorithm. Notice that $S_0 \supseteq S_1 \supseteq S_2 \dots$ and the sets $\{I_i\}_{i \geq 0}$ are all disjoint. A vertex in $S_i$ will not be added to either $I_{i+1}$ or $S_{i+1}$ if it has two or more neighbourhoods in $I_i$, in this instance it is just ignored by the algorithm. If instead those vertices in $S_i$ with edges to more than one vertex in $I_i$ we added them to $I_{i+1}$ then this procedure would describe a standard breadth-first search starting from two root vertices. Notice also that in Step 1 the order in which we consider the vertices of $S_i$ and then the edges between $S_i$ and $I_i$ is unimportant.

For each pair of vertices $I_0 \subseteq V$ the MBFS algorithm provides a filtration

$$\mathfrak{F}_i := \mathfrak{F}_i(I_0), \tag{3.2}$$

where $\mathfrak{F}_0 \subseteq \mathfrak{F}_1 \subseteq \cdots$, on the set of labelled graphs on $V$. Roughly speaking $\mathfrak{F}_i(I_0)$ only sees graphs that are the distinguishable by MBFS run up to step $i \geq 0$ from initial set $I_0$. To make this precise we must first describe an equivalence relation on graphs. Let $u, v \in V$ and $G, F$ be graphs on $V$. We say $G \cong_k^{\{u,v\}} F$ if the same $k$-sequence of sets $\{S_i, I_i, E_i\}_{1\le i\le k}$ is output when MBFS$(G, \{u, v\})$ and MBFS$(F, \{u, v\})$ are run for $k$ iterations. Let $I_0 = \{u, v\} \subseteq V$ and define $\mathfrak{F}_i(I_0)$ to be the $\sigma$-algebra where the atoms are the equivalence classes of $\cong_i^{\{u,v\}}$.

Let $x \in I_k$ where $I_k$ is produced by running MBFS$(G, I_0)$ for some given $I_0$. We shall now definition $\Gamma_i^*(x)$, the MBFS neighbourhood of $x$, for $i \ge 1$

$$\Gamma_i^*(x) = \left\{ y \in I_{k+i} : \begin{array}{c} \text{there exists } x = x_0, x_1, \ldots, x_i = y \text{ where} \\ x_{j-1}x_j \in E_{k+j} \text{ for all } j = 1, \ldots, i \end{array} \right\}, \tag{3.3}$$

and $\Gamma_0^*(x) := x$. If we restrict to the subgraph of $G$ output by MBFS$(G, I_0)$ then $\Gamma_i^*(x)$ is the subset of the $i^{th}$ neighbourhood of $x \in I_k$ which is furthest from the set $I_0$. We can also define $\Gamma_i^*(x)$ for $i \ge 1$ inductively

$$\Gamma_i^*(x) := \left\{ z \in I_{k+i} : \text{ there exists } y \in \Gamma_{i-1}^*(x) \text{ and } yz \in E_i \right\}.$$

To try and clarify (3.3) we define the following sets $S_k(x)$ which are the vertices in $S_k$ that will not cause any clashes when the $\Gamma^*$-neighbourhood of $x$ is explored,

$$S_k(x) := S_k \Big\backslash \left( \bigcup_{z\in I_k,\, z\ne x} \Gamma_1(z) \right). \tag{3.4}$$

We can then also define the neighbourhood $\Gamma_i^*(x)$ inductively by the following

$$\Gamma_i^*(x) = \bigcup_{y\in\Gamma_{i-1}^*(x)} \Gamma_1(y) \cap S_{k+i}(y).$$

Define for some constant $d$ the pruned neighbourhood $\Phi_1^d(x)$ of $x \in I_1$ by

$$\Phi_1^d(x) := \Gamma_1^*(x) \backslash \{y : \gamma_1^*(y) \le d\}, \qquad \text{and let} \qquad \varphi_1^d(x) := \left|\Phi_1^d(x)\right|. \tag{3.5}$$

This is the MBFS neighbourhood of $x$ with all the neighbours who have less than $d$ "MBFS-children" removed. Define the pruned neighbourhoods $\Psi_1^d(w)$ of $w \in I_0$ by

$$\Psi_1^d(w) := \Gamma_1^*(w) \backslash \left\{ y : \varphi_1^d(y) = 0 \right\}, \qquad \text{and let} \qquad \psi_1^d(w) := \left|\Psi_1^d(w)\right|. \tag{3.6}$$

We define $\Psi_0^d(u) = \{u\}$ and the pruned second neighbourhood $\Psi_2^d(w)$ of $w \in I_0$ by

$$\Psi_2^d(w) := \bigcup_{x\in\Psi_1^d(w)} \Phi_1^d(x) = \bigcup_{x\in\Gamma_1^*(w)} \Phi_1^d(x). \tag{3.7}$$

For MBFS$(G, \{u, v\})$ define $\Psi_i^d$, the pruned version of $I_i$ for $i = 0, 1, 2$, by

$$\Psi_i^d := \Psi_i^d(u) \cup \Psi_i^d(v), \quad i = 0, 1, 2.$$

We prune the first neighbourhoods of vertices $x \in I_1$ to obtain $\Phi_1^d(x)$ so that later on when we consider the trees induced by the union up to $i$ of the $\Gamma^*$-neighbourhoods of $y \in \Phi_1^d(x)$ we can get good control over the growth rate of the trees. We prune the first neighbourhoods of vertices $w \in I_0$ as above so that we can send flow from our source vertex $w$ to its pruned neighbourhood $\Psi_1^d(w)$ without having to worry about it getting stuck in any "dead ends".

Recall (3.2), the definition of the filtration $\mathfrak{F}_k(I_0)$. Observe that if $x \in I_k$ then $\Gamma_1^*(x)$ is $\mathfrak{F}_{k+1}$ measurable. It is worth noting however that if $y \in I_1$ then $\Phi_1^d(y)$ is $\mathfrak{F}_3$ measurable and not necessary $\mathfrak{F}_2$ measurable since $\Phi_1^d(y)$ is determined by vertices at distances 2 and 3 from $I_0$. A consequence of this is that for $w \in I_0$, $\Psi_1^d(w), \Psi_2^d(w)$ are both $\mathfrak{F}_3$ measurable as they are both determined by the $\Phi_1^d$-neighbourhoods of points in $\Gamma_1^*(w)$.

We use the sets $\Psi$ and $\Gamma^*$ returned from running the MBFS algorithm on a graph $G$ in the following definitions.

**Definition 3.3.1** (Strong $k$-path property)**.** *We say that a graph $G$ on $[n] := \{1, \ldots, n\}$ has the strong $k$-path property for integers $k, d \geq 0$ and a pair of vertices $u$,$v$ if for every pair $(x, y) \in \left(\Psi_2^d(u) \times \Psi_2^d(v)\right)$ the neighbourhoods $\Gamma_k^*(x)$ and $\Gamma_k^*(y)$ are non-empty and there is at least one edge $ij \in E(G)$ where $i \in \Gamma_k^*(x)$, $j \in \Gamma_k^*(y)$.*

For $u, v \in [n]$ and $k, d \geq 0$ we let $A_{u,v}^{n,k,d}$ be the set of graphs on $[n]$ satisfying the strong $k$-path property for $u, v$ and $d$. We refer to this as the strong $k$-path property and slightly neglect the dependence on $d$ as we shall see in the next chapter that for random graphs the value of $d$ does not make so much difference to the results.

The sets $B_w^{u,v,d}$ for $w \in \{u, v\}, d \geq 0$ are also defined using the output of MBFS$(G, \{u, v\})$:

$$B_w^{u,v,d} := \left\{G : V = [n], \Psi_1^d(w) \neq \emptyset\right\}, \quad \text{and let} \quad B_{u,v}^d := B_u^{u,v,d} \cap B_v^{u,v,d}. \qquad (3.8)$$

## 3.4 An upper bound on effective resistance for graphs satisfying the $k$-path property

The next Theorem provides an upper bound on the effective resistance for graphs satisfying the strong $k$-path property. We will show in the next chapter that the

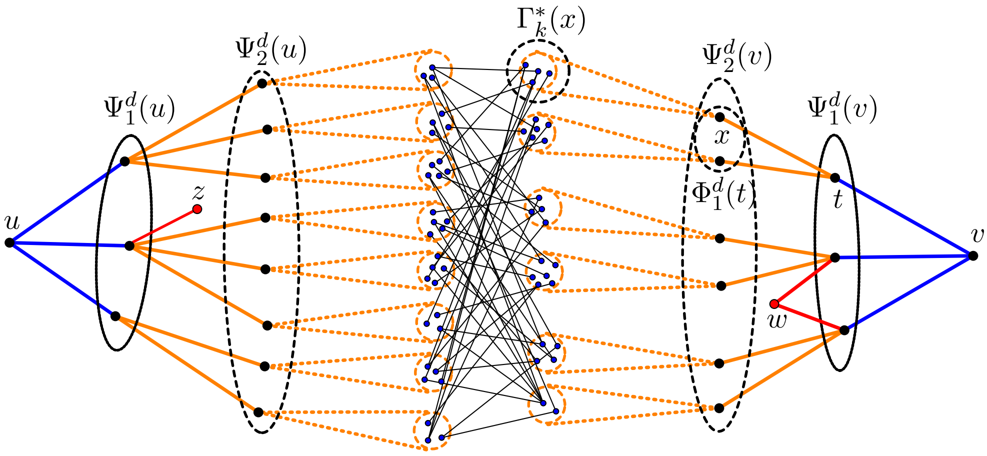


Figure 3.1: This diagram illustrates the strong $k$-path property $A^{n,k,d}_{u,v}$, see definition 3.3.1. In the above example the the vertex $z$ is not in $\Psi^d_2(u)$ since it is connected to less than $d$ vertices in $I_3$ and the vertex $w$ is not in $I_2$ as it has more than one parent in $I_1$.

strong $k$-path property is satisfied in sparsely connected $\mathcal{G}(n,p)$ for specified $k$. We then apply Theorem 3.4.1 to obtain estimates for the resistance and hitting times in $\mathcal{G}(n,p)$. Theorem 3.4.1 may potentially be applied to bound resistance in other related random graph models such as random intersection graphs [43, Ch. 11] and Chung-Lu graphs [26] in certain regimes. The regimes where this approach may be most valuable are regimes where there is constant minimum degree, the range of the degrees is large and it is hard to get good enough control on the spectral statistics to apply spectral methods to obtain estimates on the hitting times with the correct leading constant.

**Theorem 3.4.1.** *Run* $\mathrm{MBFS}(G, \{i,j\})$ *and suppose* $G \in A^{n,k,d}_{i,j}$*. Then*

$$R(i,j) \leq \frac{1}{\psi^d_1(i)} + \frac{1}{\psi^d_1(j)} + \sum_{a \in \Psi^d_1(i)} \frac{k+2}{\psi^d_1(i)^2 \varphi^d_1(a)} + \sum_{b \in \Psi^d_1(j)} \frac{k+2}{\psi^d_1(j)^2 \varphi^d_1(b)}.$$

*Proof.* We will follow the convention that $1/0 = \infty$. If $G \notin B_{i,j}$ then the bound holds trivially as at least one of the first two terms on the right is infinite.

We will now define a graph $H$ which must exist as a sub-graph of $G$ whenever $G \in A^{n,k,d}_{i,j} \cap B_{i,j}$. The sub-graph $H$ will be defined as a union of many sub-graphs of $G$ which are themselves described by the sets produced from running $\mathrm{MBFS}(G, \{i,j\})$.

Define $U_w$, $w \in I_0$ to be the graph on $V(U_w) := \Psi^d_0(w) \cup \Psi^d_1(w) \cup \Psi^d_2(w)$, where $\Psi_0^d(w) = \{w\}$, with edge set

$$E(U_w) := \left\{ yz \in E(G) : y \in \Psi_h^d(w),\ z \in \Psi_{h+1}^d(w),\ h = 0, 1 \right\}.$$

For each $x \in \Psi_2^d$ define the tree $T_k(x)$ to be the tree on $V(T_k(x)) := \bigcup_{i=0}^{k} \Gamma_i^*(x)$ with edge set

$$E(T_k(x)) := \{yz \in E(G) : y \in \Gamma_i^*(x), z \in \Gamma_{i+1}^*(x), 0 \le i \le k-1\}.$$

By the strong $k$-path property there is at least one edge $\tilde{x}\tilde{y} \in E(G)$ where $\tilde{x} \in \Gamma_k^*(x)$, $\tilde{y} \in \Gamma_k^*(y)$ for all pairs $(x, y) \in \Psi_2^d(i) \times \Psi_2^d(j)$. If there is more than one edge we select one and disregard the others. Let this set of edges be $E^*$. Let $F$ be the graph $E(F) = E^*$ and $V(F) := \{z : zw \in E^*\}$. Thus $F$ is a set of edges complete with end vertices which bridge some leaf of tree $T_k(x)$ to some leaf of $T_k(y)$ for each pair $(x, y) \in \Psi_2^d(i) \times \Psi_2^d(j)$.

With the above definitions the sub-graph $H$ is then

$$H := U_i \cup U_j \cup \left( \bigcup_{x \in \Psi_2^d} T_k(x) \right) \cup F.$$

Consult Figure 3.1 for more details. We will now describe a unit flow $\theta$ from $i$ to $j$ through the network $N = (H, C)$ where $C(e) = 1$ for all $e \in E(H)$. This flow will be used to bound from above the effective resistance $R(i, j)$ in $G$.

(i) Since $G \in B_{i,j}$ we have $\psi_1^d(i), \psi_1^d(j) \ge 1$, thus we can send a flow $\theta(\vec{ii_a}) = 1/\psi_1^d(i)$ through each edge $\vec{ii_a} \in \vec{E}(H_i)$ where $i_a \in \Psi_1^d(i)$. Likewise assign a flow $\theta(\vec{jj_b}) = -1/\psi_1^d(j)$ to each edge $\vec{jj_b} \in \vec{E}(H_j)$ where $j_b \in \Psi_1^d(j)$. Observe that one unit of flow leaves $i$ and enters $j$. The contribution to $\mathcal{E}(\theta)$ from the flow through these edges is

$$\sum_{i_a \in \Psi_1^d(i)} \frac{1}{\left(\psi_1^d(i)\right)^2} + \sum_{j_b \in \Psi_1^d(j)} \frac{1}{\left(\psi_1^d(j)\right)^2} = \frac{1}{\psi_1^d(i)} + \frac{1}{\psi_1^d(j)}.$$

(ii) For each $i_a \in \Psi_1^d(i)$ we send the flow $\theta(\vec{i_a i_{a,f}}) = 1/\left(\varphi_1^d(i_a)\psi_1^d(i)\right)$ through each edge $\vec{i_a i_{a,f}} \in \vec{E}(H_i)$ where $i_{a,f} \in \Phi_1^d(i_a)$. Likewise for each $j_b \in \Psi_1^d(j)$ we send a flow $\theta(\vec{j_b j_{b,h}}) = -1/\left(\varphi_1^d(j_b)\Psi_1^d(j)\right)$ through each edge $\vec{j_b j_{b,h}} \in \vec{E}(H_j)$ where $j_{b,h} \in \Phi_1^d(j_b)$. By definition of $\Psi_1^d(i)$, $\Psi_1^d(j)$ the sets $\Phi_1^d(i_a)$ and $\Phi_1^d(j_b)$ are non-empty so this is well defined. We see that Kirchhoff's node law is satisfied

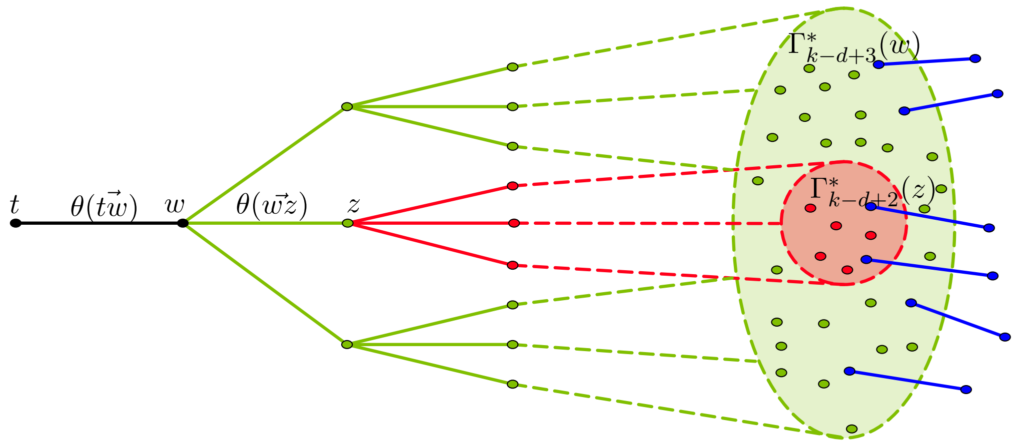


Figure 3.2: The descendants of $t \in I_{d-2}$ in the tree $T_k(i_{a,f})$ rooted at $i_{a,f}$, where the notation is consistent with Step (iv) from the proof of Theorem 3.4.1. Here the descendants of $w$ are shown in green and those that also have $z$ as an ancestor are are shown in red. The edges of $E^*$ and their endpoints are shown in blue.

at each vertex $i_a \in \Psi_1^d(i)$ since

$$\sum_{i_{a,f} \in \Phi_1^d(i_a)} \theta\left(\vec{i_a i_{a,f}}\right) = \sum_{i_{a,f} \in \Phi_1^d(i_a)} \frac{1}{\psi_1^d(i)\varphi_1^d(i_a)} = \frac{1}{\psi_1^d(i)} = \theta(\vec{i i_a}),$$

and likewise for each $j_b \in \Psi_1^d(j)$. The contribution to $\mathcal{E}(\theta)$ from these edges is

$$\sum_{w \in \{i,j\}} \sum_{\substack{w' \in \\ \Psi_1^d(w)}} \sum_{\substack{w'' \in \\ \Phi_1^d(w')}} \frac{1}{\left(\psi_1^d(w)\varphi_1^d(w')\right)^2}$$
$$= \sum_{i_a \in \Psi_1^d(i)} \frac{1}{\psi_1^d(i)^2 \varphi_1^d(i_a)} + \sum_{j_b \in \Psi_1^d(j)} \frac{1}{\psi_1^d(j)^2 \varphi_1^d(j_b)}.$$

(iii) For each edge $\vec{xy} \in \vec{E}(F)$ let $i_{a,f}$ denote the unique vertex in $\Phi_2(i)$ such that $x \in T_k(i_{a,f})$ and $j_{b,h} \in \Phi_2(j)$ denote the unique vertex such that $y \in T_k(j_{b,h})$. There is some unique $i_a \in \Psi_1^d(i)$ such that $i_{a,f} \in \Phi_1^d(i_a)$ and $j_b \in \Psi_1^d(j)$ such that $j_{b,h} \in \Phi_1^d(j_b)$. We then assign the following flow to $\vec{xy}$:

$$\theta(\vec{xy}) = \frac{1}{\varphi_1^d(i_a)\psi_1^d(i)\varphi_1^d(j_b)\psi_1^d(j)}.$$

The reason for this is that if we sum the flows leaving $T_k(i_{a,f})$ through the vertex set $\Gamma_k^*(i_{a,f})$ for some $i_a \in \Psi_1^d(i)$ with $i_{a,f} \in \Phi_1^d(i_a)$ we obtain

$$\sum_{j_b \in \Psi_1^d(j)} \sum_{j_{b,h} \in \Phi_1^d(j_b)} \frac{1}{\psi_1^d(i)\varphi_1^d(i_a)\psi_1^d(j)\varphi_1^d(j_b)} = \frac{1}{\varphi_1^d(i_a)\psi_1^d(i)},$$

which is the amount of flow entering $T_k(i_{a,f})$ at the vertex $i_{a,f}$ and likewise for the trees $T_k(j_{b,h})$ for every $j_b \in \Psi_1^d(j)$ with $j_{b,h} \in \Phi_1^d(j_b)$. In the next step we show Kirchhoff's node law will be satisfied at each vertex in $V(F)$ by virtue of the assignment of flow through the trees $T_k(i_{a,f})$, $T_k(j_{b,h})$. The contribution to $\mathcal{E}(\theta)$ by the sub-graph $F$ is

$$\sum_{i_a \in \Psi_1^d(i)} \sum_{j_b \in \Psi_1^d(j)} \sum_{i_{a,f} \in \Phi_1^d(i_a)} \sum_{j_{b,h} \in \Phi_1^d(j_b)} \frac{1}{\left(\psi_1^d(i)\varphi_1^d(i_a)\psi_1^d(j)\varphi_1^d(j_b)\right)^2}$$

$$\leq \sum_{i_a \in \Psi_1^d(i)} \frac{1}{\psi_1^d(i)^2\varphi_1^d(i_a)} + \sum_{j_b \in \Psi_1^d(j)} \frac{1}{\psi_1^d(j)^2\varphi_1^d(j_b)}.$$

The inequality above follows since when $G \in B_{i,j}$ we have $\psi_1^d(i), \psi_1^d(j) \geq 1$ and $\varphi_1^d(i_a), \varphi_1^d(j_b) \geq 1$ by definition for all $i_a \in \Psi_1^d(i), j_b \in \Psi_1^d(j)$.

(iv) For each $\vec{uz} \in \vec{E}\left(T_k(i_{a,f})\right)$ we set $\theta(\vec{uz})$ proportional to the amount of flow leaving $z$'s descendants in the set $\Gamma_k^*(i_{a,f})$, see Figure 3.2. If $z \in I_s$ then let $t$ be the parent of $u$ when $T_k(i_{a,f})$ is rooted at $i_{a,f}$ and let $t = i_a$ if $u = i_{a,f}$. We set

$$\theta(\vec{uz}) := \left( \frac{\sum_{x \in \Gamma_{k-s+2}^*(z)} \sum_{xy \in E^*} \theta(\vec{xy})}{\sum_{x \in \Gamma_{k-s+3}^*(u)} \sum_{xy \in E^*} \theta(\vec{xy})} \right) \cdot \theta(\vec{tu}).$$

Kirchhoff's node law is satisfied at each vertex $u \in V\left(T_k(i_{a,f})\right)$ since

$$\sum_{z \in \Gamma_1^*(u)} \theta(\vec{uz}) = \theta(\vec{tu}).$$

Let $\mathcal{E}\left(\theta | T_k(x)\right) := \sum_{e \in \vec{E}(T_k(x))} \theta(e)^2/2$, where $x \in \Psi_2^d$, denote the contribution to $\mathcal{E}(\theta)$ by the flow through edges of the tree $T_k(x)$. We now make the following claim to be proven later:

*Let $x \in \Psi_2^d$ and $t$ be the unique vertex in $\Psi_1^d$ connected to $x$ in $H$. Then*

$$\mathcal{E}\left(\theta | T_k(x)\right) \leq k \cdot \theta\left(t, x\right)^2. \tag{3.9}$$

Recall that $\theta\left(t, x\right) = 1/\left(\psi_1^d(w)\varphi_1^d(t)\right)$ where $w \in \{i, j\}$. Thus by (3.9) the contribution to $\mathcal{E}(\theta)$ from the edges of $\bigcup_{x \in \Psi_2^d} T_k(x)$ is then at most

$$\sum_{w \in \{i,j\}} \sum_{\substack{w' \in \\ \Psi_1^d(w)}} \sum_{\substack{w'' \in \\ \Phi_1^d(w')}} \frac{k}{\left(\psi_1^d(w)\varphi_1^d(w')\right)^2} = \sum_{i_a \in \Psi_1^d(i)} \frac{k}{\psi_1^d(i)^2 \varphi_1^d(i_a)} + \sum_{j_b \in \Psi_1^d(j)} \frac{k}{\psi_1^d(j)^2 \varphi_1^d(j_b)}.$$

(v) Finally for any edge $e \in \vec{E}(G) \backslash \vec{E}(H)$ we set $\theta(e) = 0$, this contributes 0 to $\mathcal{E}(\theta)$.

Now we collect the contributions to $\mathcal{E}(\theta)$ from the edges in $E(H)$ in Steps (i)-(v) above to obtain the following bound on $R(i,j)$ for $G \in A_{i,j}^{n,k,d} \cap B_{i,j}$

$$R(i,j) \leq \mathcal{E}(\theta) \leq \frac{1}{\psi_1^d(i)} + \frac{1}{\psi_1^d(j)} + \sum_{i_a \in \Psi_1^d(i)} \frac{k+2}{\psi_1^d(i)^2 \varphi_1^d(i_a)} + \sum_{j_b \in \Psi_1^d(j)} \frac{k+2}{\psi_1^d(j)^2 \varphi_1^d(j_b)}.$$

All that remains is to prove the claim (3.9): Consider the set $S_i$ of edges of $T_k(x)$ with closest endpoint from $x$ at distance $0 \leq i \leq k-1$ from $x$, this edge set separates $x$ from the leaves of $T_k(x)$. The combined flow through $S_i$ is $\theta(t,x)$ since this is the amount of flow entering $T_k(x)$ at $x$ and leaving the tree through its leaves. Thus since the contribution to $\mathcal{E}(\theta)$ by the edges of $S_i$ is the sum of the squares of the flows through each edge of $S_i$ we see that this cannot exceed $\theta(t,x)^2$ by convexity. The result follows by summing the contributions from the $k$ such edge sets $S_i$. □

# Chapter 4

# The strong $k$-path property for $\mathcal{G}(n,p)$

In this chapter we show that for $\mathcal{G}(n,p)$, in a range of $p$ which we call sparsely connected, with high probability the strong $k$-path property holds for a $k$ which is roughly half the diameter of $\mathcal{G}(n,p)$.

## 4.1 Neighbourhood growth bounds

In the previous section we obtained Theorem 3.4.1 which is an upper bound for the effective resistance in a graph with the strong $k$-path property. This bound is by an expression involving the pruned neighbourhoods $\Phi_1$ and $\Psi_1$, defined at (3.5) and (3.6) respectively. To apply this bound we must gain control over the distributions of $\gamma^*$, $\varphi$ and $\psi$.

A key feature of the MBFS algorithm is that the clashing vertices are removed rather than being assigned a unique parent. Though this means we are reducing the sizes of the neighbourhoods, removing clashing vertices in this way ensures that for MBFS on $\mathcal{G} \sim_d \mathcal{G}(n,p)$ the sequence $\{\gamma_1^*(x_i)\}_{i=1}^{|X|}$ for any $X \subseteq I_k$ is exchangeable.

**Lemma 4.1.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, $I_0 := \{u,v\} \subset V$ and $i,k \geq 0$. Run* MBFS$(\mathcal{G}, I_0)$.

*(i) Then $|S_1| \sim_d Bin\left(n-2,(1-p)^2\right)$ and $|I_1| \sim_d Bin\left(n-2, 2p(1-p)\right)$.*

*(ii) Conditioning on $\{x \in I_k\}$ and $|S_k(x)|$, then*

$$\gamma_1^*(x) \sim_d Bin(|S_k(x)|, p).$$

*(iii) Conditioning on* $\{x \in I_k\}, |S_{k+i}|, |I_{k+i}|$ *and* $\gamma_i^*(x)$*, then*

$$\gamma_{i+1}^*(x) \sim_d Bin\left(|S_{k+i}|, \gamma_i^*(x) \cdot p \cdot (1-p)^{|I_{k+i}|-1}\right).$$

*(iv) Let* $x \in V$*. Conditioning on* $\gamma_i(x)$ *and* $|B_i(x)|$*, then*

$$\gamma_{i+1}(x) \sim_d Bin\left(n - |B_i(x)|, 1 - (1-p)^{\gamma_i(x)}\right).$$

*Proof. Item* (i): a vertex in $S_0$ is in $S_1$ if it is not connected to either vertex in $I_0$. This happens independently with probability $(1-p)^2$ for each of the $n-2$ vertices in $S_0$ thus

$$|S_1| \sim_d Bin\left(n-2, (1-p)^2\right).$$

A vertex in $S_0$ is in $I_1$ if it is connected to exactly one vertex in $I_0$. This happens independently with probability $2p(1-p)$ for each of the $n-2$ vertices in $S_0$ thus

$$|I_1| \sim_d Bin\left(n-2, 2p(1-p)\right).$$

*Item* (ii): recall the definitions of $\Gamma_1^*(x)$ and $S_k(x)$ for $x \in I_k$, given by (3.3) and (3.4) respectively. Observe the following relation:

$$\Gamma_1^*(x) = (\Gamma_1(x) \cap S_k) \setminus \bigcup_{\substack{y \in I_k \\ y \neq x}} \Gamma_1(y) = \Gamma_1(x) \cap S_k(x).$$

Since we completely remove the vertices if they clash, and the edges of $\mathcal{G}$ are independent, the order MBFS explores the neighbourhoods of each $y \in I_k$ is unimportant. Assume that we have explored the neighbourhood of every $y \in I_k$ with $y \neq x$. We then know which vertices in the neutral set $S_k$ will not clash if included in $\Gamma_1(x)$ and these are the vertices in $S_k(x)$. Since edges occur independently with probability $p$, conditioning on $|S_k(x)|$ yields

$$\gamma_1^*(x) \sim_d Bin(|S_k(x)|, p).$$

*Item* (iii): for a vertex $v \in S_{k+i}$ we have $v \in \Gamma_{i+1}^*(x)$ when there is exactly one edge $yv \in E(\mathcal{G})$ where $y \in \Gamma_i^*(x)$ and there is no edge of the form $y'v \in E$ where $y' \in I_{k+i}$ and $y' \neq y$. Conditioning on the sizes of $I_{k+i}$ and $\Gamma_i^*(x)$ we see that each $v \in S_{k+i}$ is a member of $\Gamma_{i+1}^*(x)$ with probability $\gamma_i^*(x) \cdot p \cdot (1-p)^{|I_{k+i}|-1}$. These events are independent as each edge occurs independently. Thus, conditioning on $|S_{k+i}|, |I_{k+i}|$

and $\gamma_i^*(x)$, we have

$$\gamma_{i+1}^*(x) \sim_d Bin\left(|S_{k+i}|, \gamma_i^*(x) \cdot p \cdot (1-p)^{|I_{k+i}|-1}\right).$$

*Item* (iv): a vertex in $V \backslash B_i(x)$ is in $\Gamma_{i+1}(x)$ if it is connected to a vertex in $\Gamma_i(x)$. For each $y \in V \backslash B_i(x)$ the probability there is no $yz \in E$ where $z \in \Gamma_i(u)$ is $(1-p)^{\gamma_i(x)}$ and the events are all independent. Thus conditional on $\gamma_i(x)$ and $|B_i(x)|$,

$$\gamma_{i+1}(x) \sim_d Bin\left(n - |B_i(x)|, 1 - (1-p)^{\gamma_i(x)}\right).$$

□

Let $x \in I_k$. Choosing $i = 0$ in Lemma 4.1.1 (iii) gives the following conditional on $|S_k|$ and $|I_k|$

$$\gamma_1^*(x) \sim_d Bin\left(|S_k|, p(1-p)^{|I_k|-1}\right).$$

This appears to differ from the distribution $Bin\left(|S_k(x)|, p\right)$ given by Lemma 4.1.1 (ii). However this is not the case as, conditional on $|S_k|$ and $|I_k|$,

$$|S_k(x)| \sim_d Bin\left(|S_k|, (1-p)^{|I_k|-1}\right).$$

The following two lemmas provide tail estimates for the distributions of $\Gamma_i$ and $\Gamma_i^*$ neighbourhoods, where $i \geq 1$. We prove the Lemmas by induction where the inductive step comes from applying Chernoff bounds to the binomial distributions described in Lemma 4.1.1. For Lemma 4.1.2 this induction shows that with high probability the sequence $\gamma_1(u), \gamma_2(u), \ldots$ is bounded above by the sequence $a_1 np, a_2(np)^2, \ldots$ where the $a_i$ satisfy a recurrence relation. This recurrence can later be solved to give bounds on the sequence $a_i$ based on the exceptional probability desired. This strategy is inspired by [25].

**Lemma 4.1.2** ($\Gamma$-Neighbourhood bounds). *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where $np \to \infty$. Then for $u \in V$ and any $i \leq \log n / \log(np)$, $k > 3$ the following holds*

*(i)* $\mathbb{P}\big(\gamma_i(u) > 2k^2(np)^i\big) = o\left(e^{-3(k-3)np/2}\right)$,

*(ii)* $\mathbb{P}\big(|B_i(u)| > (2k^2+1)(np)^i\big) = o\left(e^{-3(k-3)np/2}\right)$.

*Proof. Item* (i): we wish to show the following by induction on $i \geq 0$

$$\mathbb{P}\left(\bigcup_{j=0}^{i} \{\gamma_j(u) > a_j(np)^j\}\right) \leq \sum_{j=0}^{i} e^{-j} \exp\left(-\frac{\lambda^2}{2(1+\lambda/3\sqrt{np})}\right),$$

where $a_i \geq 0$ is given by the recurrence

$$a_{i+1} = a_i + \frac{\lambda_i \sqrt{a_i}}{(np)^{(i+1)/2}},\ a_0 = 1,$$

and $\lambda_i = \sqrt{3i + \lambda^2}$ for some $\lambda$ specified later. Let $\mathcal{E}_i := \{\gamma_i(u) \leq a_i(np)^i\}$ and $\mathcal{H}_i := \bigcap_{j=0}^{i} \mathcal{E}_i$. Observe that for the base case $\gamma_0(u) = 1 = a_0$. Notice also that

$$a_i(np)^i np + \lambda_i \sqrt{a_i(np)^{i+1}} = (np)^{i+1} \left( a_i + \lambda_i \frac{\sqrt{a_i}}{(np)^{(i+1)/2}} \right) = a_{i+1}(np)^{i+1}. \tag{4.1}$$

Conditional on $\gamma_i(u)$ we have $\gamma_{i+1}(u) \preceq_1 Bin\,(\gamma_i(u) \cdot n, p)$. Thus by (4.1) above

$$\begin{aligned} \mathbb{P}((\mathcal{E}_{i+1})^c \cap \mathcal{H}_i) &:= \mathbb{P}\big(\{\gamma_{i+1}(u) > a_{i+1}(np)^{i+1}\} \cap \mathcal{H}_i\big) \\ &\leq \mathbb{E}\left[ \mathbb{P}\left( Bin\,(\gamma_i(u) \cdot n, p) > a_i(np)^i np + \lambda_i \sqrt{a_i(np)^{i+1}} \,\middle|\, \gamma_i(u) \right) \mathbf{1}_{\mathcal{H}_i} \right]. \end{aligned}$$

Now by the Chernoff bounds, Lemma 2.1.1, we have

$$\begin{aligned} \mathbb{P}((\mathcal{E}_{i+1})^c \cap \mathcal{H}_i) &\leq \mathbb{E}\left[ \exp\left( - \frac{\lambda_i^2 a_i (np)^{i+1}}{2\left( a_i(np)^{i+1} + \lambda_i \sqrt{a_i(np)^{i+1}}/3 \right)} \right) \mathbf{1}_{\mathcal{H}_i} \right] \\ &= \mathbb{E}\left[ \exp\left( - \frac{3i + \lambda^2}{2\left( 1 + \lambda_i/3\sqrt{a_i(np)^{i+1}} \right)} \right) \mathbf{1}_{\mathcal{H}_i} \right] \end{aligned}$$

Since $a_i \geq 1$ and $np = \omega(1)$ we have $\lambda_i/\sqrt{a_i(np)^{i+1}} \leq \lambda/\sqrt{np}$. Thus

$$\mathbb{P}((\mathcal{E}_{i+1})^c \cap \mathcal{H}_i) \leq e^{-i} \exp\left( - \frac{\lambda^2}{2\left(1 + \lambda/3\sqrt{np}\right)} \right) \mathbb{P}(\mathcal{H}_i)\,, \tag{4.2}$$

for $n$ large enough. Now observe that $\mathcal{H}_{i+1} \subseteq \mathcal{H}_i$ and $\mathcal{H}_i$ is the disjoint union of $\mathcal{H}_{i+1}$ and $(\mathcal{E}_{i+1})^c \cap \mathcal{H}_i$. Hence we have the following reduction by (4.2)

$$\mathbb{P}(\mathcal{H}_{i+1}) = \mathbb{P}(\mathcal{H}_i) - \mathbb{P}((\mathcal{E}_{i+1})^c \cap \mathcal{H}_i) = \left( 1 - e^{-i} \exp\left( - \frac{\lambda^2}{2\left(1 + \lambda/3\sqrt{np}\right)} \right) \right) \mathbb{P}(\mathcal{H}_i)\,.$$

If we continue this iteratively we have the following

$$\mathbb{P}(\mathcal{H}_{i+1}) = \prod_{j=0}^{i} \left( 1 - e^{-j} \exp\left( - \frac{\lambda^2}{2\left(1 + \lambda/3\sqrt{np}\right)} \right) \right) \mathbb{P}(\mathcal{H}_0)\,.$$

If we recall that $\mathcal{H}_0 = \{\gamma_0(u) \le 1\}$ so $\mathbb{P}(\mathcal{H}_0) = 1$. Thus we have

$$\mathbb{P}(\mathcal{H}_{i+1}) \ge 1 - \sum_{j=0}^{i} e^{-j} \exp\left( -\frac{\lambda^2}{2\left(1 + \lambda/3\sqrt{np}\right)} \right).$$

Let $\lambda = k\sqrt{np}$ for any $k \ge 0$ and observe that

$$\frac{k^2}{2(1+k/3)} = \frac{3k}{2} - \frac{3k}{2(1+k/3)} = \frac{3k}{2} - \frac{9}{2(1+3/k)} > \frac{3(k-3)}{2}. \tag{4.3}$$

Then, by (4.3) above and since $np = \omega(1)$, we obtain

$$\begin{aligned}\mathbb{P}((\mathcal{H}_i)^c) &\le \sum_{j=0}^{i} e^{-j} \exp\left( -\frac{\lambda^2}{2(1+\lambda/3\sqrt{np})} \right) \\ &= \frac{e}{e-1} \exp\left( -\frac{k^2 np}{2(1+k/3)} \right) \\ &= o\left( e^{-3(k-3)np/2} \right). \end{aligned} \tag{4.4}$$

By (4.4) above it makes sense to consider $k > 3$, thus we will show that $a_i \le 2k^2$ for all $i$ and any $k > 3$. Since $a_0 = 1 \le 2k^2$ assume $a_i \le 2k^2$, then by (4.1) we have

$$a_{i+1} = a_i + \frac{\lambda_i\sqrt{a_i}}{(np)^{(i+1)/2}} = 1 + \frac{\lambda_0\sqrt{a_0}}{\sqrt{np}} + \sum_{j=1}^{i} \frac{\lambda_j\sqrt{a_j}}{(np)^{(j+1)/2}}.$$

Recall that $\lambda_i = \sqrt{3i + \lambda^2}$ and observe that $\lambda_0 = \lambda = k\sqrt{np}$. Thus we have

$$\begin{aligned} a_{i+1} &= 1 + k + \sum_{j=1}^{i} \frac{\sqrt{3j + k^2 np}\sqrt{2k^2}}{(np)^{(j+1)/2}} \\ &\le 1 + k + \sqrt{2k^2} \sum_{j=1}^{i} \frac{3j/np + k^2}{(np)^{j/2}} \\ &\le 1 + k + O\Big((np)^{-1/2}\Big) \\ &\le 2k^2. \end{aligned}$$

Thus is follows by (4.4) that for any $k > 3$ we have

$$\mathbb{P}\big(\gamma_i(u) > 2k^2(np)^i\big) \le \mathbb{P}((\mathcal{H}_i)^c) = o\left(e^{-3(k-3)np/2}\right).$$

*Item* (ii): Observe that conditional on $\bigcap_{j=0}^{i}\{\gamma_j(u) \le 2k^2(np)^i\} \subseteq \mathcal{H}_i$ we have

$$|B_i(u)| = \sum_{j=0}^{i}\gamma_j(u) \le \sum_{j=0}^{i} 2k^2(np)^i \le (2k^2+1)(np)^i.$$

The result follows since $\mathbb{P}((\mathcal{H}_i)^c) = o\left(e^{-3(k-3)np/2}\right)$ by (4.4). □

Notice that what we have actually proved above is stronger than the statement of Lemma 4.1.2 and that the term $2k^2$ can be improved easily to something $O(k)$. We have stated the Lemma as it is for backwards compatibility.

We now prove some lower bounds on growth for the pruned neighbourhoods, the proof is similar to that of Lemma 4.1.2 however slightly more involved as the distribution of the $\Gamma^*$ neighbourhoods is more complicated.

**Lemma 4.1.3** ($\Gamma^*$-Neighbourhood lower bounds). *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ and $i \in \mathbb{Z}$ satisfy*

$$1 \le i \le \lfloor \log(n)/\log(np)\rfloor - 3. \tag{4.5}$$

*Let $\Psi_2^d$ be defined with respect to* $\mathrm{MBFS}(\mathcal{G},\{u,v\})$ *for $u,v \in V$ and $d \ge 0$.*

*(i) Let $c > 0$, $np \ge c\log n$ and $d \ge \max\left\{\lceil \frac{50}{c}\rceil, 50\right\}$. Then*

$$\mathbb{P}\Big(\gamma_i^*(y) < 15(np)^{i-1}\Big|\, y \in \Psi_2^d\Big) = o\left(e^{-4np}\right).$$

*(ii) If $np = \omega(\log n)$ then for any fixed $d, K \ge 0$*

$$\mathbb{P}\Big(\gamma_i^*(y) < \tfrac{9}{10}(np)^{i}\Big|\, y \in \Psi_2^d\Big) = o\left(n^{-K}\right).$$

*(iii) If $np - \log n \to \infty$ then for any $5 \le i \le \lfloor \log(n)/\log(np)\rfloor - 5$*

$$\mathbb{P}_{\mathcal{C}}\left(|B_i(v)| < 15(np)^{i-5}\right) = o\left(n^{-4}\right).$$

*Proof.* We will first set up the general framework for a neighbourhood growth bound and then apply this bound under different conditions to prove Items (i), (ii) and (iii).

Run $\mathrm{MBFS}(\mathcal{G},\{u,v\})$ and let $y \in I_h$, $n_i := |S_{i+h}|$, $p_i := p\cdot(1-p)^{|I_{i+h}|-1}$ and $r_i = \prod_{j=i_0}^{i} n_j p_j$. We wish to show that there exists some $i_0 \in \mathbb{Z}$, $i_0 \ge 0$ such that for all $i \ge i_0$:

$$\mathbb{P}\big(\gamma_{i+1}^*(y) < a_{i+1}r_i\big) \le (i+1)\exp\left(-\lambda^2/2\right), \tag{4.6}$$

where $a_i \geq 0$ satisfies $a_{i+1} = a_i - \lambda\sqrt{a_i}/\sqrt{r_i}$, for some initial $a_{i_0}$ we will find later. Observe

$$a_i r_{i-1} n_i p_i - \lambda\sqrt{a_i r_{i-1} n_i p_i} = \left(a_i - \lambda\frac{\sqrt{a_i}}{\sqrt{r_i}}\right) r_i = a_{i+1} r_i.$$

Applying Lemma 4.1.1 (iii) and conditioning on $\mathfrak{F}_{i+h}$ yields

$$\gamma^*_{i+1}(y) \sim_d Bin\left(n_i, \gamma^*_i(y) p_i\right).$$

Let $\mathcal{H}_i := \{\gamma^*_i(u) \geq a_i r_{i-1}\} \in \mathfrak{F}_{i+h}$. Now by Lemma 2.1.1 (i) and the inductive hypothesis, this is the bound $\mathbb{P}(\gamma^*_i(y) < a_i r_{i-1}) \leq i \exp\left(-\lambda^2/2\right)$, we have the following

$$\begin{aligned}
\mathbb{P}\big(\gamma^*_{i+1}(y) < a_{i+1} r_i\big) &= \mathbb{E}\big[\mathbb{P}\big(\gamma^*_{i+1}(y) < a_i r_{i-1} n_i p_i - \lambda\sqrt{a_i r_{i-1} n_i p_i}\big|\mathfrak{F}_{i+h}\big)\big] \\
&\leq \mathbb{E}\big[\mathbb{P}\big(Bin\left(n_i, \gamma^*_i(y)p_i\right) < a_i r_i - \lambda\sqrt{a_i r_i}\big|\mathfrak{F}_{i+h}\big)\,\mathbf{1}_{\mathcal{H}_i}\big] + \mathbb{P}(\mathcal{H}^c_i) \\
&\leq \exp\left(-\lambda^2 a_i r_i/(2a_i r_i)\right) + i\exp\left(-\lambda^2/2\right) \\
&= (i+1)\exp\left(-\lambda^2/2\right).
\end{aligned}$$

The above always holds, however it may be vacuous as if $i$ is too large then $a_i$ may be negative. This can also happen for an incorrect choice of the starting time $i_0$ and initial value $a_{i_0}$. We address this in the application making sure to condition on events where everything is well defined. In this spirit let $l := \lfloor \log(n)/\log(np) \rfloor - h - 1$ and define the event

$$\mathcal{D} := \bigcap_{i=0}^{l} \left\{|I_{i+h}| \leq 144(np)^{i+h}\right\} \cap \left\{\gamma^*_i(y) \leq 72(np)^i\right\} \cap \left\{|S_{i+h}| \geq n - 146(np)^{i+h}\right\}.$$

Conditioning on the event $\mathcal{D}$ and the filtration $\mathfrak{F}_{i+h}$ for any $i \leq l$ ensures that $Bin\left(n_i, \gamma^*_i(y)p_i\right)$ is a valid probability distribution and $n_i p_i = (1-o(1))np$. By Lemma 4.1.2 with $k = 6$,

$$\begin{aligned}
\mathbb{P}(\mathcal{D}^c) &\leq 2\sum_{i=0}^{l}\mathbb{P}\big(B_{i+2}(u) > 73(np)^{i+2}\big) + 2\sum_{i=0}^{l}\mathbb{P}\big(\Gamma(u)_{i+2}(u) > 72(np)^{i+2}\big) \\
&\quad + \sum_{i=0}^{l}\mathbb{P}\big(\Gamma_i(u) > 72(np)^i\big) \\
&= o\left(\exp\left(-4np\right)\right). \qquad (4.7)
\end{aligned}$$

*Item* (i): recall from (3.7) that if $y \in \Psi^d_2(u) \cup \Psi^d_2(v) \subseteq I_2$ then $\gamma^*_1(y) > d$ for any fixed $d$. Conditional on $\mathcal{D} \cap \mathfrak{F}_3$, $\gamma^*_2(y) \succeq_1 Bin\left(n(1-\varepsilon), dp(1-\varepsilon)\right)$ for any fixed

$1 > \varepsilon > 0$ provided $n$ is large enough. If we choose $\lambda = 3\sqrt{np}$, $d \geq \max\left\{\lceil \frac{50}{c} \rceil, 50\right\}$ then Lemma 2.1.1 (i) yields

$$\begin{aligned}
\mathbb{P}(\gamma_2^*(y) < dn_1p_1/2) &= \mathbb{E}\big[\mathbb{P}\left(\gamma_2^*(y) < dn_1p_1/2 \big| \mathfrak{F}_3\right)(\mathbf{1}_{\mathcal{D}} + \mathbf{1}_{\mathcal{D}^c})\big] \\
&\leq e^{-dnp/10} + \mathbb{P}(\mathcal{D}^c) \\
&\leq e^{-\lambda^2/2}.
\end{aligned}$$

Take $i_0 = 1$ and $a_2 = d/3$ since on $\mathcal{D}$ we have $d/2n_1p_1 \geq dnp/3$. Now $a_2 \geq \cdots \geq a_i$ so on the event $\mathcal{D}$ we have the following for any $\varepsilon > 0$ and $3 \leq i \leq \lfloor \log(n)/\log(np) \rfloor - 3$

$$a_i = a_2 - \sum_{k=2}^{i-1} \frac{\lambda\sqrt{a_k}}{\sqrt{r_k}} \geq \frac{d}{3} - (3+\varepsilon)\sqrt{\frac{d}{3np}} \geq 16,$$

since conditional on $\mathcal{D}$ we have $r_i = \prod_{j=i_0}^{i} n_jp_j \geq (1-\varepsilon)(np)^i$ for any $\varepsilon > 0$ and $1 \leq i \leq \lfloor \log(n)/\log(np) \rfloor - 3$ when $n$ is large. Notice also that $\gamma_1^*(y) > d > 15(np)^0$ so by (4.6)

$$\begin{aligned}
\mathbb{P}\big(\gamma_{i+1}^*(y) < 15(np)^i\big) &\leq \mathbb{P}\big(\gamma_{i+1}^*(y) < a_{i+1}r_i\big) + \mathbb{P}(\mathcal{D}^c) \\
&\leq (i+1)e^{-\lambda^2/2} + o\left(e^{-9np/2}\right) \leq e^{-4np}.
\end{aligned}$$

*Item* (ii): in this case on the event $\mathcal{D}$ we have $n_ip_i = (1-o(1))np = \omega(\log n)$ for every $0 \leq i \leq l$, so we do not need to rely on the fact that $\gamma_1^*(y) \geq d$ to start the branching, thus the following holds for any $d \geq 0$. Let $\lambda = \sqrt{3K\log n}$ where $K > 0$ is any fixed constant. As before conditioning on $\mathcal{D} \cap \mathfrak{F}_3$ ensures that for any fixed $1 > \varepsilon > 0$ we have $\gamma_1^*(y) \sim_d Bin\,(n_0, p_0) \succeq_1 Bin\,(n(1-\varepsilon), p(1-\varepsilon))$ when $n$ is large enough. By Lemma 2.1.1 (i),

$$\begin{aligned}
\mathbb{E}\big[\mathbb{P}\big(\gamma_1^*(y) < r_0 - (5/4)\lambda\sqrt{r_0}\big|\mathfrak{F}_3\big)(\mathbf{1}_{\mathcal{D}} + \mathbf{1}_{\mathcal{D}^c})\big] &\leq e^{-25\lambda^2/32} + \mathbb{P}(\mathcal{D}^c) \\
&\leq \exp\left(-\lambda^2/2\right).
\end{aligned}$$

Take $i_0 = 0$, $a_1 = 19/20$ since on $\mathcal{D}$ we have $r_0 - (5/4)\lambda\sqrt{r_0} \geq 19np/20$. Now $a_1 \geq \cdots \geq a_i$ so on the event $\mathcal{D}$ we have the following for any $\varepsilon > 0$ and $2 \leq i \leq \lfloor \log(n)/\log(np) \rfloor - 3$

$$a_i = a_1 - \sum_{k=1}^{i-1} \frac{\lambda\sqrt{a_k}}{\sqrt{r_k}} \geq \frac{19}{20} - (1+\varepsilon)\frac{\sqrt{19 \cdot 3K\log n}}{\sqrt{20np}} = \frac{19}{20} - o(1) \geq \frac{9}{10}.$$

Thus for any $1 \le i \le \lfloor \log(n)/\log(np) \rfloor - 3,\ K > 0$ we have

$$\begin{aligned}\mathbb{P}\big(\gamma_i^*(y) < 9/10(np)^i\big) &\le \mathbb{P}(\gamma_i^*(y) < a_i r_{i-1}) + \mathbb{P}(\mathcal{D}^c)\\ &\le (i+1)e^{-\lambda^2/2} + e^{-9np/2} \le o\left(n^{-K}\right).\end{aligned}$$

*Item* (iii): since $\mathcal{G} \in \mathcal{C}$ there exists a path $u := u_0, u_1, \ldots, u_l$ with $u_{j-1}u_j \in E$ for each $1 \le j \le l$. Let $f(u_j)$ be the number of $u_j v \in E$ where $v \in V \backslash \{u_0, \ldots, u_l\}$. Then for any fixed $d$

$$\begin{aligned}\mathbb{P}(f(u_j) < d) &= \sum_{i=1}^{d-1} \binom{n-l-1}{i} p^i (1-p)^{n-i-l-1}\\ &\le \sum_{i=1}^{d-1} \frac{(np)^i}{i!} e^{-(n-d-l)p}\\ &\le (np)^d e^{1-(1-(d-l)/n)np}\\ &= e^{d\log(np)-(1-(d-l)/n)np}\\ &= e^{-(1-o(1))np}.\end{aligned}$$

Let $\mathcal{E}$ be the event $\{\gamma_1(u_{j_0}) \ge d$ for some $0 \le j_0 \le 4\}$. As $np \ge \log n$ and $\{f(u_j)\}_{j=0}^{l}$ are i.i.d.:

$$\mathbb{P}_{\mathcal{C}}(\mathcal{E}^c) \le \mathbb{P}(f(u_j) < d)^5 / \mathbb{P}(\mathcal{C}) \le e^{-5(1-o(1))\log n} \le o\left(n^{-4}\right). \tag{4.8}$$

On $\mathcal{E}$ there is some $u_{j_0} \in V$ with $d(u, u_{j_0}) = j_0 \le 4$ and $\gamma_1(u_{j_0}) > d$. We use the stochastic domination $\gamma_i(u_{j_0}) \succeq_1 \gamma_i^*(u_{j_0})$ to bound the growth of $|B_{i+j_0}(u)|$ from below by that of $\gamma_i^*(u_{j_0})$. Here we consider $u_{j_0} \in I_{j_0}$ defined with respect to $\mathrm{MBFS}(\mathcal{G}, \{u, v\})$ for some $v \in V$. Let $\lambda = 3\sqrt{\log n},\ d \ge \max\left\{\lceil \frac{50}{c} \rceil, 50\right\}$. On $\mathcal{D}$, $r_{j_0+1} \ge .99np$ when $n$ is large. By Lemma 2.1.1 (i):

$$\mathbb{E}\big[\mathbb{P}\left(\gamma_{j_0+2}^*(u) < dn_{j_0+1}p_{j_0+1}/2 \middle| \mathfrak{F}_{j_0+1}\right) \mathbf{1}_{\mathcal{D}\cap\mathcal{E}}\big] \le \mathbb{E}\Big[e^{-dr_{j_0+1}/8}\mathbf{1}_{\mathcal{D}\cap\mathcal{E}}\Big] \le e^{-\lambda^2/2}.$$

Take $i_0 = j_0 + 1$ and $a_{j_0+2} = d/3$ since on $\mathcal{D} \cap \mathcal{E}$ we have $dn_{j_0+1}p_{j_0+1}/2 \ge dnp/3$. Now $a_{i_0} \ge \cdots \ge a_i$ and on the event $\mathcal{D} \cap \mathcal{E}$ we have $r_i = (1 - o(1))(np)^{i-j_0}$. Thus we have the following for any $\varepsilon > 0$ and $j_0 + 3 \le i \le \lfloor \log(n)/\log(np) \rfloor - j_0 - 1$:

$$a_i = a_{j_0+2} - \sum_{k=j_0+2}^{i-1} \frac{\lambda\sqrt{a_k}}{\sqrt{r_k}} \ge \frac{d}{3} - (3+\varepsilon)\sqrt{\frac{d\log n}{3(np)^2}} \ge 16.$$

Notice also $\gamma^*_{j_0+1}(y) > d > 15(np)^0$. Thus for any $4 \le i \le \lfloor \log(n)/\log(np) \rfloor - 5$:

$$\begin{aligned}
&\mathbb{P}_{\mathcal{C}}\big(|B_{i+1}(u)| < 15(np)^{i-5}\big) \\
&\qquad \le \mathbb{P}_{\mathcal{C}}\big(\gamma_{i+1}(u) < 15(np)^{i-j_0-1}\big|\mathcal{D},\mathcal{E}\big) + \mathbb{P}_{\mathcal{C}}(\mathcal{D}^c) + \mathbb{P}_{\mathcal{C}}(\mathcal{E}^c) \\
&\qquad \le \mathbb{P}\big(\gamma^*_{i+1}(y) < a_{i+1}r_i\big|\mathcal{D},\mathcal{E}\big)\,/\mathbb{P}(\mathcal{C}) + \mathbb{P}(\mathcal{D}^c)\,/\mathbb{P}(\mathcal{C}) + \mathbb{P}_{\mathcal{C}}(\mathcal{E}^c)\,.
\end{aligned}$$

By the bounds on $\mathbb{P}(\mathcal{C})$, $\mathbb{P}(\mathcal{D})$ and $\mathbb{P}(\mathcal{E}^c)$ by (2.15), (4.7) and (4.8) respectively

$$\mathbb{P}_{\mathcal{C}}\big(|B_{i+1}(u)| < 15(np)^{i-5}\big) \le 2(i+1)e^{-\lambda^2/2} + o\left(e^{-4np}\right) + o\left(n^{-4}\right) = o\left(n^{-4}\right).$$

□

## 4.2 The strong $k$-path property for $\mathcal{G}(n,p)$

Recall $A^{n,k,d}_{u,v}$ is the set of graphs on $[n]$ satisfying the strong $k$-path property for $u, v \in [n]$, see Definition 3.3.1. Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ and define the function

$$D = \begin{cases} \max\left\{\lceil \frac{50}{c} \rceil, 50\right\} & \text{if } np = (c \pm o(1))\log(n) \text{ where } c > 0 \\ 0 & \text{if } np = \omega\left(\log(n)\right) \end{cases}. \tag{4.9}$$

The significance of $D$ to the pruned neighbourhoods should be evident from the statement of Lemma 4.1.3. From now on whenever we denote any of the pruned neighbourhoods $\Phi_1(u), \Psi_1(u), \phi_1(u), \varphi_1(u), \dots$ without the index $d$ we take this to mean $d = D$, for example $\varphi_2(u) := \varphi_2^D(u)$, where $D$ is given by (4.9). Define the following event:

$$\mathcal{A}^n_{u,v} := \{ \text{ exists } k \le \log n/(2\log np) + 2 \text{ such that } \mathcal{G} \in A^{n,k,D}_{u,v}\}. \tag{4.10}$$

Recall the definition (3.8) of $B^{u,v,d}_w := \left\{G : V = [n], \Psi^d_1(w) \neq \emptyset\right\}$ for $w \in \{u, v\} \subset V$ and also the intersection $B^d_{u,v} := B^{u,v}_u \cap B^{u,v,d}_v$. For the same $D$ we define

$$\mathcal{B}^{u,v}_w := \left\{\mathcal{G} \in B^{u,v,D}_w\right\}, \qquad \mathcal{B}_{u,v} = \mathcal{B}^{u,v}_u \cap \mathcal{B}^{u,v}_v. \tag{4.11}$$

We are now in a position to show that the strong $k$-path property holds in sparsely connected binomial random graph with high probability.

**Lemma 4.2.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where for any $c > 0$, $c\log n \le np < n^{1/10}$. Then* *for* $u, v \in V$, $u \neq v$

$$\mathbb{P}\big((\mathcal{A}^n_{u,v})^c\big) = o\left(e^{-7\min\{np,\log n\}/2}\right) \qquad \textit{and} \qquad \mathbb{P}((\mathcal{B}_{u,v})^c) = e^{-(1-o(1))np}.$$

*Proof.* Run MBFS$(\mathcal{G}, \{u, v\})$, $u, v \in V$. For $k \geq 0$ let $\mathcal{T} := \mathcal{T}_{u,v,k} = \mathcal{T}_1 \cap \mathcal{T}_2$ where

$$\mathcal{T}_1 := \left\{|S_{k+2}| \geq n - n^{5/6}\right\},$$

and

$$\mathcal{T}_2 := \{|\Gamma^*_k(x) \times \Gamma^*_k(y)| \geq 4n, \text{for all } (x, y) \in \Psi_2(u) \times \Psi_2(v)\}.$$

On the event $\mathcal{T}_1$ when MBFS$(\mathcal{G}, \{u, v\})$ has run for $k + 2$ iterations there is still a lot of the graph yet to explore and the algorithm will run for at least one more iteration. The $k$ in the definition of $\mathcal{T}$ will be the one occurring in the description of $\mathcal{A}^n_{u,v}$. Set the value of $k$ to be

$$k := k(n, p) = \begin{cases} \left\lceil \log\left(\frac{4n}{(15)^2}\right) / \left(2\log(np)\right)\right\rceil + 1 & \text{if } np = c\log n \text{ where } c > 0 \\ \left\lceil \log\left(\frac{400n}{81}\right) / \left(2\log(np)\right)\right\rceil & \text{if } np = \omega\left(\log n\right), \end{cases} \tag{4.12}$$

and notice $k \leq \log(np)/\left(2\log n\right) + 2$ for large $n$. It remains to show that for this $k$, given by (4.12), we have

$$\mathbb{P}\Big(\mathcal{G} \notin A^{n,k,D}_{u,v}\Big) = o\left(e^{-7\min\{np,\log n\}/2}\right).$$

Let $\mathcal{R} := \mathcal{R}_{u,v}$ be the event $\left\{|\Psi_2(u) \times \Psi_2(v)| \leq (72(np)^2)^2\right\}$. Since $\psi^d_2(u) \leq \gamma_2(u)$ for any $d \geq 0, u \in V$ an application of Lemma 4.1.2 with $k = 6$ yields

$$\mathbb{P}(\mathcal{R}^c) \leq \mathbb{P}\big(\psi_2(u) > 72(np)^2\big) + \mathbb{P}\big(\psi_2(v) > 72(np)^2\big) = o\left(e^{-9np/2}\right). \tag{4.13}$$

We have the following by the tower property and the bound (4.13) for $\mathbb{P}(\mathcal{R}^c)$

$$\begin{aligned} \mathbb{P}(\mathcal{T}^c) &\leq \mathbb{E}\big[\mathbb{P}\big(\mathcal{T}^c_2 \big| \mathfrak{F}_3\big)\, \mathbf{1}_{\mathcal{R}}\big] + \mathbb{P}(\mathcal{T}^c_1) + \mathbb{P}(\mathcal{R}^c) \\ &\leq 2\mathbb{E}\Big[\psi_2(u)\psi_2(v)\mathbf{1}_{\mathcal{R}}\mathbb{P}\Big(\gamma^*_k(w) < 2n^{1/2}\big|\{w \in \Psi_2\}, \mathfrak{F}_3\Big)\Big] \\ &\quad + 2\mathbb{P}\Big(\gamma_{k+2}(u) > n^{5/6}/2\Big) + o\left(e^{-9np/2}\right), \end{aligned}$$

as $\{\gamma_1^*(x)\gamma_1^*(y) < k\} \subseteq \{\gamma_1^*(x) \text{ or } \gamma_1^*(y) < \sqrt{k}\}$. Observe that provided $np \leq n^{1/10}$ the choice of $k$ given by (4.12) satisfies (4.5) in Lemma 4.1.3. Thus by Lemmas 4.1.2, 4.1.3 (i) and 4.1.3 (ii) we have

$$\begin{aligned}\mathbb{P}(\mathcal{T}^c) &\leq 2(72(np)^2)^2\mathbb{P}\Big(\gamma_k^*(w) < 2n^{1/2}\big|w \in \Psi_2\Big) + 3 \cdot o\left(e^{-9np/2}\right) \\ &\leq e^{5\log(np) - 4\min\{np,\log n\}} + o\left(e^{-9np/2}\right) \\ &= o\left(e^{-7\min\{np,\log n\}/2}\right). \qquad (4.14)\end{aligned}$$

The bound $\mathbb{P}\big(\gamma_k^*(w) < 2n^{1/2}\big|w \in \Psi_2\big) \leq e^{-4\min\{np,\log n\}}$ comes from an amalgamation of Lemmas 4.1.3 (i) and 4.1.3 (ii), where we have chosen $K = 4$ for Lemma 4.1.3 (ii). This is so we can cover the different values of $np$ with one bound. Recall that $\Psi_2 = \Psi_2^D$, the use of this bound is valid since $D$, given by (4.9), satisfies the assumptions on $d$ in the statement of Lemma 4.1.3.

Let $\mathcal{L}_{x,y}$ be the following event indexed by $(x, y) \in \Psi_2(u) \times \Psi_2(v)$,

$$\mathcal{L}_{x,y} := \{x'y' \notin E, \text{for every pair } (x', y') \in \Gamma_k^*(x) \times \Gamma_k^*(y)\}.$$

This is independent of $\mathfrak{F}_{k+2}$ as each $x'y'$ has not been checked up to iteration $k+2$, thus

$$\begin{aligned}\mathbb{P}\big(\mathcal{L}_{x,y}\big|\mathfrak{F}_{k+2}\big)\,\mathbf{1}_{\mathcal{T}} &= \mathbb{P}\big(x'y' \notin E\big)^{\gamma_k^*(x)\gamma_k^*(y)}\,\mathbf{1}_{\mathcal{T}} \\ &\leq (1-p)^{4n}\mathbf{1}_{\mathcal{T}} \\ &\leq 2\exp\left(-4np\right)\mathbf{1}_{\mathcal{T}}. \qquad (4.15)\end{aligned}$$

Recall Definition 3.3.1 of the strong $k$-path property $A_{u,v}^{n,k,D}$ and observe

$$\left\{\mathcal{G} \notin A_{u,v}^{n,k,D}\right\} = \bigcup_{(x,y)\in\Psi_2(u)\times\Psi_2(v)} \{\Gamma_k^*(x) = \emptyset\} \cup \{\Gamma_k^*(y) = \emptyset\} \cup \mathcal{L}_{x,y}.$$

Observe that for each $i, j \geq 0$ the random variables $\{\gamma_j^*(w)\}_{w\in I_i}$ are identically distributed. Recall also that $\Psi_1(u), \Psi_1(v), \mathcal{R} \in \mathfrak{F}_3$. Now by the union bound, tower property and since $\psi_1(u)\psi_1(v) \le (72(np)^2)^2$ on $\mathcal{R}$, we have the following

$$
\begin{aligned}
\mathfrak{P} &:= \mathbb{P}\Big(\Big\{\mathcal{G} \notin A_{u,v}^{n,k,D}\Big\} \cap \mathcal{R} \cap \mathcal{T}\Big) \\
&\le \mathbb{E}\left[\sum_{(x,y)\in\Psi_2(u)\times\Psi_2(v)} \mathbb{E}\Big[\Big(\mathbf{1}_{\mathcal{L}_{x,y}\cup\{\gamma_k^*(x)=0\}\cup\{\gamma_k^*(y)=0\}}\Big)\mathbf{1}_{\mathcal{R}}\mathbf{1}_{\mathcal{T}}\Big|\mathfrak{F}_3\Big]\right] \\
&\le \mathbb{E}\Big[(72(np)^2)^2\mathbf{1}_{\mathcal{R}}\mathbb{E}\Big[\Big(\mathbf{1}_{\mathcal{L}_{x,y}} + \mathbf{1}_{\{\gamma_k^*(x)=0\}} + \mathbf{1}_{\{\gamma_k^*(y)=0\}}\Big)\mathbf{1}_{\mathcal{T}}\Big|\mathfrak{F}_3\Big]\Big].
\end{aligned}
$$

Now since $x, y \in \Psi_2$ and $\gamma_j^*(x), \gamma_j^*(y)$ are identically distributed for any $j \ge 0$:

$$
\mathfrak{P} \le (72(np)^2)^2\left(\mathbb{E}\Big[\mathbb{E}\Big[\mathbf{1}_{\mathcal{L}_{x,y}}\mathbf{1}_{\mathcal{T}}\Big|\mathfrak{F}_{k+2}\Big]\Big] + 2\mathbb{P}\Big(\gamma_k^*(w) = 0\Big|\{w \in \Psi_2\}, \mathfrak{F}_3\Big)\right).
$$

By Lemma (4.1.3) (i), (4.15) and since $\mathcal{T} \in \mathfrak{F}_{k+2}$ we have

$$
\begin{aligned}
\mathfrak{P} &\le (72(np)^2)^2\left(\mathbb{E}\Big[\mathbb{P}\Big(\mathcal{L}_{x,y}\Big|\mathfrak{F}_{k+2}\Big)\mathbf{1}_{\mathcal{T}}\Big] + 2e^{-4\min\{np,\log n\}}\right) \\
&= o\left(e^{-7\min\{np,\log n\}/2}\right).
\end{aligned}
$$

By (4.13), (4.14) and the bound on $\mathbb{P}\Big(\Big\{\mathcal{G} \notin A_{u,v}^{n,k,D}\Big\} \cap \mathcal{R} \cap \mathcal{T}\Big)$ directly above:

$$
\begin{aligned}
\mathbb{P}\Big(\mathcal{G} \notin A_{u,v}^{n,k,D}\Big) &\le \mathbb{P}\Big(\Big\{\mathcal{G} \notin A_{u,v}^{n,k,D}\Big\} \cap \mathcal{R} \cap \mathcal{T}\Big) + \mathbb{P}((\mathcal{R} \cap \mathcal{T})^c) \\
&\le o\left(e^{-7\min\{np,\log n\}/2}\right).
\end{aligned}
$$

For $\mathbb{P}((\mathcal{B}_{u,v})^c)$, use Lemma 2.1.2 to bound the difference between the $\psi$ and $\gamma^*$-distributions:

$$
\begin{aligned}
\mathbb{P}((\mathcal{B}_{u,v})^c) &\le \mathbb{P}(\psi_1(u) = 0) + \mathbb{P}(\psi_1(v) = 0) \\
&\le 2\mathbb{P}(\gamma_1^*(u) = 0) + 2\mathbb{P}(\psi_1(u) \ne \gamma_1^*(u)).
\end{aligned}
$$

Then since $\mathbb{P}(\psi_1(u) \ne \gamma_1^*(u))$ is known by Lemma 4.3.1 we have

$$
\mathbb{P}((\mathcal{B}_{u,v})^c) \le 2\mathbb{P}(\gamma_1^*(u) = 0 \mid \gamma_1^*(v) \le 32np) + 2\mathbb{P}(\gamma_1(v) > 32np) + 2e^{-(1-o(1))np}.
$$

Applying Lemma 4.1.1 (ii) to the first term and Lemma 4.1.2 (i) with $k = 4$ to the second:

$$
\mathbb{P}((\mathcal{B}_{u,v})^c) \le 2(1-p)^{n-32np-1} + o\left(e^{-3(4-1)np/2}\right) + e^{-(1-o(1))np} = e^{-(1-o(1))np}.
$$

□

The event $\mathcal{A}^n_{u,v}$ is generally applied in combination with $\mathcal{B}_{u,v}$ as $\mathcal{G} \in A^{n,k,D}_{u,v}$ can be meaningless if $\mathcal{G} \notin B^D_{u,v}$. However, we have defined the event $\mathcal{B}_{u,v}$ and $\mathcal{A}^n_{u,v}$ separately as sometimes it is necessary to condition on something stronger than the event $\mathcal{B}_{u,v}$. The bound on $R(u,v)$ for $G \in A^{n,k,D}_{u,v}$, Theorem 3.4.1, is sensitive to the $\Psi$-neighbourhoods being empty and so we will also need the following crude but resilient bound on effective resistance in connected $\mathcal{G}(n,p)$ when calculating errors.

**Lemma 4.2.2.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ be such that $np - \log n \to \infty$. Then for $i,j \in V$,*

$$\mathbb{P}_{\mathcal{C}}(R(i,j) > 3\log n/\log(np)) = o(n^{-4}).$$

*Proof.* Since $G \in \mathcal{C}$ the effective resistance between two points is bounded from above by the graph distance. Let $\mathcal{J}_{i,j} := \{|B_k(i)| \cdot |B_k(j)| \geq 4n\}$ where

$$k := k(n,p) = \left\lceil \frac{\log\left(\frac{4n}{15^2}\right)}{2\log(np)} \right\rceil + 5.$$

Using Lemma 4.1.3 (iii) to bound $\mathbb{P}_{\mathcal{C}}\left(\mathcal{J}^c_{i,j}\right)$, since $5 \leq k \leq \lfloor \log(n)/\log(np) \rfloor - 5$ when $n$ large:

$$\begin{aligned}
&\mathbb{P}_{\mathcal{C}}(R(i,j) > 2k+1) \\
&\quad \leq \mathbb{P}_{\mathcal{C}}\left(d(i,j) > 2k+1 \middle| \mathcal{J}_{i,j}\right) + \mathbb{P}_{\mathcal{C}}\left(\mathcal{J}^c_{i,j}\right) \\
&\quad \leq \mathbb{P}\left(xy \notin E,\ \forall (x,y) \in B_k(i) \times B_k(j), B_k(i) \cap B_k(j) = \emptyset \middle| \mathcal{J}_{i,j}\right) / \mathbb{P}(\mathcal{C}) \\
&\qquad\qquad + 2\mathbb{P}_{\mathcal{C}}\left(|B_k(j)| < 2\sqrt{n}\right) \leq 2(1-p)^{4n} + 2\,o\left(n^{-4}\right) \\
&\quad = o\left(n^{-4}\right).
\end{aligned}$$

The result follows since $2k+1 = 2\left(\left\lceil \frac{\log\left(\frac{4n}{15^2}\right)}{2\log(np)} \right\rceil + 5\right) + 1 \leq \frac{3\log n}{\log(np)}$ for large $n$. □

## 4.3 Neighbourhood couplings

In this section we state and prove Lemma 4.3.1 which bounds the difference in distribution between the sizes of the pruned neighbourhoods $\Psi^d_1$ and $\Psi^d_2$ and the original neighbourhoods $\gamma^d_1$ and $\gamma^d_2$. Lemma 4.3.1, in combination with Lemma 2.1.2, will allow us to gain control over the $\Psi^d_1$ and $\Phi^d_1$ neighbourhood distributions in $\mathcal{G}(n,p)$ by relating them to the $\Gamma^*$-neighbourhood distributions which are known by Lemma 4.1.1.

**Lemma 4.3.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ and for any $c > 0$ let $c \log n \le np \le o(n^{1/3})$. Let $d \ge 0$ be fixed and let $I_1$ and the $\varphi_1^d$, $\psi_1^d$, $\psi_2^d$ and $\gamma^*$-distributions be defined with respect to* $\mathrm{MBFS}(\mathcal{G}, \{u,v\})$, $u, v \in V$. *Then*

*(i)* $\mathbb{P}\big(\varphi_1^d(x) \neq \gamma_1^*(x) \big| x \in I_1\big) = e^{-(1-o(1))np}$,

*(ii)* $\mathbb{P}\big(\psi_1^d(u) \neq \gamma_1^*(u)\big) = e^{-(1-o(1))np}$,

*(iii)* $\mathbb{P}\big(\psi_1^d(u) \neq \gamma_1(u)$ *or* $\psi_1^d(v) \neq \gamma_1(v)\big) \le 2np^2 + e^{-(1-o(1))np}$,

*(iv)* $\mathbb{P}\big(\psi_2^d(u) \neq \gamma_2(u)$ *or* $\psi_2^d(v) \neq \gamma_2(v)\big) \le 4n^3p^4 + e^{-(1-o(1))np} + O\big(n^2p^3\big)$.

*Proof. Item* (i): run $\mathrm{MBFS}(\mathcal{G}, \{u,v\})$ and let $x \in I_1$. By the definition (3.5) of $\psi_1^d(x)$, if $\gamma_1^*(\tilde{x}) > d$ for all $\tilde{x} \in \Gamma_1^*(x)$ then $\varphi_1^d(x) = \gamma_1^*(x)$. Hence for $x \in I_1$,

$$\begin{aligned}\mathfrak{A} &:= \mathbb{P}\Big(\varphi_1^d(x) \neq \gamma_1^*(x) \big| \mathfrak{F}_2\Big) \\ &= \mathbb{P}\big(\gamma_1^*(\tilde{x}) \le d \text{ for some } \tilde{x} \in \Gamma_1^*(x) \big| \mathfrak{F}_2\big) \\ &\le \sum_{\tilde{x} \in \Gamma_1^*(x)} \mathbb{P}\big(\gamma_1^*(\tilde{x}) \le d \big| \mathfrak{F}_2\big).\end{aligned}$$

If $\tilde{x} \in \Gamma_1^*(x), x \in I_1$ then $\tilde{x} \in I_2$. Knowing the parent of $\tilde{x}$ does not affect the $\gamma_1^*$-distribution conditioned on $\{\tilde{x} \in I_2\}$, so by Lemmas 4.1.1 (iii) as $|S_2|, |I_2| \in \mathfrak{F}_2$

$$\begin{aligned}\mathfrak{A} &\le \gamma_1^*(x)\mathbb{P}\Big(Bin\left(|S_2|, p(1-p)^{|I_2|-1}\right) \le d \Big| |S_2|, |I_2|\Big) \\ &= \gamma_1^*(x) \sum_{j=0}^{d} \binom{|S_2|}{j} \left(p(1-p)^{|I_2|-1}\right)^j \left(1 - p(1-p)^{|I_2|-1}\right)^{|S_2|-j}.\end{aligned}$$

Now using the bounds $\binom{n}{j} \le n^j/j!$ and $(1-p)^n \le \exp(-np)$ we have

$$\begin{aligned}\mathfrak{A} &\le \gamma_1^*(x) \sum_{j=0}^{d} \frac{\left(|S_2| p(1-p)^{|I_2|}\right)^j}{j!} \exp\left(-p(1-p)^{|I_2|}\left(|S_2| - j\right)\right) \\ &\le d\gamma_1^*(x) \left(|S_2| p(1-p)^{|I_2|}\right)^d \exp\left(-p(1-p)^{|I_2|}\left(|S_2| - d\right)\right).\end{aligned}$$

Let $\mathcal{E}_x := \big\{|I_2| \le 66(np)^2\big\} \cap \{\gamma_1^*(x) \le 32np\} \cap \{|S_2| \ge n - 66(np)^2\}$ for $x \in I_1$, then

$$\begin{aligned}\mathbb{P}\Big(\varphi_1^d(x) = \gamma_1^*(x) \big| \mathfrak{F}_2\Big) \mathbf{1}_{\mathcal{E}_x} &\le 32de^{(d+1)\log(np) - p(1-66n^2p^3)\left(n - 66(np)^2 - d\right)} \\ &= e^{-(1-o(1))np}. \end{aligned} \tag{4.16}$$

Recall that $|I_2| \le |B_2(u)| + |B_2(v)|$, $\gamma_1^*(x) \le \gamma_1(x)$ and $|S_2| \ge n - |B_2(u)| - |B_2(v)|$.

We have the following for $x \in I_1$, $np = \omega(\log\log n)$ by Lemma 4.1.2 (ii) with $k = 4$:

$$\begin{aligned}\mathbb{P}(\mathcal{E}_x^c) &\le \mathbb{P}(\gamma_1(x) > 32np) + 2\mathbb{P}\big(|B_2(u)| > 33(np)^2\big)\\ &\le 3 \cdot o\left(e^{-3(4-3)np/2}\right)\\ &= e^{-(1-o(1))np}.\end{aligned}$$

Now for $x \in I_1$, by the tower property, (4.16) and the above bound on $\mathbb{P}(\mathcal{E}_x^c)$, we have

$$\mathbb{P}\Big(\varphi_1^d(x) \neq \gamma_1^*(x)\Big) \le \mathbb{E}\Big[\mathbb{P}\Big(\varphi_1^d(x) = \gamma_1^*(x)\big|\mathfrak{F}_2\Big)\,\mathbf{1}_{\mathcal{E}_x}\Big] + \mathbb{P}(\mathcal{E}_x^c) = e^{-(1-o(1))np}.$$

*Item* (ii): for $\tilde{u} \in I_1$ the distribution of $\gamma_1^*(\tilde{u})$ conditioned on $|S_1|, |I_1|$ is known by 4.1.1 (iii). Thus using the bound $(1-p)^n \le \exp(-np)$ we obtain the following for $\tilde{u} \in I_1$,

$$\begin{aligned}\mathbb{P}\big(\gamma_1^*(\tilde{u}) = 0\big|\mathfrak{F}_1\big) &= \mathbb{P}\Big(Bin\left(|S_1|, p(1-p)^{|I_1|-1}\right) = 0\Big||S_1|, |I_1|\Big)\\ &= \left(1 - p(1-p)^{|I_1|-1}\right)^{|S_1|}\\ &\le \exp\left(-|S_1|p(1-p)^{|I_1|}\right).\end{aligned} \tag{4.17}$$

Recall the definition (3.6) of $\Psi_1^d(u)$. If $\tilde{u} \in \Gamma_1^*(u)$ then $\tilde{u} \in I_1$ and knowing the parent of $\tilde{u}$ does not affect the $\gamma_1^*$-distribution conditioned $\{\tilde{u} \in I_1\}$. So by Lemma 2.1.2 and (4.17) we have

$$\begin{aligned}\mathbb{P}\Big(\psi_1^d(u) \neq \gamma_1^*(u)\big|\mathfrak{F}_1\Big) &= \mathbb{P}\Big(\varphi_1^d(\tilde{u}) = 0 \text{ for some } \tilde{u} \in \Gamma_1^*(u)\big|\mathfrak{F}_1\Big)\\ &\le \sum_{\tilde{u} \in \Gamma_1^*(u)} \mathbb{P}\Big(\varphi_1^d(\tilde{u}) = 0\big|\mathfrak{F}_1\Big).\end{aligned}$$

Now using the coupling lemma, Lemma 2.1.2, yields the following for $\tilde{u} \in I_1$

$$\begin{aligned}\mathbb{P}\Big(\psi_1^d(u) \neq \gamma_1^*(u)\big|\mathfrak{F}_1\Big) &\le \gamma_1^*(u)\mathbb{P}\big(\gamma_1^*(\tilde{u}) = 0\big|\mathfrak{F}_1\big) + \gamma_1^*(u)\mathbb{P}\Big(\varphi_1^d(\tilde{u}) \neq \gamma_1^*(\tilde{u})\big|\mathfrak{F}_1\Big)\\ &\le \gamma_1^*(u)\exp\left(-|S_1|p(1-p)^{|I_1|}\right)\\ &\quad + \gamma_1^*(u)\mathbb{P}\Big(\varphi_1^d(\tilde{u}) \neq \gamma_1^*(\tilde{u})\big|\mathfrak{F}_1\Big).\end{aligned} \tag{4.18}$$

Let $\mathcal{E}_u := \{|I_1| \le 64np\} \cap \{\gamma_1^*(u) \le 32np\} \cap \{|S_1| \ge n - 66np\}$, $u \in I_0$. Now by

(2.3):

$$\gamma_1^*(u)\exp\left(-|S_1|p(1-p)^{|I_1|}\right)\mathbf{1}_{\mathcal{E}_u} \le e^{\log(32np)-(n-66np)p(1-64np^2)}$$
$$= e^{-(1-o(1))np}. \tag{4.19}$$

Recall $|I_1| \preceq_1 \gamma_1(u) + \gamma_1(v)$ and $|S_1| \ge n - |B_1(u)| - |B_1(u)|$. Lemmas 2.1.1 and 4.1.2 (ii) yield

$$\mathbb{P}(\mathcal{E}_u^c) \le 3\mathbb{P}(\gamma_1(u) > 32np) + 2\mathbb{P}(|B_1(u)| > 33np) = o\left(e^{-3np/2}\right). \tag{4.20}$$

For $u \in I_0, \tilde{u} \in I_1$ we use the tower property and (4.18) to give

$$\begin{aligned}\mathfrak{P} := \mathbb{P}\Big(\psi_1^d(u) \ne \gamma_1^*(u)\Big) &\le \mathbb{E}\Big[\mathbb{P}\Big(\psi_1^d(u) \ne \gamma_1^*(u)\big|\mathfrak{F}_1\Big)\mathbf{1}_{\mathcal{E}_u}\Big] + \mathbb{P}(\mathcal{E}_u^c)\\ &\le \mathbb{E}\Big[\gamma_1^*(u)\exp\left(-|S_2|p(1-p)^{|I_1|}\right)\mathbf{1}_{\mathcal{E}_u}\Big]\\ &\quad + \mathbb{E}\Big[\gamma_1^*(u)\mathbf{1}_{\mathcal{E}_u}\mathbb{P}\Big(\varphi_1^d(\tilde{u}) \ne \gamma_1^*(\tilde{u})\big|\mathfrak{F}_1\Big)\Big] + \mathbb{P}(\mathcal{E}_u^c)\,.\end{aligned}$$

Using the bounds from (4.19), Item (ii) and (4.20) on the above three terms respectively we obtain

$$\mathfrak{P} \le e^{-(1-o(1))np} + 32np\mathbb{P}\Big(\varphi_1^d(\tilde{u}) \ne \gamma_1^*(\tilde{u})\big|\tilde{u} \in I_1\Big) + o\left(e^{-3np/2}\right) = e^{-(1-o(1))np}.$$

*Item* (iii): let $I_0 = \{u, v\}$ and $\mathcal{H} := \{\gamma_1(u) = \gamma_1^*(u),\ \gamma_1(v) = \gamma_1^*(v)\}$. By Item (ii)

$$\begin{aligned}&\mathbb{P}\Big(\psi_1^d(u) \ne \gamma_1(u) \text{ or } \psi_1^d(v) \ne \gamma_1(v)\Big)\\ &\qquad \le \mathbb{P}\Big(\{\psi_1^d(u) \ne \gamma_1(u) \text{ or } \psi_1^d(v) \ne \gamma_1(v)\} \cap \mathcal{H}\Big) + \mathbb{P}(\mathcal{H}^c)\\ &\qquad \le \mathbb{P}\Big(\psi_1^d(u) \ne \gamma_1^*(u) \text{ or } \psi_1^d(v) \ne \gamma_1^*(u)\Big) + \mathbb{P}(\mathcal{H}^c)\\ &\qquad \le 2e^{-(1-o(1))np} + \mathbb{P}(\mathcal{H}^c)\,.\end{aligned} \tag{4.21}$$

To calculate $\mathbb{P}(\mathcal{H}^c)$ in the above recall the definition (3.3) of $\gamma_1^*(u)$ and observe

$$\begin{aligned}\mathbb{P}(\mathcal{H}^c) &= \mathbb{P}(\{uv \in E\} \cup \{xu \in E \text{ and } xv \in E \text{ for some } x \in V\setminus I_0\})\\ &\le \mathbb{P}(uv \in E) + \sum\nolimits_{x \in V\setminus I_0}\mathbb{P}(xu \in E \text{ and } xv \in E) = p + (n-2)p^2.\end{aligned} \tag{4.22}$$

Then combining (4.21) and (4.22) yields the bound

$$\begin{aligned}\mathbb{P}\Big(\psi_1^d(u) \neq \gamma_1(u) \text{ or } \psi_1^d(v) \neq \gamma_1(v)\Big) &\leq 2e^{-(1-o(1))np} + p + (n-2)p^2 \\ &\leq 2np^2 + e^{-(1-o(1))np}.\end{aligned}$$

*Item* (iv): let $I_0 = \{u, v\}$ and $\mathcal{L} := \{\gamma_2(u) = \gamma_2^*(u),\ \gamma_2(v) = \gamma_2^*(v)\}$. Then

$$\mathcal{L} := \left(\bigcap_{x\in\gamma_1(u),y\in\gamma_1(v)} \{xy \notin E\}\right) \cap \left(\bigcap_{z\in S_1} \{|\{x \in I_1 : xz \in E\}| \leq 1\}\right) \cap \mathcal{H}, \qquad (4.23)$$

by the definition (3.3) of $\gamma_2^*(u)$. Observe that by the Bernoulli inequality (2.3),

$$\begin{aligned}\mathbb{P}\big(|\{x \in I_1 : xz \in E\}| > 1 \big| \mathfrak{F}_1\big) &= 1 - \sum\nolimits_{a=0,1} \mathbb{P}\big(|\{x \in I_1 : xz \in E\}| = a \big| \mathfrak{F}_1\big) \\ &= 1 - (1-p)^{|I_1|} - |I_1|p(1-p)^{|I_1|-1} \\ &\leq 1 - (1 - |I_1|p) - |I_1|p(1 - |I_1|p) \\ &= (|I_1|p)^2 .\end{aligned}$$

By (4.23), the above estimate on $\mathbb{P}\big(|\{x \in I_1 : xz \in E\}| > 1 \big| \mathfrak{F}_1\big)$ and $\mathcal{H} \in \mathfrak{F}_1$, we have

$$\begin{aligned}\mathbb{P}\big(\mathcal{L}^c \big| \mathfrak{F}_1\big) \leq& \sum_{x\in\gamma_1(u),y\in\gamma_1(v)} \mathbb{P}\big(xy \in E \big| \mathfrak{F}_1\big) \\ &+ \sum_{z\in S_1} \mathbb{P}\big(|\{x \in I_1 : xz \in E\}| > 1 \big| \mathfrak{F}_1\big) + \mathbb{P}\big(\mathcal{H}^c \big| \mathfrak{F}_1\big) \\ \leq& \gamma_1(u)\gamma_1(v)p + |S_1| \left(|I_1|p\right)^2 + \mathbf{1}_{\mathcal{H}^c}.\end{aligned}$$

Then by the bound on $\mathbb{P}\big(\mathcal{L}^c \big| \mathfrak{F}_1\big)$ above, the tower property and Hölder's inequality (2.2) we have the following

$$\mathbb{P}(\mathcal{L}^c) = \mathbb{E}\big[\mathbb{P}\big(\mathcal{L}^c \big| \mathfrak{F}_1\big)\big] \leq p\sqrt{\mathbb{E}[\gamma_1(u)^2]\,\mathbb{E}[\gamma_1(v)^2]} + p^2\sqrt{\mathbb{E}[|S_1|^2]\,\mathbb{E}[|I_1|^4]} + \mathbb{E}[\mathbf{1}_{\mathcal{H}^c}] .$$

By Lemma 4.1.1 $|S_1| \sim_d Bin\left(n-2, (1-p)^2\right)$, $|I_1| \sim_d Bin\left(n-2, 2p(1-p)\right)$. Thus applying the bound on moments of binomial random variables from (2.16) yields

$$\begin{aligned}\mathbb{P}(\mathcal{L}^c) &\leq p\left((np)^2 + O(np)\right) + p^2\sqrt{\left(n^2 - 4n^2p + O(n)\right)\left(16(np)^4 + O((np)^3)\right)} + \mathbb{P}(\mathcal{H}^c) \\ &\leq n^2p^3 + O\big(np^2\big) + 4n^3p^4 + O\big(n^2p^3\big) + p + (n-2)p^2 \\ &= 4n^3p^4 + O\big(n^2p^3\big) .\end{aligned}$$

Let $\mathcal{F} := \big\{\psi_2^d(u) = \gamma_2(u),\ \psi_2^d(v) = \gamma_2(v)\big\}$. Then by the definitions (3.6) and (3.7) of the vertex sets $\Psi_1(u)$ and $\Psi_2(u)$ we have

$$\mathcal{F} := \Big\{\psi_1^d(u) = \gamma_1(u),\ \psi_1^d(v) = \gamma_1(v)\Big\} \cap \Big(\bigcap\nolimits_{x\in I_1} \Big\{\varphi_1^d(x) = \gamma_1^*(x)\Big\}\Big) \cap \mathcal{L}.$$

Let $\mathcal{D} = \{\gamma_1(u), \gamma_1(v) \leq (1 + 7/\min\{c, 1\})np, \}$. Then, for $x \in I_1$, by Items (i),(iii):

$$\begin{aligned}\mathbb{P}(\mathcal{F}^c) &\leq \mathbb{P}\Big(\psi_1^d(u) \neq \gamma_1(u) \text{ or } \psi_1^d(v) \neq \gamma_1(u)\Big) \\ &\quad + \mathbb{E}\Big[|I_1|\mathbb{P}\Big(\varphi_1^d(x) = \gamma_1^*(x)\big|\mathfrak{F}_1\Big)\left(\mathbf{1}_{\mathcal{D}} + \mathbf{1}_{\mathcal{D}^c}\right)\Big] + \mathbb{P}(\mathcal{L}^c) \\ &\leq 2np^2 + e^{-(1-o(1))np} + O\Big(npe^{-(1-o(1))np}\Big) + n\mathbb{P}(\mathcal{D}^c) + 4n^3p^4 + O\big(n^2p^3\big) \\ &\leq 4n^3p^4 + e^{-(1-o(1))np} + O\big(n^2p^3\big).\end{aligned}$$

The last inequality holds since

$$\mathbb{P}(\mathcal{D}^c) \leq 2\exp\left(-\frac{7^2}{2(\min\{c,1\})^2\,(1 + 7/\min\{c, 1\})}np\right) = o\left(1/n^2\right)$$

by Lemma 2.1.1 (ii). □

# Chapter 5

# Effective Resistance and Hitting times in $\mathcal{G}(n,p)$

In this Chapter we will apply the results of the previous chapter to prove Theorems 5.1.1, 5.2.1, 5.3.1 and 5.4.2. We must first state and prove Lemma 5.0.1 which provides higher reciprocal moments of $\Psi_1$ and $\Phi_1$ and provides other bounds which are of use to us when we prove the main theorems. These moments arise in the proof of Theorem 5.1.1 when we apply Hölder's inequality (2.2) to the resistance bound in Theorem 3.4.1.

Recall the definitions of $\mathcal{A}^k_{u,v}$ and $\mathcal{B}_{u,v}$ from (4.10) and (4.11) respectively. Recall also the function $D$ (4.9) in the definitions of $\mathcal{A}^k_{u,v}$ and $\mathcal{B}_{u,v}$ and that if we state a pruned neighbourhood without reference to $d$ then we take $d = D$, i.e. $\Psi_1 := \Psi_1^D$.

**Lemma 5.0.1.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$ *where* $\log n + \log\log\log n \leq np < o(n^{1/3})$. *Let* $\alpha \geq 1$ *and* $\Psi_1(u), \Psi_1(v)$ *be defined with respect to* $\mathrm{MBFS}(\mathcal{G}, \{u,v\})$, $u, v \in V$. *Then*

*(i)* $$\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\psi_1(u)^\alpha}\right]^{1/\alpha} = \frac{1}{np} + O\left(\frac{\log n}{(np)^2\log(np)}\right),$$

*(ii)* $$\mathbb{E}_{\mathcal{C}}\left[\left(\sup_{x\in\Psi_1(u)} \frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\varphi_1(x)}\right)^\alpha\right]^{1/\alpha} \leq O\left(\frac{1}{np}\right).$$

*(iii) If* $c\log n \leq np \leq n^{1/10}$, *for any fixed* $c > 0$, *then*

$$\mathbb{P}\left(R(u,v) > \left(\frac{1}{\psi_1(u)} + \frac{1}{\psi_1(v)}\right)\left(1 + \frac{9\log n}{np\log(np)}\right)\right) = o\left(e^{-np/4}\right) + o\left(n^{-7/2}\right).$$

*Proof. Item* (i): we restrict to the event $\mathcal{B}_u^{u,v} = \{\mathcal{G} \in \mathcal{B}_u^{u,v,D}\}$ to ensure the expectation is bounded,

$$\mathfrak{E} := \mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\psi_1(u)^\alpha}\right] = \sum_{k=1}^{n}\frac{1}{k^\alpha}\mathbb{P}_{\mathcal{C}}\big(\psi_1^D(u)=k\big) = \sum_{k=1}^{n}\frac{1}{k^\alpha}\frac{\mathbb{P}\big(\{\psi_1^D(u)=k\}\cap\mathcal{C}\big)}{\mathbb{P}(\mathcal{C})}.$$

Applying the coupling lemma, Lemma 2.1.2, and then Lemma 4.3.1 to bound $\mathbb{P}\big(\gamma_1^*(u)\neq\psi_1^D(u)\big)$ gives the following upper bound for the last term in the line above

$$\mathfrak{E} \leq \sum_{k=1}^{n}\frac{1}{k^\alpha}\frac{\mathbb{P}(\gamma_1^*(u)=k)+\mathbb{P}(\gamma_1^*(u)\neq\psi_1(u))}{\mathbb{P}(\mathcal{C})} = \sum_{k=1}^{n}\frac{1}{k^\alpha}\frac{\mathbb{P}(\gamma_1^*(u)=k)+e^{-(1-o(1))np}}{\mathbb{P}(\mathcal{C})}.$$

Let $\tilde{\gamma}_1(v) := |\Gamma_1(v)\cap S_0| \sim_d Bin(n-2,p)$. By Lemma 4.1.1 we have $\gamma_1^*(u) \sim_d Bin(n-2-h,p)$ conditional on $\{\tilde{\gamma}_1(v)=h\}$. The law of total expectation and the fact that the harmonic series diverges at rate $\log(n)$ yield the following

$$\mathfrak{E} \leq \sum_{k=1}^{n}\frac{1}{k^\alpha}\frac{\sum_{h=0}^{n-2}\mathbb{P}\Big(\gamma_1^*(u)=k\Big|\tilde{\gamma}_1(v)=h\Big)\mathbb{P}(\tilde{\gamma}_1(v)=h)}{\mathbb{P}(\mathcal{C})} + O\left(\frac{(\log n)e^{-(1-o(1))np}}{\mathbb{P}(\mathcal{C})}\right).$$

Now by writing out $\mathbb{P}\Big(\gamma_1^*(u)=k\Big|\tilde{\gamma}_1(v)=h\Big)\mathbb{P}(\tilde{\gamma}_1(v)=h)$ explicitly we have

$$\begin{aligned}\mathfrak{E} &\leq \sum_{k=1}^{n}\frac{1}{k^\alpha}\frac{\sum_{h=0}^{n-2}\binom{n-2-h}{k}p^k(1-p)^{n-2-h-k}\cdot\binom{n-2}{h}p^h(1-p)^{n-2-h}}{\mathbb{P}(\mathcal{C})} + e^{-(1-o(1))np}\\ &= \sum_{h=0}^{n-3}\binom{n-2}{h}\frac{p^h(1-p)^{n-2-h}}{\mathbb{P}(\mathcal{C})}\left(\sum_{k=1}^{n-2-h}\frac{1}{k^\alpha}\binom{n-2-h}{k}p^k(1-p)^{n-2-h-k}\right)\\ &\quad + e^{-(1-o(1))np}.\end{aligned}$$

Applying Proposition 2.4.2 to the bracketed sum above where we let $X_h$ be a random variable with distribution $Bin(n-h-3,p)$ yields

$$\mathfrak{E} \leq \frac{np}{\mathbb{P}(\mathcal{C})}\sum_{h=0}^{n-3}\binom{n-2}{h}p^h(1-p)^{n-2-h}\mathbb{E}\left[\frac{1}{(X_h+1)^{\alpha+1}}\right] + e^{-(1-o(1))np}.$$

The weight in front of the expectation term is the density of a $Bin(n-2,p)$ random variable. Split the sum at $t := \sqrt{3np(\alpha+2)\log(np)}$ and bound the expectation to give

$$\mathfrak{E} \le \frac{np}{\mathbb{P}(\mathcal{C})}\left(\mathbb{P}(Bin(n-2,p)\le t)\,\mathbb{E}\left[\frac{1}{(X_t+1)^{\alpha+1}}\right] + \mathbb{P}(Bin(n-2,p)>t)\right) \\ + e^{-(1-o(1))np}.$$

Using Lemma 2.1.1 to bound $\mathbb{P}(Bin(n-2,p)>t)$ and Lemma 2.4.3 to calculate the expectation term $\mathbb{E}\left[\frac{1}{(X_t+1)^{\alpha+1}}\right]$ we have

$$\mathfrak{E} \le \frac{np}{\mathbb{P}(\mathcal{C})}\left[\left(\frac{1}{((n-t-3)p)^{\alpha+1}} + O\left(\frac{1}{((n-t-3)p)^{\alpha+2}}\right)\right) + o\left(\frac{1}{(np)^{\alpha+2}}\right)\right] \\ + e^{-(1-o(1))np}.$$

By (2.15), $\mathbb{P}(\mathcal{C}^c) \le O(\log n/(np\log(np)))$ whenever $np \ge \log n + \log\log\log n$. Thus

$$\mathfrak{E} \le \frac{1}{(1-\mathbb{P}(\mathcal{C}^c))(np)^\alpha} + O\left(\frac{1}{(np)^{\alpha+1}}\right) = \frac{1}{(np)^\alpha}\left(1+O(\mathbb{P}(\mathcal{C}^c))\right) + O\left(\frac{1}{(np)^{\alpha+1}}\right).$$

Recall the Bernoulli inequality (2.3): $(1+x)^r \le 1+rx$, for any $x > -1$ and $0 \le r \le 1$. Applying this yields

$$\begin{aligned}
\mathfrak{E}^{1/\alpha} &\le \left(\frac{(1+O(\mathbb{P}(\mathcal{C}^c)))}{(np)^\alpha} + O\left(\frac{1}{(np)^{\alpha+1}}\right)\right)^{1/\alpha} \\
&= \frac{(1+O(\mathbb{P}(\mathcal{C}^c)))^{1/\alpha}}{np}\left(1+O\left(\frac{1}{np}\right)\right)^{1/\alpha} \\
&= \frac{1+O(\mathbb{P}(\mathcal{C}^c))}{np}\left(1+O\left(\frac{1}{np}\right)\right) \\
&= \frac{1}{np} + O\left(\frac{\log n}{(np)^2\log(np)}\right).
\end{aligned}$$

*Item* (ii): Let $\mathcal{H}$ be the event $\{\varphi_1^D(x) = \gamma_1^*(x) \text{ for all } x \in I_1\} \in \mathfrak{F}_3$ and define

$$K_p := \left(1-\sqrt{3}/2\right)np(1-66np^2). \tag{5.1}$$

Recall $\Psi_1(u) \subset I_1$ for $u \in I_0$ and switch between the $\varphi$ and $\gamma_1^*$ distributions on the event $\mathcal{H}$:

$$\mathfrak{P}_u := \mathbb{P}\left(\inf_{x\in\Psi_1(u)} \varphi_1(x) < K_p\right) \le \mathbb{P}\left(\left\{\inf_{x\in I_1} \gamma_1^*(x) < K_p\right\} \cap \mathcal{H}\right) + \mathbb{P}(\mathcal{H}^c)\,.$$

Now by the tower property and the definition of $\mathcal{H}$ we have

$$\mathfrak{P}_u \le \mathbb{E}\left[\mathbb{P}\left(\inf_{x\in I_1} \gamma_1^*(x) < K_p \middle| \mathfrak{F}_1\right)\right] + \mathbb{E}\left[\mathbb{P}\left(\varphi_1(x) \neq \gamma_1^*(x) \text{ for some } x \in I_1 \middle| \mathfrak{F}_1\right)\right].$$

Applying the union bound since $I_1 \in \mathfrak{F}_1$ yields

$$\mathfrak{P}_u \le \mathbb{E}\left[|I_1|\mathbb{P}\left(\gamma_1^*(x) < K_p \middle| x \in I_1, \mathfrak{F}_1\right)\right] + \mathbb{E}\left[|I_1|\mathbb{P}\left(\varphi_1(x) \neq \gamma_1^*(x) \middle| x \in I_1, \mathfrak{F}_1\right)\right].$$

Let $a := 4/\min\{c, 1\}$ where $c > 0$ is any fixed positive real number such that $np \geq c\log n$. Separate the expectations into parts $\{|I_1| \le 4a^2np\}$ and $\{|I_1| > 4a^2np\}$:

$$\mathfrak{P}_u \le 4a^2np\mathbb{E}\left[\mathbb{P}\left(\gamma_1^*(x) < K_p \middle| x \in I_1, \mathfrak{F}_1\right)\right] \\ + 4a^2np\mathbb{P}\left(\varphi_1(x) \neq \gamma_1^*(x) \middle| x \in I_1\right) + 2n\mathbb{P}\left(|I_1| > 4a^2np\right).$$

Since $\gamma_1^*(x) \sim_d Bin(|S_1(x)|, p)$ by Lemma 4.1.1, $S_1(x) \in \mathfrak{F}_2$, and by Lemma 4.3.1 (i) we have

$$\mathfrak{P}_u \le 4a^2np\mathbb{E}\left[\mathbb{P}\left(Bin(|S_1(x)|, p) < K_p \middle| \mathfrak{F}_2\right)\right] \\ + 4a^2(np)e^{-(1-o(1))np} + 4n\mathbb{P}\left(\gamma_1(u) > 2a^2np\right).$$

Applying Lemma 2.1.1 to the first term and Lemma 4.1.2 (i) with $k = a$ to the last yields

$$\mathfrak{P}_u \le 4a^2np\mathbb{E}\left[e^{-(|S_1(x)|p-K_p)^2/(2|S_1(X)|p)}\right] \\ + 4a^2(np)e^{-(1-o(1))np} + 4n \cdot o\left(e^{-3(a-3)np/2}\right).$$

Separating the expectation into the two disjoint parts $\{|S_1(x)| \le n - 66(np)^2\}$ and $\{|S_1(x)| > n - 66(np)^2\}$ we have the following bound from above

$$\mathfrak{P}_u \le 4a^2(np) \cdot e^{-\left((n-66(np)^2)p - \left(1-\sqrt{3}/2\right)np(1-66np^2)\right)^2/(2np)} \\ + 2\mathbb{P}\left(|B_2(u)| > 33(np)^2\right) + o\left(e^{-np/2}\right).$$

Rearranging the first term and applying Lemma 4.1.2 (ii) with $k = 4$ to the middle term yields the following

$$\mathfrak{P}_u \le 4a^2npe^{-np/3} + o\left(e^{-3(4-3)np/2}\right) + o\left(e^{-np/2}\right) = o\left(e^{-np/4}\right). \tag{5.2}$$

Recall $\sup_{x\in\Psi_1(u)} \mathbf{1}_{\mathcal{B}_u^{u,v}}/\varphi_1(x) < 1/D$, see (3.5) & (3.6). Applying the Bernoulli inequality (2.3) we obtain

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[\left(\sup_{x\in\Psi_1(u)}\frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\varphi_1(x)}\right)^{\alpha}\right]^{1/\alpha} &\le \left(\frac{1}{(K_p)^{\alpha}}+\frac{1}{D^{\alpha}\mathbb{P}(\mathcal{C})}\mathbb{P}\Big(\inf_{x\in\Psi_1(u)}\varphi_1(x)<K_p\Big)\right)^{1/\alpha}\\ &\le \frac{1}{(K_p)}\left(1+(K_p)^{\alpha}e^{-np/4}/D^{\alpha}\mathbb{P}(\mathcal{C})\right)^{1/\alpha}\\ &\le O\Big(\frac{1}{np}\Big). \end{aligned} \tag{5.3}$$

Note that the bound (5.2) on $\mathfrak{P}_{\mathfrak{u}}$ holds for any $np \ge c\log n$, $c > 0$ fixed. The restriction on $np$ to $np \ge \log n$ comes from (5.3), where we need $\mathbb{P}(\mathcal{C})$ bounded below by a constant.
*Item* (iii): conditioning on the event $\mathcal{A}^n_{u,v}$ and applying Theorem 3.4.1 yields

$$\begin{aligned}R(u,v) &\le \frac{1}{\psi_1(u)}+\frac{1}{\psi_1(v)}+\sum_{a\in\Psi_1(u)}\frac{k+2}{\psi_1(u)^2\varphi_1(a)}+\sum_{b\in\Psi_1(v)}\frac{k+2}{\psi_1(v)^2\varphi_1(b)}\\ &\le \frac{1}{\psi_1(u)}\left(1+(k+2)\cdot\sup_{x\in\Psi_1(u)}\frac{1}{\varphi_1(x)}\right)+\frac{1}{\psi_1(v)}\left(1+(k+2)\cdot\sup_{x\in\Psi_1(v)}\frac{1}{\varphi_1(x)}\right).\end{aligned}$$

Recall $K_p$ from (5.1) and bound on $k$ in description of the event $\mathcal{A}^n_{u,v}$, note that $K_p \ge np/9$ and $k \le \log(n)/\log(np)$ for large $n$. Thus inserting these bounds, conditional on $\mathcal{H} := \{\varphi_1(a) \ge K_p \text{ for all } a \in \Psi_1\} \cap \mathcal{A}^n_{u,v}$ we have

$$R(u,v) \le \left(\frac{1}{\psi_1(u)}+\frac{1}{\psi_1(v)}\right)\left(1+\frac{k+2}{K_p}\right) \le \left(\frac{1}{\psi_1(u)}+\frac{1}{\psi_1(v)}\right)\left(1+\frac{9\log n}{np\log(np)}\right).$$

Applying the bounds on $\mathfrak{P}_u$ from (5.2) and on $\mathbb{P}\big((\mathcal{A}^n_{u,v})^c\big)$ by Lemma 4.2.1 yield

$$\mathbb{P}(\mathcal{H}^c) \le 2\mathbb{P}\Big(\inf_{x\in\Psi_1(u)}\varphi_1(x)<K_p\Big)+\mathbb{P}\big((\mathcal{A}^n_{u,v})^c\big) = o\left(e^{-np/4}\right)+o\left(n^{-7/2}\right).$$

□

## 5.1 Expectation for hitting times, Theorem 5.1.1

Let $\mathcal{S}_{i,j}$ be the event $\{R(i,j) > 3\log n/\log(np)\}$. By Lemma 4.2.2, if $np-\log n \to \infty$ then we have the following

$$\mathbb{P}_{\mathcal{C}}(\mathcal{S}_{i,j}) := \mathbb{P}_{\mathcal{C}}(R(i,j) > 3\log n/\log(np)) = o(n^{-4}). \tag{5.4}$$

If $G \in \mathcal{C}$ then there is a path of length at most $n-1$ between any $i, j \in V$. Since effective resistance is bounded by graph distance for all $i, j \in V$ we have the bound

$$R(i,j) \leq n-1. \tag{5.5}$$

Let $\mathcal{C} := \mathcal{C}_n$ be the event that $\mathcal{G} \sim_d \mathcal{G}(n,p)$ is connected. Let $a(n), b(n) : \mathbb{N} \to \mathbb{R}$, then for ease of presentation we use the notation

$$a(n) \overset{O}{=} b(n) \quad \text{to denote} \quad a(n) = \left(1 \pm O\left(\frac{\log n}{np\log(np)}\right)\right) b(n).$$

We shall now (re)state one of the main theorems of this chapter and then prove it.

**Theorem 5.1.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ with $\log n + \log\log\log n \leq np \leq n^{1/10}$. Then for any $i, j \in V(\mathcal{G})$ where $i \neq j$,*

*(i)* $\mathbb{E}\big[R(i,j)\big|\mathcal{C}\big] \overset{O}{=} \dfrac{2}{np}$, $\quad \mathbb{E}\big[h(i,j)\big|\mathcal{C}\big] \overset{O}{=} n$, $\quad \mathbb{E}\big[\kappa(i,j)\big|\mathcal{C}\big] \overset{O}{=} 2n$,

*(ii)* $\mathbb{E}\big[K(\mathcal{G})\big|\mathcal{C}\big] \overset{O}{=} \dfrac{n}{p}$, $\quad \mathbb{E}\big[\overline{cc}(\mathcal{G})\,\big|\mathcal{C}\big] \overset{O}{=} n$, $\quad \mathbb{E}\big[cc_i(\mathcal{G})\big|\mathcal{C}\big] \overset{O}{=} n$,

*(iii)* $\mathbb{E}\big[K(\mathcal{G})^2\big|\mathcal{C}\big] \overset{O}{=} \dfrac{n^2}{p^2}$, $\quad \mathbb{E}\big[h(i,j)^2\big|\mathcal{C}\big] \overset{O}{=} n^2$, $\quad \mathbb{E}\Big[cc_i(\mathcal{G})^2\,\Big|\mathcal{C}\Big] \overset{O}{=} n^2$,

*(iv)* $\mathbb{E}\big[H_i(\mathcal{G})\big|\mathcal{C}\big] \overset{O}{=} n$, $\quad \mathbb{E}\big[H(\mathcal{G})\big|\mathcal{C}\big] \overset{O}{=} n$, $\quad \mathbb{E}\big[T(\mathcal{G})\big|\mathcal{C}\big] \overset{O}{=} n$,

*(v)* $\mathbb{E}\big[H_i(\mathcal{G})^2\big|\mathcal{C}\big] \overset{O}{=} n^2$, $\quad \mathbb{E}\big[H(\mathcal{G})^2\big|\mathcal{C}\big] \overset{O}{=} n^2$, $\quad \mathbb{E}\big[T(\mathcal{G})^2\big|\mathcal{C}\big] \overset{O}{=} n^2$.

We shall arrange the proof as follows: the proof of each item in the statement of Theorem 5.1.1 will appear as a subsection and within this subsection each moment calculation will be contained in a proof environment.

### 5.1.1 Proof of Theorem 5.1.1 (i)

*Proof of* $\mathbb{E}_{\mathcal{C}}[R(i,j)]$. Let $\mathcal{C}_1$ be the event $\mathcal{A}^n_{i,j} \cap \mathcal{B}_{i,j}$ and $k^* = \log(n)/\log(np)$. Observe conditional on $\mathcal{C}_1$, for large $n$, we have $k+2 \leq k^*$, where $k$ is as in Theorem 3.4.1. First we apply the bound on resistance from Theorem 3.4.1 to bound $\mathbb{E}_{\mathcal{C}}[R(i,j)\mathbf{1}_{\mathcal{C}_1}]$ from above

$$\mathbb{E}_{\mathcal{C}}[R(i,j)\mathbf{1}_{\mathcal{C}_1}] \leq \sum_{x\in\{i,j\}} \left( \mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}^{i,j}_x}}{\psi_1(x)}\right] + k^* \cdot \mathbb{E}_{\mathcal{C}}\left[\sum_{a\in\Psi_1(x)} \frac{\mathbf{1}_{\mathcal{A}^n_{i,j}\cap\mathcal{B}_{i,j}}}{\psi_1(x)^2\varphi_1(a)}\right]\right). \tag{5.6}$$

By Lemma 5.0.1 (i) the first term in the sum is $1/(np) + O\big(\log n/(np)^2\log(np)\big)$. To bound the second term, start by pulling out $\sup_{a\in\Psi_1(x)} 1/\varphi_1(a)$ from the sum over

$a \in \Psi_1(x)$ to obtain

$$\mathfrak{E} := k^* \cdot \mathbb{E}_{\mathcal{C}}\left[\sum_{a \in \Psi_1(x)} \frac{\mathbf{1}_{\mathcal{A}^n_{i,j} \cap \mathcal{B}_{i,j}}}{\psi_1(x)^2 \varphi_1(a)}\right] \leq k^* \cdot \mathbb{E}_{\mathcal{C}}\left[\left(\sup_{a \in \Psi_1(x)} \frac{1}{\varphi_1(a)}\right) \frac{\mathbf{1}_{\mathcal{A}^n_{i,j} \cap \mathcal{B}_{i,j}}}{\psi_1(x)}\right].$$

Using Hölder's inequality (2.2) on the product of random variables in the expectation gives

$$\mathfrak{E} \leq k^* \cdot \mathbb{E}_{\mathcal{C}}\left[\sup_{a \in \Psi_1(x)} \frac{\mathbf{1}_{\mathcal{B}^{i,j}_x}}{\varphi_1(a)^2}\right]^{1/2} \mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}^{i,j}_x}}{\psi_1(x)^2}\right]^{1/2}.$$

Upper bounds for each of the expectation terms can be found in Lemma 5.0.1, yielding

$$\mathfrak{E} \leq \left(\frac{\log n}{\log(np)}\right) \cdot O\left(\frac{1}{np}\right) \cdot \left(\frac{1}{np} + O\left(\frac{\log n}{(np)^2 \log(np)}\right)\right) = O\left(\frac{\log n}{(np)^2 \log(np)}\right).$$

Combining the estimates on $\mathfrak{E}$ above with the bound on $\mathbb{E}_{\mathcal{C}}\left[\mathbf{1}_{\mathcal{B}^{i,j}_x}/\psi_1(x)\right]$ by Lemma 5.0.1 (i) yields the following

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[R(i,j)\mathbf{1}_{\mathcal{C}_1}\right] &\leq 2\left(\frac{1}{np} + O\left(\frac{\log n}{(np)^2 \log(np)}\right) + O\left(\frac{\log n}{(np)^2 \log(np)}\right)\right) \\ &= \frac{2}{np} + O\left(\frac{\log n}{(np)^2 \log(np)}\right).\end{aligned}$$

When $np \geq c \log n$ and $c > 3$ we have the following for $\mathbb{E}_{\mathcal{C}}\left[R(i,j)\mathbf{1}_{(\mathcal{C}_1)^c}\right]$ by first applying the effective resistance bound (5.5) then bounds on $\mathbb{P}[\mathcal{C}_1^c]$ from Lemma 4.2.1:

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[R(i,j)\mathbf{1}_{(\mathcal{C}_1)^c}\right] &\leq \frac{(n-1)}{\mathbb{P}(\mathcal{C})}\mathbb{P}\left(\left(\mathcal{A}^n_{i,j} \cap \mathcal{B}^{i,j}_{i,j}\right)^c\right) \\ &= n\left(e^{-(1-o(1))np} + o\left(1/n^{7/2}\right)\right) \\ &= o\left(1/n^2\right).\end{aligned}$$

If $\log n + \log\log\log n \leq np \leq 3\log n$ then we further partition using $\mathcal{S}_{i,j}$ from (5.4) to obtain

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[R(i,j)\mathbf{1}_{(\mathcal{C}_1)^c}\left(\mathbf{1}_{\mathcal{S}_{i,j}} + \mathbf{1}_{\mathcal{S}^c_{i,j}}\right)\right] &\leq \frac{3\log n}{\mathbb{P}(\mathcal{C})\log np}\mathbb{P}\left(\left(\mathcal{A}^n_{i,j} \cap \mathcal{B}^{i,j}_{i,j}\right)^c\right) + n\mathbb{P}_{\mathcal{C}}\left(\mathcal{S}^c_{i,j}\right) \\ &= o\left(1/n^{4/5}\right).\end{aligned}$$

The upper bound follows as $\mathbb{E}_{\mathcal{C}}[R(i,j)] = \mathbb{E}_{\mathcal{C}}[R(i,j)\mathbf{1}_{\mathcal{C}_1}] + \mathbb{E}_{\mathcal{C}}\left[R(i,j)\mathbf{1}_{(\mathcal{C}_1)^c}\right]$. Let $\mathcal{D}$ be the event $\{\gamma_1(i), \gamma_1(j) \geq np - a\sqrt{np}\}$ where $a = 3\sqrt{\log n}$ if $np = \omega(\log n)$ and $a = 3\sqrt{\log\log n}$ if $np = O(\log n)$. Then applying Lemma 3.1.3 and $1 \geq \mathbf{1}_{\mathcal{D}}$ yields

$$\mathbb{E}_{\mathcal{C}}[R(i,j)] \geq \mathbb{E}_{\mathcal{C}}\left[\frac{1}{\gamma_1(i)+1} + \frac{1}{\gamma_1(j)+1}\right] \geq 2\frac{1}{np + a\sqrt{np}}\mathbb{P}_{\mathcal{C}}(\mathcal{D})$$

when $i \neq j$. Making use of the inequality $\mathbb{P}_{\mathcal{C}}(\mathcal{D}^c) \leq \mathbb{P}(\mathcal{D}^c)/\mathbb{P}(\mathcal{C})$ and then bounding $\mathbb{P}(\mathcal{D}^c)$ from below by Lemma 2.1.1 we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[\gamma_1(u) \cdot R(i,j)] &\geq \left(\frac{2}{np} - O\left(\frac{a}{(np)^{3/2}}\right)\right)\left(1 - \frac{1}{\mathbb{P}(\mathcal{C})}(e^{-\frac{a^2}{2}} - 2e^{-\frac{a^2}{3}})\right) \\ &= \frac{2}{np} - O\left(\frac{\log n}{np\log np}\right).\end{aligned}$$

□

*Proof of* $\mathbb{E}_{\mathcal{C}}[h(i,j)]$. We have the following expression for hitting times from (2.14):

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[h(i,j)] &= \mathbb{E}_{\mathcal{C}}[mR(i,j)] + \frac{1}{2}\sum_{u\in V}\left(\mathbb{E}_{\mathcal{C}}[d(u)R(u,j)] - \mathbb{E}_{\mathcal{C}}[d(u)R(u,i)]\right) \\ &= \mathbb{E}_{\mathcal{C}}[mR(i,j)],\end{aligned}$$

when $i \neq j$, by symmetry. We will calculate $\mathbb{E}_{\mathcal{C}}[\gamma_1(u)R(i,j)]$ and apply

$$\mathbb{E}_{\mathcal{C}}[m \cdot R(i,j)] = \frac{1}{2}\sum_{u\in V}\mathbb{E}_{\mathcal{C}}[\gamma_1(u)R(i,j)].$$

Let $\mathcal{M}$ be the event $\{\gamma_1(u) \leq 5np, \text{ for all } u \in V\}$. Then for each $\{i,j\} \subset V$ we define the disjoint events

$$\mathcal{C}_1 := \mathcal{A}_{i,j}^n \cap \mathcal{B}_{i,j}, \quad \mathcal{C}_2 := (\mathcal{C}_1)^c \cap \mathcal{M}, \quad \mathcal{C}_3 := (\mathcal{C}_1)^c \cap \mathcal{M}^c.$$

We will now bound $\mathbb{E}_{\mathcal{C}}[\gamma_1(u) \cdot R(i,j) \cdot \mathbf{1}_{\mathcal{C}_1}]$ from above using the Hölder inequality (2.2). This is almost identical to the calculation for $\mathbb{E}_{\mathcal{C}}[R(i,j) \cdot \mathbf{1}_{\mathcal{C}_1}]$, see (5.6). However, we also use (2.16) to give bounds of the form $\mathbb{E}[\gamma_1(u)^\alpha] = (np)^\alpha + O\big((np)^{\alpha-1}\big)$

where $\alpha$ is a positive real number. We have

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\cdot R(i,j)\cdot \mathbf{1}_{\mathcal{C}_1}]\\
&\le \sum_{x\in\{i,j\}}\left(\mathbb{E}_{\mathcal{C}}\left[\frac{\gamma_1(u)\mathbf{1}_{\mathcal{B}_x^{i,j}}}{\psi_1(x)}\right]+k^*\cdot\mathbb{E}_{\mathcal{C}}\left[\sum_{x_a\Psi_1(x)}\frac{\gamma_1(u)\mathbf{1}_{\mathcal{C}_1}}{\psi_1(x)^2\varphi_1(x_a)}\right]\right)\\
&\le 2\left(1+O\left(\frac{\log n}{(np)\log(np)}\right)+O\left(\frac{\log n}{(np)\log(np)}\right)\right)\\
&=2+O\left(\frac{\log n}{(np)\log(np)}\right).
\end{aligned}
$$

When $np \ge c\log n$ and $c>3$ for expectation on $\mathcal{C}_2 := \mathcal{C}_1^c\cap\mathcal{M}$ we apply the trivial effective resistance bound $R(i,j)\le n-1$ (5.5) and $\gamma_1(u)\mathbf{1}_{\mathcal{M}}\le 5np$, then bound $\mathbb{P}(\mathcal{C}_1^c)$ by Lemma 4.2.1 yielding

$$
\mathbb{E}_{\mathcal{C}}[\gamma_1(u)R(i,j)\mathbf{1}_{\mathcal{C}_2}] \le \frac{(5np)(n-1)}{\mathbb{P}(\mathcal{C})}\mathbb{P}\left(\left(\mathcal{A}_{i,j}^n\cap\mathcal{B}_{i,j}^{i,j}\right)^c\right)=o\left(1/n^2\right).
$$

If $\log n+\log\log\log n\le np\le 3\log n$ then we further partition using $\mathcal{S}_{i,j}$ from (5.4), to obtain

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)R(i,j)\mathbf{1}_{\mathcal{C}_2}\left(\mathbf{1}_{\mathcal{S}_{i,j}}+\mathbf{1}_{\mathcal{S}_{i,j}^c}\right)\right] &\le (5np)\,\frac{3\log n}{\mathbb{P}(\mathcal{C})\log np}\mathbb{P}((\mathcal{C}_1)^c)+5n^2p\mathbb{P}_{\mathcal{C}}\left(\mathcal{S}_{i,j}^c\right)\\
&= o\left(1/n^{4/5}\right).
\end{aligned}
$$

Since $\mathbb{P}_{\mathcal{C}}\left(\mathcal{M}^c\right)\le n\cdot\exp\left(-3\cdot 4^2np/8\right)/\mathbb{P}(\mathcal{C})=o(1/n^5)$ by Lemma 2.1.1 we have

$$
\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\cdot R(i,j)\mathbf{1}_{\mathcal{C}_3}]\le (n-1)^2\mathbb{P}_{\mathcal{C}}(\mathcal{M}^c)=o(n^{-3}).
$$

Combining expectations over $\mathcal{C}_1,\mathcal{C}_2$ and $\mathcal{C}_3$ yields the following for any $u,i,j\in V$ and any $i\ne j$

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\cdot R(i,j)] &= \mathbb{E}_{\mathcal{C}}[\gamma_1(u)\cdot R(i,j)\left(\mathbf{1}_{\mathcal{C}_1}+\mathbf{1}_{\mathcal{C}_2}+\mathbf{1}_{\mathcal{C}_3}\right)]\\
&\le 2+O\left(\frac{\log n}{np\log np}\right).
\end{aligned}
\tag{5.7}
$$

Let $\mathcal{D}$ be the event $\{\gamma_1(u)\ge np-a\sqrt{np}\}\cap\{\gamma_1(i),\gamma_1(j)\le np+a\sqrt{np}\}$ where $a=3\sqrt{\log n}$ if $np=\omega(\log n)$ and $a=3\sqrt{\log\log n}$ if $np=O(\log n)$. Then by Lemma 3.1.3 and $1\ge\mathbf{1}_{\mathcal{D}}$ we have the following

$$
\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\cdot R(i,j)]\ge\mathbb{E}_{\mathcal{C}}\left[\frac{\gamma_1(u)}{\gamma_1(i)+1}+\frac{\gamma_1(u)}{\gamma_1(j)+1}\right]\ge 2\frac{np-a\sqrt{np}}{np+a\sqrt{np}}\mathbb{P}_{\mathcal{C}}(\mathcal{D})
$$

when $i \neq j$. Since $\mathbb{P}_{\mathcal{C}}(\mathcal{D}^c) \leq \mathbb{P}(\mathcal{D}^c)\,/\mathbb{P}(\mathcal{C})$ and bounding $\mathbb{P}(\mathcal{D}^c)$ by Lemma 2.1.1 we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\cdot R(i,j)] &\geq \left(2 - O\left(\frac{a}{\sqrt{np}}\right)\right)\left(1 - \frac{1}{\mathbb{P}(\mathcal{C})}(e^{-\frac{a^2}{2}} - 2e^{-\frac{a^2}{3}})\right)\\ &= 2 - O\left(\frac{\log n}{np\log np}\right). \end{aligned} \tag{5.8}$$

Summing (5.7) and (5.8) over $u \in V$ yields the required bounds for $\mathbb{E}_{\mathcal{C}}[h(i,j)]$. □

Recall that for functions $a(n), b(n)$ we use

$$a(n) \overset{O}{=} b(n) \quad \text{to denote} \quad a(n) = \left(1 \pm O\left(\frac{\log n}{np\log(np)}\right)\right) b(n).$$

*Proof of* $\mathbb{E}_{\mathcal{C}}[\kappa(i,j)]$. This follows from the result for $\mathbb{E}_{\mathcal{C}}[h(i,j)]$ as by (2.12) we have

$$\mathbb{E}_{\mathcal{C}}[\kappa(i,j)] = \mathbb{E}_{\mathcal{C}}[h(i,j) + h(j,i)] = 2\mathbb{E}_{\mathcal{C}}[h(i,j)] \overset{O}{=} 2n.$$

□

### 5.1.2 Proof of Theorem 5.1.1 (ii)

*Proof of* $\mathbb{E}_{\mathcal{C}}[K(\mathcal{G})]\,, \mathbb{E}_{\mathcal{C}}[cc_i(\mathcal{G})]\,, \mathbb{E}_{\mathcal{C}}[\overline{cc}(\mathcal{G})]$. We will use linearity of expectation to express the expectations of these indices in terms of quantities we have already calculated. The bounds for $\mathbb{E}_{\mathcal{C}}[R(i,j)]$ in Theorem 5.1.1 (i) hold for all $\{i,j\} \subseteq V$. Hence by (2.10) we have

$$\mathbb{E}_{\mathcal{C}}[K(\mathcal{G})] = \sum_{\{i,j\}\subseteq V} \mathbb{E}_{\mathcal{C}}[R(i,j)] \overset{O}{=} \frac{n(n-1)}{2} \cdot \frac{2}{np} \overset{O}{=} \frac{n}{p}.$$

The bounds for $\mathbb{E}_{\mathcal{C}}[h(i,j)]$ in Theorem 5.1.1 (i) hold for all $i, j \in V,\ i \neq j$. So by (2.11) we have

$$\mathbb{E}_{\mathcal{C}}[cc_i(\mathcal{G})] = \frac{1}{n-1} \sum_{j\in V\setminus\{i\}} \mathbb{E}_{\mathcal{C}}[h(i,j)] \overset{O}{=} \frac{1}{n-1} \cdot (n-1) \cdot n \overset{O}{=} n.$$

The bounds for $\mathbb{E}_{\mathcal{C}}[\kappa(i,j)]$ in Theorem 5.1.1 (i) hold for all $\{i,j\} \subseteq V$. Thus by (2.13) we have

$$\mathbb{E}_{\mathcal{C}}[\overline{cc}(\mathcal{G})] = \frac{1}{n(n-1)} \sum_{\{i,j\}\subseteq V} \mathbb{E}_{\mathcal{C}}[\kappa(i,j)] \overset{O}{=} \frac{1}{n(n-1)} \cdot \frac{n(n-1)}{2} \cdot 2n \overset{O}{=} n.$$

$\square$

### 5.1.3 Proof of Theorem 5.1.1 (iii)

*Proof of* $\mathbb{E}_{\mathcal{C}}\big[K(\mathcal{G})^2\big]$. Observe that by (2.10) we have

$$\mathbb{E}_{\mathcal{C}}\big[K(\mathcal{G})^2\big] = \sum_{\{i,j\}\subseteq V}\sum_{\{w,z\}\subseteq V} \mathbb{E}_{\mathcal{C}}[R(i,j)R(w,z)]\,. \tag{5.9}$$

For each pair $\{i,j\},\{w,z\} \subset V$ define the following disjoint events

$$\mathcal{C}_1 := \mathcal{A}^n_{i,j} \cap \mathcal{A}^n_{w,z} \cap \mathcal{B}_{i,j} \cap \mathcal{B}_{w,z}, \qquad \mathcal{C}_2 := (\mathcal{C}_1)^c\,.$$

Let $\mathfrak{E} := \mathbb{E}_{\mathcal{C}}[R(i,j)R(w,z)\mathbf{1}_{\mathcal{C}_1}]$. The effective resistance bound from Theorem 3.4.1 yields

$$\begin{aligned}
\mathfrak{E} &\le \mathbb{E}_{\mathcal{C}}\left[\prod_{\substack{(x,y)\in\\ \{(i,j),(w,z)\}}}\left(\frac{1}{\psi_1(x)}+\frac{1}{\psi_1(y)}+\sum_{b\in\{x,y\}}\sum_{a\in\Psi_1(b)}\frac{k^*}{\psi_1(b)^2\varphi_1(a)}\right)\mathbf{1}_{\mathcal{C}_1}\right]\\
&\le \sum_{\substack{x\in\{i,j\}\\ y\in\{w,z\}}}\left(\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}^{i,j}_x}\mathbf{1}_{\mathcal{B}^{w,z}_y}}{\psi_1(x)\psi_1(y)}\right]+\sum_{\substack{f,g\in\{x,y\}\\ f\neq g}}k^*\cdot\mathbb{E}_{\mathcal{C}}\left[\sum_{a\in\Psi_1(f)}\frac{\mathbf{1}_{\mathcal{C}_1}}{\psi_1(f)^2\psi_1(g)\varphi_1(a)}\right]\right)\\
&\quad + (k^*)^2\cdot\sum_{\substack{x\in\{i,j\}\\ y\in\{w,z\}}}\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{a\in\Psi_1(x)}\frac{\mathbf{1}_{\mathcal{C}_1}}{\psi_1(x)^2\varphi_1(a)}\right)\left(\sum_{a\in\Psi_1(y)}\frac{\mathbf{1}_{\mathcal{C}_1}}{\psi_1(y)^2\varphi_1(a)}\right)\right].
\end{aligned}$$

By removing $\sup 1/\varphi(a)$ from the sums over $a\in\Psi_1(x),\Psi_1(y)$ and by symmetry we have

$$\begin{aligned}
\mathfrak{E} &\le 4\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}^{w,z}_w}\mathbf{1}_{\mathcal{B}^{i,j}_i}}{\psi_1(i)\psi_1(w)}\right]+8k^*\cdot\mathbb{E}_{\mathcal{C}}\left[\sup_{a\in\Psi_1(i)}\frac{\mathbf{1}_{\mathcal{C}_1}}{\varphi_1(a)\psi_1(i)\psi_1(w)}\right]\\
&\quad + 4(k^*)^2\cdot\mathbb{E}_{\mathcal{C}}\left[\left(\sup_{a\in\Psi_1(i)}\frac{1}{\varphi_1(a)}\right)\left(\sup_{b\in\Psi_1(w)}\frac{1}{\varphi_1(a)}\right)\frac{\mathbf{1}_{\mathcal{C}_1}}{\psi_1(i)\psi_1(w)}\right].
\end{aligned}$$

Then applying Hölder's inequality (2.2) and substituting like terms yields

$$\begin{aligned}\mathfrak{E} \le 4\,\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\psi_1(i)^2}\right] + 8k^* \cdot \mathbb{E}_{\mathcal{C}}\left[\sup_{a\in\Psi_1(i)} \frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\varphi_1(a)^3}\right]^{1/3} \mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\psi_1(i)^3}\right]^{2/3} \\ + 4(k^*)^2 \cdot \mathbb{E}_{\mathcal{C}}\left[\sup_{a\in\Psi_1(i)} \frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\varphi_1(a)^4}\right]^{1/2} \mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\psi_1(i)^4}\right]^{1/2}.\end{aligned}$$

Now applying the estimates in Lemma 5.0.1 to the expectations above we obtain

$$\begin{aligned}\mathfrak{E} &\le 4\left(\frac{1}{(np)^2} + O\left(\frac{\log n}{(np)^3\log(np)}\right)\right) + 8\cdot O\left(\frac{\log n}{(np)^3\log(np)}\right) \\ &\quad + 4\cdot O\left(\left(\frac{\log n}{(np)^2\log(np)}\right)^2\right) \\ &= \frac{4}{(np)^2} + O\left(\frac{\log n}{(np)^3\log(np)}\right).\end{aligned}$$

When $np \ge c\log n$ and $c > 3$ we have the following for expectation on $\mathcal{C}_2$ by first applying the effective resistance bound (5.5) then using Lemma 4.2.1 to bound from above the results of applying the union bound to $\mathbb{P}(\mathcal{C}_2) = \mathbb{P}(\mathcal{C}_1^c)$:

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[R(i,j)R(w,z)\mathbf{1}_{\mathcal{C}_2}] &\le \frac{(n-1)^2}{\mathbb{P}(\mathcal{C})}\mathbb{P}(\mathcal{C}_2) \\ &\le n^2\left(2e^{-(1-o(1))np} + o\left(1/n^{7/2}\right)\right) \\ &= o\left(1/n\right).\end{aligned}$$

If $\log n + \log\log\log n \le np \le 3\log n$ then we further partition using $\mathcal{S}_{i,j}$ from (5.4) to obtain

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left[R(i,j)R(w,z)\mathbf{1}_{\mathcal{C}_2}\left(\mathbf{1}_{\mathcal{S}_{i,j}\cap\mathcal{S}_{w,z}} + \mathbf{1}_{(\mathcal{S}_{i,j}\cap\mathcal{S}_{w,z})^c}\right)\right] \\ &\quad\le \left(3\log(n)/\log(np)\right)^2 \mathbb{P}(\mathcal{C}_2)/\mathbb{P}(\mathcal{C}) + 2(n-1)^2\mathbb{P}_{\mathcal{C}}\left(\mathcal{S}_{i,j}^c\right) \\ &\quad\le O\left((\log n)^2 e^{-(1-o(1))np}\right) + n^2 o(1/n^4) \\ &\quad= o\left(1/n^{4/5}\right).\end{aligned}$$

Combining expectations over $\mathcal{C}_1$ and $\mathcal{C}_2$ gives the upper bound on $\mathbb{E}_{\mathcal{C}}[R(i,j)R(w,z)]$.

Let $\mathcal{D}$ be the event $\{\gamma_1(i), \gamma_1(j), \gamma_1(w), \gamma_1(z) \le np + a\sqrt{np}\}$ where $a = 3\sqrt{\log\log n}$ if $np = O(\log n)$ and $a = 3\sqrt{\log n}$ if $np = \omega(\log n)$. By Lemma 3.1.3 and

$1 \geq \mathbf{1}_{\mathcal{D}}$ we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[R(i,j)R(w,z)] &\geq \sum_{x\in\{i,j\},\, y\in\{w,z\}} \mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{D}}}{(\gamma_1(x)+1)\,(\gamma_1(y)+1)}\right] \\ &\geq \frac{4}{\left(np+a\sqrt{np}\right)^2}\mathbb{P}_{\mathcal{C}}(\mathcal{D})\end{aligned}$$

for $i \neq j, w \neq z$. By applying the bound $\mathbb{P}_{\mathcal{C}}(\mathcal{D}^c) \leq \mathbb{P}(\mathcal{D}^c)\,/\mathbb{P}(\mathcal{C})$ and and then using the Chernoff bound, Lemma 2.1.1, to bound $\mathbb{P}(\mathcal{D}^c)$ from above we obtain

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[R(i,j)R(w,z)] &\geq \left(\frac{4}{np} - O\left(\frac{a}{(np)^{3/2}}\right)\right)^2\left(1-\frac{4}{\mathbb{P}(\mathcal{C})}e^{-a^2/3}\right) \\ &\geq \frac{4}{(np)^2} - O\left(\frac{\log n}{(np)^2\log np}\right).\end{aligned}$$

The result follows from the above bounds and (5.9). ☐

*Proof of* $\mathbb{E}_{\mathcal{C}}\left[h(i,j)^2\right]$. Let $g(a,b,c,d) := \mathbb{E}_{\mathcal{C}}[\gamma_1(u)\gamma_1(v)R(a,b)R(c,d)]$, if we apply Tetali's formula (2.14) and expand out $\mathbb{E}_{\mathcal{C}}[h(i,j)h(i,a)]$ we obtain the following for any $i,j,a \in V$:

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[h(i,j)h(i,a)] &= \mathbb{E}_{\mathcal{C}}\left[\left(\sum_{u\in V}\frac{\gamma_1(u)}{2}\left(R(i,j)+R(j,u)-R(u,i)\right)\right)\right. \\ &\qquad\left.\cdot\left(\sum_{v\in V}\frac{\gamma_1(v)}{2}\left(R(i,a)+R(a,v)-R(v,i)\right)\right)\right] \\ &= \frac{1}{4}\sum_{u,v\in V}\left(g(i,j,i,a) + \sum_{\substack{(w,z)\in\\ \{(u,i),(j,a)\}}} g(i,w,v,z) - \sum_{w\in\{i,u\}} g(w,j,i,v)\right) \\ &\quad + \frac{1}{4}\sum_{u,v\in V}\sum_{w\in\{i,v\}}\left(g(u,j,w,a) - g(w,a,i,u)\right) \\ &= \frac{1}{4}\sum_{u,v\in V}\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\gamma_1(v)R(i,j)R(i,a)]\,. \end{aligned} \tag{5.10}$$

To see the above, observe that $R(a,b)R(c,d) = 0$ if and only $a = b$ or $c = d$. Thus only the first term, $g(i,j,i,a)$, will always be non-zero. All the other terms contain one or more input from $\{u,v\}$ so will be zero at different times. Of the eight other terms there are two positive and two negative terms containing one of $\{u,v\}$, then two positive and two negative terms containing both $u$ and $v$ as inputs.

Thus by symmetry when the sums are expanded everything apart from the first term $g(i,j,i,a)$ cancels.

For $(u,v,i,j,w,z) \in V^6$ with $i \neq j,\ w \neq z$ define the event

$$\mathcal{M}_{u,v} := \{\gamma_1(v), \gamma_1(u) \leq 7np\}.$$

We define the following disjoint events

$$\mathcal{C}_1 := \mathcal{A}^n_{i,j} \cap \mathcal{A}^n_{w,z} \cap \mathcal{B}_{i,j} \cap \mathcal{B}_{w,z}, \quad \mathcal{C}_2 := \mathcal{C}_1^c \cap \mathcal{M}_{u,v}, \quad \mathcal{C}_3 := \mathcal{C}_1^c \cap \mathcal{M}_{u,v}^c.$$

Recall $k^* = \log(n)/\log(np)$ and that conditional on $\mathcal{C}_1$, for large $n$, we have $k + 2 \leq k^*$, where $k$ is as in Theorem 3.4.1. Let $\mathfrak{E} := \mathbb{E}_{\mathcal{C}}[\gamma_1(u)\gamma_1(v)R(i,j)R(w,z)\mathbf{1}_{\mathcal{C}_1}]$. Applying Theorem 3.4.1 yields

$$\begin{aligned}
\mathfrak{E} &\leq \mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)\gamma_1(v) \prod_{\substack{(x,y)\in \\ \{(i,j),(w,z)\}}} \left(\frac{1}{\psi_1(x)} + \frac{1}{\psi_1(y)} + \sum_{b\in\{x,y\}} \sum_{b_a\in\Psi_1(b)} \frac{k^*}{\psi_1(b)^2\varphi_1(b_a)}\right)\mathbf{1}_{\mathcal{C}_1}\right] \\
&\leq \sum_{\substack{x\in\{i,j\}\\ y\in\{w,z\}}} \mathbb{E}_{\mathcal{C}}\left[\frac{\gamma_1(u)\gamma_1(v)\mathbf{1}_{\mathcal{B}_{i,j}\cap\mathcal{B}_{w,z}}}{\psi_1(x)\psi_1(y)}\right] \\
&\quad + k^* \cdot \sum_{\substack{x\in\{i,j\}\\ y\in\{w,z\}}} \sum_{\substack{f,g\in\{x,y\}\\ f\neq g}} \mathbb{E}_{\mathcal{C}}\left[\sum_{a\in\Psi_1(f)} \frac{\gamma_1(u)\gamma_1(v)\mathbf{1}_{\mathcal{C}_1}}{\psi_1(f)^2\psi_1(g)\varphi_1(a)}\right] \\
&\quad + (k^*)^2 \cdot \sum_{\substack{x\in\{i,j\}\\ y\in\{w,z\}}} \mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)\gamma_1(v)\left(\sum_{a\in\Psi_1(x)} \frac{\mathbf{1}_{\mathcal{C}_1}}{\psi_1(x)^2\varphi_1(a)}\right)\left(\sum_{b\in\Psi_1(y)} \frac{\mathbf{1}_{\mathcal{C}_1}}{\psi_1(y)^2\varphi_1(b)}\right)\right].
\end{aligned}$$

By removing $\sup 1/\varphi(a)$ from the sums and reducing using symmetry we have

$$\begin{aligned}
\mathfrak{E} &\leq 4\mathbb{E}_{\mathcal{C}}\left[\frac{\gamma_1(u)\gamma_1(v)\mathbf{1}_{\mathcal{B}_i^{i,j}}\mathbf{1}_{\mathcal{B}_w^{w,z}}}{\psi_1(i)\psi_1(w)}\right] + 8k^* \cdot \mathbb{E}_{\mathcal{C}}\left[\sup_{a\in\psi_1(i)} \frac{\gamma_1(u)\gamma_1(v)\mathbf{1}_{\mathcal{C}_1}}{\varphi_1(a)\psi_1(i)\psi_1(w)}\right] \\
&\quad + 4(k^*)^2 \cdot \mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)\gamma_1(v)\left(\sup_{a\in\psi_1(i)} \frac{1}{\varphi_1(a)}\right)\left(\sup_{b\in\psi_1(w)} \frac{1}{\varphi_1(a)}\right)\frac{\mathbf{1}_{\mathcal{C}_1}}{\psi_1(x)\psi_1(y)}\right].
\end{aligned}$$

Then applying Hölder's inequality (2.2) and collecting similar terms we obtain

$$\begin{aligned}\mathfrak{E} \le\, & 4\,\mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)^4\right]^{1/2}\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\psi_1(i)^4}\right]^{1/2} \\ & + 8k^* \cdot \mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)^5\right]^{2/5}\mathbb{E}_{\mathcal{C}}\left[\sup_{a\in\Psi_1(i)}\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\varphi_1(a)^5}\right]^{1/5}\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\psi_1(i)^5}\right]^{2/5} \\ & + 4(k^*)^2 \cdot \mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)^6\right]^{1/3}\mathbb{E}_{\mathcal{C}}\left[\sup_{a\in\Psi_1(i)}\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\varphi_1(a)^6}\right]^{1/3}\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_i^{i,j}}}{\psi_1(i)^6}\right]^{1/3}.\end{aligned}$$

Now applying the estimates in Lemma 5.0.1 to the expectations above yields

$$\begin{aligned}\mathfrak{E} &\le 4\left(1 + O\left(\frac{\log n}{np\log(np)}\right)\right) + 8\cdot O\left(\frac{\log n}{np\log(np)}\right) + 4\cdot O\left(\left(\frac{\log n}{np\log(np)}\right)^2\right) \\ &= 4 + O\left(\frac{\log n}{np\log(np)}\right).\end{aligned}$$

For $\mathcal{C}_2 := \mathcal{C}_1^c \cap \mathcal{M}$ and $np \ge 3\log n$ we apply the effective resistance bound (5.5) and $\gamma_1(u)\mathbf{1}_{\mathcal{M}}, \gamma_1(v)\mathbf{1}_{\mathcal{M}} \le 7np$, then bound $\mathbb{P}(\mathcal{C}_1^c)$ by Lemma 4.2.1 yielding

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\gamma_1(v)R(i,j)R(w,z)\mathbf{1}_{\mathcal{C}_2}] &\le \frac{7(n-1)^4p^2}{\mathbb{P}(\mathcal{C})}\mathbb{P}(\mathcal{C}_1^c) \\ &\le 7n^4p^2 \cdot e^{-(1-o(1))np} \\ &= o\left(1/n^{4/5}\right).\end{aligned}$$

If $\log n + \log\log\log n \le np \le 3\log n$ we further partition using $\mathcal{S}_{i,j}, \mathcal{S}_{w,z}$ from (5.4) to obtain

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)\gamma_1(v)R(i,j)R(w,z)\mathbf{1}_{\mathcal{C}_2}\left(\mathbf{1}_{\mathcal{S}_{i,j}\cap\mathcal{S}_{w,z}} + \mathbf{1}_{(\mathcal{S}_{i,j}\cap\mathcal{S}_{w,z})^c}\right)\right] \\ &\quad\le (7np)^2\left(3\log n/\log np\right)^2\mathbb{P}(\mathcal{C}_1^c)/\mathbb{P}(\mathcal{C}) + 7n^4p^2\left(\mathbb{P}_{\mathcal{C}}\left(\mathcal{S}_{i,j}^c\right) + \mathbb{P}_{\mathcal{C}}\left(\mathcal{S}_{w,z}^c\right)\right) \\ &\quad= o\left(1/n^{4/5}\right).\end{aligned}$$

Since $\mathbb{P}_{\mathcal{C}}\left(\mathcal{M}^c\right) \le \exp\left(-3\cdot 6^2 np/18\right)/\mathbb{P}(\mathcal{C}) = o(1/n^6)$ by Lemma 2.1.1 we have

$$\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\gamma_1(v)R(i,j)R(w,z)\mathbf{1}_{\mathcal{C}_3}] \le (n-1)^4\mathbb{P}_{\mathcal{C}}(\mathcal{M}^c) = o(n^{-2}).$$

Combining expectations over $\mathcal{C}_1$ ,$\mathcal{C}_2$ & $\mathcal{C}_3$ gives

$$\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\gamma_1(v)R(i,j)R(w,z)] \le 4 + O\left(\frac{\log n}{np\log(np)}\right). \tag{5.11}$$

Let $\mathcal{D}$ be the event $\{\gamma_1(u), \gamma_1(v) \geq np - a\sqrt{np}\} \cap \{\gamma_1(i), \gamma_1(j), \gamma_1(w), \gamma_1(z) \leq np + a\sqrt{np}\}$ where $a = 3\sqrt{\log\log n}$ if $np = O(\log n)$ and $a = 3\sqrt{\log n}$ if $np = \omega(\log n)$. By Lemma 3.1.3:

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}[\gamma_1(u)\gamma_1(v)R(i,j)R(w,z)] &\geq \sum_{\substack{x\in\{i,j\},\\ y\in\{w,z\}}} \mathbb{E}_{\mathcal{C}}\left[\frac{\gamma_1(u)\gamma_1(v)\mathbf{1}_{\mathcal{D}}}{(\gamma_1(x)+1)\,(\gamma_1(y)+1)}\right] \\
&\geq 4\frac{\left(np - a\sqrt{np}\right)^2}{\left(np + a\sqrt{np}\right)^2}\mathbb{P}_{\mathcal{C}}(\mathcal{D}) \\
&\geq \left(4 - O\left(\frac{a}{\sqrt{np}}\right)\right)\left(1 - \frac{2}{\mathbb{P}(\mathcal{C})}e^{-\frac{a^2}{2}} - \frac{4}{\mathbb{P}(\mathcal{C})}e^{-\frac{a^2}{3}}\right) \\
&= 4 - O\left(\frac{\log n}{np\log np}\right), \qquad (5.12)
\end{aligned}$$

for $i \neq j, w \neq z$. The bound on $\mathbb{P}_{\mathcal{C}}(\mathcal{D})$ is by Lemma 2.1.1. Combining (5.10)–(5.12) yields

$$\mathbb{E}_{\mathcal{C}}[h(i,j)h(i,a)] = n^2 \pm O\left(\frac{n\log n}{p\log(np)}\right), \qquad (5.13)$$

for any $i, j, w, z \in V,\ i \neq j, w \neq z$. Thus we have the result for $\mathbb{E}_{\mathcal{C}}\left[h(i,j)^2\right]$. □

*Proof of* $\mathbb{E}_{\mathcal{C}}\left[cc_i(\mathcal{G})^2\right]$. This follows from (5.13) above as by the definition (2.11) of $cc_i(G)$,

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left[cc_i(\mathcal{G})^2\right] &= \frac{1}{(n-1)^2}\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{j\in V} h(i,j)\right)^2\right] \\
&= \frac{1}{(n-1)^2}\sum_{j,k\in V; j,k\neq i} \mathbb{E}_{\mathcal{C}}[h(i,j)h(i,k)] \\
&\overset{O}{=} n^2.
\end{aligned}$$

□

### 5.1.4 Proof of Theorem 5.1.1 (iv)

*Proof of* $\mathbb{E}_{\mathcal{C}}[H(\mathcal{G})], \mathbb{E}_{\mathcal{C}}[H_i(\mathcal{G})], \mathbb{E}_{\mathcal{C}}[T(\mathcal{G})]$. Recall the definitions (2.8),(2.9) for $i \in V$:

$$H_i(G) := \sum_{j\in V} \frac{\gamma_1(j)}{2m} h(j,i),$$
$$H(G) := \sum_{j\in V} \frac{\gamma_1(j)}{2m} h(i,j),$$
$$T(G) := \sum_{i,j\in V} \frac{\gamma_1(i)\gamma_1(j)}{4m^2} h(i,j),$$

where $m := |E| \sim_d Bin\left(\binom{n}{2}, p\right)$. Let $h = \binom{n}{2} - 1$, $m^* \sim_d Bin\,(h,p)$. Then we have the following for any given $k \in \mathbb{Z},\ k \geq 1$ using Proposition 2.4.2 and the fact that $\mathcal{C} \subset \{m \geq 1\}$ we have

$$\mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^k}\right] = \mathbb{E}\left[\frac{\mathbf{1}_{\mathcal{C}}}{m^k}\right]\frac{1}{\mathbb{P}(\mathcal{C})} \leq \mathbb{E}\left[\frac{\mathbf{1}_{\{m\geq 1\}}}{m^k}\right]\frac{1}{\mathbb{P}(\mathcal{C})} = \mathbb{E}\left[\frac{\binom{n}{2}p}{(m^*+1)^{k+1}}\right]\frac{1}{\mathbb{P}(\mathcal{C})}.$$

Observe that by (2.15), $\mathbb{P}(\mathcal{C}^c) \leq O(\log n/(np\log(np)))$ whenever $np \geq \log n + \log\log\log n$. Using Lemma 2.4.3 to bound the expectation term we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^k}\right] &= \left(\frac{1}{(hp+1)^{k+1}} + O\left(\frac{1}{(hp+1)^{k+2}}\right)\right)\frac{\binom{n}{2}p}{\mathbb{P}(\mathcal{C})} \\ &= \frac{2^k}{n^{2k}p^k} + O\left(\frac{\log n}{n^{2k+1}p^{k+1}\log(np)}\right).\end{aligned}$$

Now by the Bernoulli inequality (2.3) for any given $a, k \in \mathbb{Z},\ a,k \geq 1$ we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^k}\right]^{1/a} &= \frac{2^{k/a}}{n^{2k/a}p^{k/a}}\left(1 + O\left(\frac{\log n}{np\log(np)}\right)\right)^{1/a} \\ &\leq \frac{2^{k/a}}{n^{2k/a}p^{k/a}} + O\left(\frac{\log n}{n^{2k/a+1}p^{k/a+1}\log(np)}\right).\end{aligned} \tag{5.14}$$

Using Hölder's inequality to break the product of random variables in the expectation yields

$$\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})] \leq \frac{1}{4}\sum_{i,j\in V} \mathbb{E}_{\mathcal{C}}\left[\gamma_1(i)^6\right]^{1/6} \mathbb{E}_{\mathcal{C}}\left[\gamma_1(j)^6\right]^{1/6} \mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^{12}}\right]^{1/6} \mathbb{E}_{\mathcal{C}}\left[h(i,j)^2\right]^{1/2}.$$

Then applying (2.16), (5.14) and the upper bound on $\mathbb{E}_{\mathcal{C}}\big[h(i,j)^2\big]$ from Theorem 5.1.1 (iii) we obtain

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})] &\leq \frac{n^2}{4}\big(n^2p^2 + O(np)\big)\left(\frac{4}{n^4p^2} + O\left(\frac{\log n}{n^5p^3\log(np)}\right)\right)\left(n + O\left(\frac{\log n}{p\log(np)}\right)\right) \\ &= n + O\left(\frac{\log n}{p\log(np)}\right).\end{aligned}$$

The same upper bounds for $\mathbb{E}_{\mathcal{C}}[H_i(\mathcal{G})]$ and $\mathbb{E}_{\mathcal{C}}[H(\mathcal{G})]$ follow similarly. By (2.14) we have

$$\begin{aligned}T(G) &= \sum_{i,j\in V}\frac{\gamma_1(i)\gamma_1(j)}{4m^2}\left(mR(i,j) + \sum_{u\in V}\frac{\gamma_1(u)}{2}\left[R(u,j) - R(u,i)\right]\right) \\ &= \sum_{i,j\in V}\frac{\gamma_1(i)\gamma_1(j)R(i,j)}{4m} \\ &\quad + \sum_{i,j,u\in V}\left(\frac{\gamma_1(i)\gamma_1(j)\gamma_1(u)}{8m^2}R(u,j) - \frac{\gamma_1(i)\gamma_1(j)\gamma_1(u)}{8m^2}R(u,i)\right),\end{aligned}$$

for $G$ connected. As $G$ is connected the effective resistance bound, Lemma 3.1.3, yields

$$\begin{aligned}T(G) \geq &\sum_{\substack{i,j\in V\\ i\neq j}}\sum_{w\in\{i,j\}}\frac{\gamma_1(i)\gamma_1(j)}{4m(\gamma_1(w)+1)} + \sum_{\substack{j,u\in V\\ j\neq u}}\sum_{w\in\{u,j\}}\frac{\gamma_1(j)\gamma_1(u)}{4m(\gamma_1(w)+1)} \\ &- \sum_{i,u\in V}\frac{\gamma_1(i)\gamma_1(u)}{4m}R(u,i).\end{aligned}$$

Rearranging and reducing sums using the bound $\gamma_1(i)/(\gamma_1(i)+1) \leq 1$ we have

$$\begin{aligned}T(G) &\geq \sum_{i\in V}\left(\frac{\gamma_1(i)}{2(\gamma_1(i)+1)} - \frac{\gamma_1(i)}{4m}\right) + \sum_{j\in V}\left(\frac{\gamma_1(j)}{\gamma_1(j)+1} - \frac{\gamma_1(j)}{2m}\right) \\ &\quad + \sum_{u\in V}\left(\frac{\gamma_1(u)}{2(\gamma_1(u)+1)} - \frac{\gamma_1(u)}{4m}\right) - \sum_{i,u\in V}\frac{\gamma_1(i)\gamma_1(u)}{4m}R(u,i) \\ &= 2n - 2\sum_{i\in V}\frac{1}{\gamma_1(i)+1} - 2 - \sum_{i,u\in V}\frac{\gamma_1(i)\gamma_1(u)}{4m}R(u,i).\end{aligned}$$

Manipulating the sums and bounding terms in a similar manner yields

$$\begin{aligned}
H(G) &= \sum_{j\in V} \frac{\gamma_1(j)}{2m} \left( mR(i,j) + \sum_{u\in V} \frac{\gamma_1(u)}{2} \left[R(u,j) - R(u,i)\right] \right) \\
&\geq \frac{n-1}{2} - \sum_{j\in V} \frac{1}{2(\gamma_1(j)+1)} + \frac{m-1}{(\gamma_1(i)+1)} - 1 \\
&\quad + \sum_{\substack{j,u\in V \\ j\neq u}} \sum_{w\in\{u,j\}} \frac{\gamma_1(j)\gamma_1(u)}{4m(\gamma_1(w)+1)} - \sum_{j,u\in V} \frac{\gamma_1(u)\gamma_1(j)}{4m} R(u,i) \\
&\geq \frac{3n}{2} + \frac{m-1}{\gamma_1(i)+1} - \sum_{u\in V} \frac{3}{2(\gamma_1(u)+1)} - \frac{5}{2} - \sum_{u\in V} \frac{\gamma_1(u)}{2} R(u,i).
\end{aligned}$$

Again by a similar procedure we have the following for the stationary hitting time $H_i(G)$

$$\begin{aligned}
H_i(G) &= \sum_{j\in V} \frac{\gamma_1(j)}{2m} \left( mR(i,j) + \sum_{u\in V} \frac{\gamma_1(u)}{2} \left[R(u,i) - R(u,j)\right] \right) \\
&\geq \frac{n-1}{2} - \sum_{j\in V} \frac{1}{2(\gamma_1(j)+1)} + \frac{m-1}{(\gamma_1(i)+1)} - 1 \\
&\quad + \sum_{u\in V, u\neq i} \frac{\gamma_1(u)}{2} \left( \frac{1}{\gamma_1(i)+1} + \frac{1}{\gamma_1(u)+1} \right) - \sum_{j,u\in V} \frac{\gamma_1(u)\gamma_1(j)}{4m} R(u,j) \\
&\geq n + \frac{2m-2}{\gamma_1(i)+1} - \sum_{u\in V} \frac{1}{\gamma_1(u)+1} - \frac{7}{2} - \sum_{j,u\in V} \frac{\gamma_1(u)\gamma_1(j)}{4m} R(u,j).
\end{aligned}$$

Let $\mathcal{D}$ be the event $\{m \geq n^2p/2 - a\sqrt{n^2p/2}\} \cap \{\gamma_1(j) \leq np + a\sqrt{np}\}$ where $a = 3\sqrt{\log\log n}$ if $np = O(\log n)$ and $a = 3\sqrt{\log n}$ if $np = \omega(\log n)$. Now by Lemma 2.1.1 we obtain

$$\begin{aligned}
\mathbb{P}_{\mathcal{C}}(\mathcal{D}) &= \left(1 - \exp\left(-a^2/2\right)/\mathbb{P}(\mathcal{C}) - \exp\left(-a^2/2(1 + a/3\sqrt{np})\right)/\mathbb{P}(\mathcal{C})\right) \\
&= 1 - o\left(1/np\right).
\end{aligned}$$

By Hölder's inequality (2.2), $1 \geq \mathbf{1}_{\mathcal{D}}$ and the bound on $\mathbb{P}_{\mathcal{C}}(\mathcal{D})$ in the line above we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}[H_i(\mathcal{G})] &\geq n + 2\frac{\binom{n}{2}p - a\sqrt{\binom{n}{2}p} - 1}{np + a\sqrt{np} + 1}\mathbb{P}_{\mathcal{C}}(\mathcal{D}) - n \cdot \mathbb{E}_{\mathcal{C}}\left[\frac{1}{\gamma_1(u)+1}\right] - \frac{7}{2} \\
&\quad - \frac{n}{4}\mathbb{E}_{\mathcal{C}}\left[\gamma_1(j)^4\right]^{1/4}\mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^4}\right]^{1/4}\mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)^2 R(u,j)^2\right]^{1/2} \\
&= n - O\left(\frac{\log n}{p\log(np)}\right).
\end{aligned}$$

The last equality comes from applying estimates to the expectation terms which are given by Lemma 2.4.3, (2.16), (5.14) and (5.11) respectively. Similarly we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}[H(\mathcal{G})] &\geq \frac{3n}{2} + \frac{\binom{n}{2}p - a\sqrt{\binom{n}{2}p} - 1}{np + a\sqrt{np} + 1}\mathbb{P}_{\mathcal{C}}(\mathcal{D}) - \frac{3n}{2} \cdot \mathbb{E}_{\mathcal{C}}\left[\frac{1}{\gamma_1(u)+1}\right] - \frac{5}{2} \\
&\quad - \frac{n}{2}\mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)^2 R(u,i)^2\right]^{1/2} \\
&= n - O\left(\frac{\log n}{p\log(np)}\right),
\end{aligned}$$

and also,

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})] &\geq 2n - 2n \cdot \mathbb{E}_{\mathcal{C}}\left[\frac{1}{\gamma_1(i)+1}\right] - 2 \\
&\quad - \frac{n}{4}\mathbb{E}_{\mathcal{C}}\left[\gamma_1(i)^4\right]^{1/4}\mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^4}\right]^{1/4}\mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)^2 R(u,i)^2\right]^{1/2} \\
&= n - O\left(\frac{\log n}{p\log(np)}\right).
\end{aligned}$$

□

### 5.1.5 Proof of Theorem 5.1.1 (v)

*Proof of* $\mathbb{E}_{\mathcal{C}}\left[H(\mathcal{G})^2\right], \mathbb{E}_{\mathcal{C}}\left[H_i(\mathcal{G})^2\right], \mathbb{E}_{\mathcal{C}}\left[T(\mathcal{G})^2\right]$. We will first bound $\mathbb{E}_{\mathcal{C}}\left[h(i,j)^3\right]$ from above. By Tetali's formula (2.14) we obtain the following for any $i, j, a \in V$

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[h(i,j)^3\right] &= \mathbb{E}_{\mathcal{C}}\left[\prod_{w\in\{x,y,z\}}\left(\sum_{w\in V}\frac{\gamma_1(w)}{2}\left(R(i,j)+R(j,w)-R(w,i)\right)\right)\right]\\ &= \frac{1}{8}\sum_{x,y,z\in V}\mathbb{E}_{\mathcal{C}}\left[\gamma_1(x)\gamma_1(y)\gamma_1(z)R(i,j)^3\right]. \qquad (5.15)\end{aligned}$$

Similarly to (5.10) when the product is expanded everything apart from the only term with effective resistances not dependent on the indices of summation cancels. There are three positive and three negative terms containing one of $\{x, y, z\}$, then six positive and six negative terms containing two of $\{x, y, z\}$, finally four positive and four negative terms containing all three indices $\{x, y, z\}$. When the sum over $x, y, z$ is taken all the terms containing at least one of $x, y, z$ cancel.

For each $(x, y, z) \in V^3$ let $\mathcal{M}_{x,y,z}$ be the event $\{\gamma_1(x), \gamma_1(y), \gamma_1(z) \leq 8np\}$ and define the following disjoint events

$$\mathcal{C}_1 := \mathcal{A}^n_{i,j}\cap\mathcal{B}_{i,j}, \quad \mathcal{C}_2 := \mathcal{C}_1^c\cap\mathcal{M}_{x,y,z}, \quad \mathcal{C}_3 := \mathcal{C}_1^c\cap\mathcal{M}_{x,y,z}^c.$$

Let $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ be a multi-index and $\binom{3}{\alpha} = 3/\alpha_1!\alpha_2!\alpha_3!\alpha_4!$ be the multinomial coefficient. Let $\mathfrak{E} := \mathbb{E}_{\mathcal{C}}\left[\gamma_1(x)\gamma_1(y)\gamma_1(z)R(i,j)^3\mathbf{1}_{\mathcal{C}_1}\right]$. Applying the resistance bound, Theorem 3.4.1, to $\mathfrak{E}$ yields the following

$$\begin{aligned}\mathfrak{E} &\leq \mathbb{E}_{\mathcal{C}}\left[\gamma_1(x)\gamma_1(y)\gamma_1(z)\left(\frac{1}{\psi_1(i)}+\frac{1}{\psi_1(j)}+\sum_{b\in\{i,j\}}\sum_{a\in\Psi_1(b)}\frac{k^*}{\psi_1(b)^2\varphi_1(a)}\right)^3\mathbf{1}_{\mathcal{C}_1}\right]\\ &\leq \sum_{|\alpha|=3}\binom{3}{\alpha}\mathbb{E}_{\mathcal{C}}\left[\frac{\gamma_1(x)\gamma_1(y)\gamma_1(z)}{\psi(i)^{\alpha_1}\psi(j)^{\alpha_2}}\left(\sum_{a\in\Psi_1(i)}\frac{k^*}{\psi_1(i)^2\varphi_1(a)}\right)^{\alpha_3}\left(\sum_{a\in\Psi_1(j)}\frac{k^*}{\psi_1(j)^2\varphi_1(a)}\right)^{\alpha_4}\mathbf{1}_{\mathcal{C}_1}\right].\end{aligned}$$

Again, by taking supremums to remove the random sum in each of the last three terms and then applying Hölder's inequality to all the terms as was done for (5.6) we obtain

$$\begin{aligned}\mathfrak{E} &\leq 8\left(1+O\left(\frac{\log n}{np\log(np)}\right)\right)+24\cdot O\left(\frac{\log n}{np\log(np)}\right)+24\cdot O\left(\left(\frac{\log n}{np\log(np)}\right)^2\right)\\ &\quad+8\cdot O\left(\left(\frac{\log n}{np\log(np)}\right)^3\right)=8+O\left(\frac{\log n}{np\log(np)}\right).\end{aligned}$$

For $\mathcal{C}_2$ and $np \geq 4\log n$ applying the bound $\gamma_1(x), \gamma_1(v), \gamma_1(z) \leq 7np$ on $\mathcal{M}_{x,y,z} \subseteq \mathcal{C}_2$, the effective resistance bound (5.5) and then bounding $\mathbb{P}(\mathcal{C}_1^c)$ by Lemma 4.2.1 yields

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[\gamma_1(x)\gamma_1(y)\gamma_1(z)R(i,j)^3\mathbf{1}_{\mathcal{C}_2}\right] &\leq \frac{8(n-1)^6p^3}{\mathbb{P}(\mathcal{C})}\mathbb{P}(\mathcal{C}_1^c)\\ &\leq 8n^6p^3\cdot o\left(1/n^{7/2}\right)\\ &= o\left(1/n^{1/5}\right).\end{aligned}$$

If $\log n + \log\log\log n \leq np \leq 4\log n$ then we further partition using $\mathcal{S}_{i,j}$ from (5.4) to obtain

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left[\gamma_1(x)\gamma_1(y)\gamma_1(z)R(i,j)^3\mathbf{1}_{\mathcal{C}_2}\left(\mathbf{1}_{\mathcal{S}_{i,j}} + \mathbf{1}_{(\mathcal{S}_{i,j})^c}\right)\right]\\ &\quad\leq (8np)^3\left(3\log n/\log np\right)^3\mathbb{P}(\mathcal{C}_1^c)/\mathbb{P}(\mathcal{C}) + 8n^6p^3\mathbb{P}_{\mathcal{C}}\left(\mathcal{S}_{i,j}^c\right)\\ &\quad= o\left(1/n^{4/5}\right).\end{aligned}$$

Since $\mathbb{P}_{\mathcal{C}}\left(\mathcal{M}_{x,y,z}^c\right) \leq \exp\left(-3\cdot 7^2np/20\right)/\mathbb{P}(\mathcal{C}) = o(1/n^7)$ by Lemma 2.1.1 we have

$$\mathbb{E}_{\mathcal{C}}\left[\gamma_1(u)\gamma_1(v)R(i,j)^3\mathbf{1}_{\mathcal{C}_3}\right] \leq (n-1)^6\mathbb{P}_{\mathcal{C}}\left(\mathcal{M}_{x,y,z}^c\right) = o(1/n).$$

Inserting the combined expectations over $\mathcal{C}_1$ ,$\mathcal{C}_2$ and $\mathcal{C}_3$ into (5.15) yields

$$\mathbb{E}_{\mathcal{C}}\left[h(i,j)^3\right] \leq n^3\left(1 + O\left(\frac{\log n}{np\log(np)}\right)\right). \tag{5.16}$$

By the definition (2.9) of $T(\mathcal{G})^2$ and Hölder's inequality (2.2) with exponent 3, we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[T(\mathcal{G})^2\right] &= \mathbb{E}_{\mathcal{C}}\left[\left(\sum_{i,j\in V}\frac{\gamma_1(i)\gamma_1(j)}{(2m)^2}h(i,j)\right)^2\right]\\ &= \mathbb{E}_{\mathcal{C}}\left[\sum_{i,j,x,y\in V}\frac{\gamma_1(i)\gamma_1(j)\gamma_1(x)\gamma_1(y)}{(2m)^4}h(i,j)h(x,y)\right]\\ &\leq \sum_{i,j,x,y\in V}\left(\mathbb{E}_{\mathcal{C}}\left[\left(\frac{\gamma_1(i)\gamma_1(j)\gamma_1(x)\gamma_1(y)}{(2m)^4}\right)^3\right]\mathbb{E}_{\mathcal{C}}\left[h(i,j)^3\right]\mathbb{E}_{\mathcal{C}}\left[h(x,y)^3\right]\right)^{1/3}.\end{aligned}$$

Using Hölder's inequality again this time with exponent 5 and collecting like terms yields

$$\mathbb{E}_{\mathcal{C}}\left[T(\mathcal{G})^2\right] \le \frac{n^4}{2^4}\left(\mathbb{E}_{\mathcal{C}}\left[\gamma_1(i)^{15}\right]^4 \mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^{60}}\right]\right)^{1/15}\left(\mathbb{E}_{\mathcal{C}}\left[h(i,j)^3\right]^2\right)^{1/3}.$$

Applying the bounds (2.16), (5.14) and (5.16) respectively then Bernoulli's inequality (2.3) gives

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[T(\mathcal{G})^2\right] &\le \frac{n^4}{2^4}\left((np)^{60} + O\big((np)^{59}\big)\right)^{1/15}\left(\frac{2^{60}}{n^{120}p^{60}} + O\left(\frac{\log n}{n^{121}p^{61}\log(np)}\right)\right)^{1/15} \\ &\quad \cdot \left(n^6 + O\left(\frac{n^6 \log n}{np\log(np)}\right)\right)^{1/3} \\ &= n^2 + O\left(\frac{n^2\log n}{np\log(np)}\right).\end{aligned}$$

Then by Jensen's inequality and the lower bound on $\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})]$ from Theorem 5.1.1 (iv)

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[T(\mathcal{G})^2\right] &\ge \mathbb{E}_{\mathcal{C}}[T(\mathcal{G})]^2 \\ &\ge \left(n - O\left(\frac{n\log n}{np\log(np)}\right)\right)^2 \\ &= n^2\left(1 - O\left(\frac{\log n}{np\log(np)}\right)\right).\end{aligned}$$

Similar calculations yield the same bounds for $\mathbb{E}_{\mathcal{C}}\left[H(\mathcal{G})^2\right]$ and $\mathbb{E}_{\mathcal{C}}\left[H_i(\mathcal{G})^2\right]$. □

## 5.2 Concentration for hitting times, Theorem 5.2.1

The previous theorem, Theorem 5.1.1, provided second moments for the hitting times and some other indices. We can apply the second moment method to prove the following concentration result.

**Theorem 5.2.1.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$ *with* $\log n + \log\log\log n \le np \le n^{1/10}$, $f(n) : \mathbb{N} \to \mathbb{R}_+$. *Then for* $X \in \{h(i,j),\ \kappa(i,j),\ K(\mathcal{G}),\ H_i(\mathcal{G}),\ H(\mathcal{G}),\ T(\mathcal{G}),\ cc_i(\mathcal{G}),\ \overline{cc}(\mathcal{G})\}$, $i,j \in V$, $i \ne j$,

$$\mathbb{P}\left(\left|X - \mathbb{E}\left[X\middle|\mathcal{C}\right]\right| > \mathbb{E}\left[X\middle|\mathcal{C}\right]\sqrt{\frac{f(n)\log n}{np\log(np)}}\right) = O\left(\frac{1}{f(n)}\right) + \mathbb{P}(\mathcal{C}^c).$$

*Proof of Theorem 5.2.1.* Let $X \in \{h(i,j),\ \kappa(i,j),\ H_i(\mathcal{G}),\ H(\mathcal{G}),\ T(\mathcal{G}),\ cc_i\}$ where

$i, j \in V$ and recall $\mathbb{E}_{\mathcal{C}}[\cdot] = \mathbb{E}[\cdot|\mathcal{C}]$. We have the following for these $X$ by Theorem 5.1.1

$$\mathrm{Var}\left(X\middle|\mathcal{C}\right) = n^2 + O\left(\frac{n\log n}{p\log(np)}\right) - \left(n + O\left(\frac{\log n}{p\log(np)}\right)\right)^2 = O\left(\frac{n\log n}{p\log(np)}\right).$$

We can also calculate the conditional variance of $K(\mathcal{G})$ by Theorem 5.1.1, this yields

$$\begin{aligned}\mathrm{Var}\left(K(\mathcal{G})\middle|\mathcal{C}\right) &= \frac{n^2}{p^2} \pm O\left(\frac{n\log n}{p^3\log(np)}\right) - \left(\frac{n}{p} \pm O\left(\frac{\log n}{p^2\log(np)}\right)\right)^2 \\ &= O\left(\frac{n\log n}{p^3\log(np)}\right).\end{aligned}$$

By the Chebyshev inequality (2.6) for each of the above

$$\mathbb{P}\left(\left|X - \mathbb{E}\left[X\middle|\mathcal{C}\right]\right| \geq \lambda(n)\sqrt{\mathrm{Var}\left(X|\mathcal{C}\right)}\,\middle|\,\mathcal{C}\right) \leq \frac{1}{\lambda(n)^2}.$$

For $X$ above we have $\mathrm{Var}\left(X|\mathcal{C}\right) = O\left(\mathbb{E}\left[X\middle|\mathcal{C}\right]^2 \frac{\log n}{np\log(np)}\right)$ by Theorem 5.1.1, thus there exists some $K$ independent of $n$ and $X$ such that

$$\sqrt{\mathrm{Var}\left(X|\mathcal{C}\right)} < \mathbb{E}\left[X\middle|\mathcal{C}\right]\sqrt{\frac{K\log n}{np\log(np)}},$$

for large $n$. By choosing $\lambda(n) = \sqrt{f(n)/K}$ for any function $f(n)$ we have

$$\mathbb{P}\left(\left|X - \mathbb{E}\left[X\middle|\mathcal{C}\right]\right| > \mathbb{E}\left[X\middle|\mathcal{C}\right]\sqrt{\frac{f(n)\log n}{np\log(np)}}\,\middle|\,\mathcal{C}\right) \leq \frac{K}{f(n)} = O\left(\frac{1}{f(n)}\right). \tag{5.17}$$

The result follows since $\mathbb{P}(\mathcal{A}) \leq \mathbb{P}(\mathcal{A}|\mathcal{C}) + \mathbb{P}(\mathcal{C}^c)$, for any event $\mathcal{A}$.

For $\overline{cc}(\mathcal{G})$ we will obtain concentration by comparison with $K(\mathcal{G})$. For any function $f(n)$ let $\mathcal{E}$ be the event $\left\{|m - \binom{n}{2}p| \leq \sqrt{3\log(f(n))\binom{n}{2}p}\right\}$. Recall that

$$\overline{cc}(\mathcal{G}) = \frac{2mK(\mathcal{G})}{n(n-1)},$$

by (2.13), where $m := |E| \sim_d Bin(\binom{n}{2}, p)$. Then conditional on event $\mathcal{E}$ we have

$$\begin{aligned}|\overline{cc}(\mathcal{G})/p - K(\mathcal{G})| &= |K(\mathcal{G}) \cdot (2m/n(n-1)p) - K(\mathcal{G})| \\ &\leq K(G)\sqrt{6\log(f(n))/n^2p}.\end{aligned}$$

Let $\mathcal{T}$ be the event $\left\{\left|K(\mathcal{G}) - \mathbb{E}_{\mathcal{C}}[K(\mathcal{G})]\right| \leq \mathbb{E}_{\mathcal{C}}[K(\mathcal{G})]\sqrt{\frac{f(n)\log n}{2np\log(np)}}\right\}$, where

$\mathbb{P}(\mathcal{T}|\mathcal{C})$ is given by (5.17). Observe that $\mathbb{E}\big[\overline{cc}(\mathcal{G})\big|\mathcal{C}\big] \overset{O}{=} p\mathbb{E}\big[K(\mathcal{G})\big|\mathcal{C}\big]$ by Theorem 5.1.1, so conditional on the event $\mathcal{E} \cap \mathcal{T}$ we have

$$\begin{aligned}\left|\frac{\overline{cc}(\mathcal{G})}{p} - K(\mathcal{G})\right| &\leq \mathbb{E}_{\mathcal{C}}[K(\mathcal{G})]\left(1+\sqrt{\frac{f(n)\log n}{2np\log(np)}}\right)\sqrt{\frac{6\log(f(n))}{n^2p}} \\ &\leq \frac{3\mathbb{E}\big[\overline{cc}(\mathcal{G})\big|\mathcal{C}\big]}{p}\sqrt{\frac{\log(f(n))}{n^2p}}.\end{aligned}$$

Finally by the estimates above and the triangle inequality, conditional on the event $\mathcal{E} \cap \mathcal{T}$ we have the following

$$\begin{aligned}\left|\frac{\overline{cc}(\mathcal{G})}{p} - \frac{\mathbb{E}_{\mathcal{C}}[\overline{cc}(\mathcal{G})]}{p}\right| &\leq \left|\frac{\overline{cc}(\mathcal{G})}{p} - K(\mathcal{G})\right| + \Big|K(\mathcal{G}) - \mathbb{E}_{\mathcal{C}}[K(\mathcal{G})]\Big| \\ &\quad + \left|\mathbb{E}_{\mathcal{C}}[K(\mathcal{G})] - \frac{\mathbb{E}_{\mathcal{C}}[\overline{cc}(\mathcal{G})]}{p}\right| \\ &\leq \frac{3\mathbb{E}\big[\overline{cc}(\mathcal{G})\big|\mathcal{C}\big]}{p}\sqrt{\frac{\log(f(n))}{n^2p}} + \frac{\mathbb{E}\big[\overline{cc}(\mathcal{G})\big|\mathcal{C}\big]}{p}\sqrt{\frac{f(n)\log n}{2np\log(np)}} \\ &\quad + \frac{n}{p}\cdot O\left(\frac{\log n}{np\log(np)}\right) \\ &\leq \frac{\mathbb{E}_{\mathcal{C}}[\overline{cc}(\mathcal{G})]}{p}\sqrt{\frac{f(n)\log n}{np\log(np)}}. \end{aligned} \tag{5.18}$$

We then apply the Chernoff bounds, Lemma 2.1.1, to obtain

$$\begin{aligned}\mathbb{P}(\mathcal{E}^c) &\leq \exp\left(\frac{-3\log(f(n))\binom{n}{2}p}{2\binom{n}{2}p}\right) + \exp\left(\frac{-3\log(f(n))\binom{n}{2}p}{2\left(\binom{n}{2}p + \sqrt{3\log(f(n))\binom{n}{2}p/3}\right)}\right) \\ &\leq o\left(\frac{1}{f(n)}\right).\end{aligned}$$

Finally by (5.18), the above bound on $\mathbb{P}(\mathcal{E}^c)$ and the bound (5.17) on $\mathbb{P}(\mathcal{T}^c)$ we have

$$\begin{aligned}\mathbb{P}\left(\Big|\overline{cc}(\mathcal{G}) - \mathbb{E}_{\mathcal{C}}[\overline{cc}(\mathcal{G})]\Big| > \mathbb{E}_{\mathcal{C}}[\overline{cc}(\mathcal{G})]\sqrt{\frac{f(n)\log n}{np\log(np)}}\right) &\leq \mathbb{P}(\mathcal{E}^c) + \mathbb{P}(\mathcal{T}^c) \\ &\leq O\left(\frac{1}{f(n)}\right) + \mathbb{P}(\mathcal{C}^c).\end{aligned}$$

□

## 5.3 Resistance and degree, Theorem 5.3.1

We prove the following theorem using the resistance bounds from Chapter 3 and the results of Chapter 4 which relate the pruned neighbourhoods to the original neighbourhoods in $\mathcal{G}(n,p)$.

**Theorem 5.3.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ and $i, j \in V, i \neq j$.*

*(i) For every $c > 0$ if $c \log n \leq np \leq n^{1/10}$, then*

$$\mathbb{P}\left(\left|R(i,j) - \left(\frac{1}{\gamma_1(i)} + \frac{1}{\gamma_1(j)}\right)\right| > \max\left\{\frac{1}{\gamma_1(i)^2} + \frac{1}{\gamma_1(j)^2}, \frac{9(\gamma_1(i)+\gamma_1(j))\log n}{\gamma_1(i)\gamma_1(j)np\log(np)}\right\}\right) \leq 2np^2 + o\left(e^{-np/4}\right).$$

*(ii) For every $c > 0$ if $np = c \log n$, then for any fixed $k > 0$*

$$\mathbb{P}\left(\left|R(i,j) - \frac{2}{c\log n}\right| > \frac{10}{c^2 \log(n) \log\log(n)}\right) \leq \frac{5}{(\log n)^k}.$$

*(iii) If $np = \omega(\log n)$ and $np \leq n^{1/10}$, then*

$$\mathbb{P}\left(\left|R(i,j) - \frac{2}{np}\right| > \frac{7\sqrt{\log n}}{(np)^{3/2}}\right) = o\left(\frac{1}{n^{7/2}}\right).$$

*Proof of Theorem 5.3.1.* Define the following three functions for ease of notation

$$\begin{aligned} r_{i,j} &:= \frac{1}{\gamma_1(i)} + \frac{1}{\gamma_1(j)}, \\ f_{i,j} &:= \frac{1}{\gamma_1(i)^2} + \frac{1}{\gamma_1(j)^2}, \\ g_{i,j} &:= \frac{9(\gamma_1(i)+\gamma_1(j))\log n}{\gamma_1(i)\gamma_1(j)np\log(np)}. \end{aligned}$$

*Item* (i): we wish to show that $R(i,j)$ differs from $r_{i,j}$ by at most $\max\{f_{i,j}, g_{i,j}\}$. Let $\mathcal{H}$ be the event $\{|R(i,j) - r_{i,j}| \leq \max\{f_{i,j}, g_{i,j}\}\}$. By Lemma 3.1.3 we have

$$\begin{aligned} R(i,j) - r_{i,j} &\geq -\left(\frac{1}{\gamma_1(i)^2 + \gamma_1(i)} + \frac{1}{\gamma_1(j)^2 + \gamma_1(j)}\right) \\ &> -\left(\frac{1}{\gamma_1(i)^2} + \frac{1}{\gamma_1(j)^2}\right) \\ &= -f_{i,j}. \end{aligned}$$

Let $\mathcal{L}$ be the event $\{\psi_1(i) = \gamma_1(i), \psi_1(j) = \gamma_1(j)\}$, where $\mathbb{P}(\mathcal{L}) = 2np^2 + e^{-(1-o(1))np}$

by Lemma 4.3.1. We also define the following event

$$\mathcal{F} := \left\{ R(i,j) \le \left( \frac{1}{\psi_1(u)} + \frac{1}{\psi_1(v)} \right) \left( 1 + \frac{9 \log n}{np \log(np)} \right) \right\}.$$

Observe $\mathbb{P}(\mathcal{F}^c) = o\left(e^{-np/4}\right) + o\left(n^{-7/2}\right)$ by Lemma 5.0.1 (iii). Conditional on $\mathcal{L} \cap \mathcal{F}$

$$R(i,j) - r_{i,j} \le \left( \frac{1}{\gamma_1(u)} + \frac{1}{\gamma_1(v)} \right) \left( 1 + \frac{9 \log n}{np \log(np)} \right) - \left( \frac{1}{\gamma_1(u)} + \frac{1}{\gamma_1(v)} \right) \le g_{i,j}.$$

Thus combining the bounds on $R(i,j)$ conditional on $\mathcal{L} \cap \mathcal{F}$ above we have

$$\mathbb{P}(\mathcal{H}^c) \le \mathbb{P}(R(i,j) - r_{i,j} < -f_{i,j}) + \mathbb{P}(R(i,j) - r_{i,j} > g_{i,j}) \le \mathbb{P}(\mathcal{L}^c) + \mathbb{P}(\mathcal{F}^c).$$

Applying the bounds on $\mathbb{P}(\mathcal{L}^c)$ and $\mathbb{P}(\mathcal{F}^c)$ from Lemmas 4.3.1 (iii) and 5.0.1 (iii) respectively:

$$\mathbb{P}(\mathcal{H}^c) \le o\left(e^{-np/4}\right) + o\left(n^{-7/2}\right) + 2np^2 + e^{-(1-o(1))np} \le 2np^2 + o\left(e^{-np/4}\right).$$

*Item* (ii): we seek to bound the tails of $|R(i,j) - 2/np|$ when $np = O(\log n)$. Let $\mathcal{E}(\lambda(n))$ be the event $\left\{ |\gamma_1(i) - np|, |\gamma_1(j) - np| \le \sqrt{np \cdot \lambda(n)} \right\}$, for $\lambda(n) = o(np)$. By Lemma 2.1.1:

$$\begin{aligned}
\mathbb{P}(\mathcal{E}(\lambda(n))^c) &\le 2 \exp\left( \frac{-(\sqrt{np \cdot \lambda(n)} - p)^2}{2(n-1)p} \right) + 2 \exp\left( \frac{-(\sqrt{np \cdot \lambda(n)} - p)^2}{2\left(np + \sqrt{np \cdot \lambda(n)}/3\right)} \right) \\
&\le 4e^{-\lambda(n)/3}.
\end{aligned}$$

Choose $\lambda(n) = 3k \log \log n$, $k \in \mathbb{R}_+$, then conditional on the event $\mathcal{E}(\lambda(n)) \cap \mathcal{H}$ we have

$$\begin{aligned}
\left| R(i,j) - \frac{2}{np} \right| &\le \left| R(i,j) - r_{i,j} \right| + \left| r_{i,j} - \frac{2}{np} \right| \\
&\le \frac{19 \log(n)}{2(np)^2 \log(np)} + \frac{2\sqrt{k \log \log n}}{(np)^{3/2}} \\
&\le \frac{10 \log(n)}{(np)^2 \log(np)},
\end{aligned}$$

since $\max\{f_{i,j}, g_{i,j}\} \leq 19\log(n)/\left(2(np)^2\log(np)\right)$ on $\mathcal{E}(\lambda(n))\cap\mathcal{H}$. Thus by Item (i):

$$\begin{aligned}\mathbb{P}\left(\left|R(i,j)-\frac{2}{np}\right| > \frac{10\log n}{(np)2\log(np)}\right) &\leq \mathbb{P}((\mathcal{H}\cap\mathcal{E})^c)\\ &\leq \mathbb{P}(\mathcal{H}^c)+\frac{4}{(\log n)^k}\\ &\leq \frac{5}{(\log n)^k}.\end{aligned}$$

*Item* (iii): our aim is now to bound the tails of $|R(i,j)-2/np|$ when $np=\omega(\log n)$. Run MBFS$(\mathcal{G},\{i,j\})$ and let $\mathcal{T}$ be the event

$$\left\{\psi_1(i)\geq np-3\sqrt{np\log n}\right\}\cap\left\{\psi_1(j)\geq np-3\sqrt{np\log n}\right\}.$$

Recall $\gamma_1^*(i)\sim_d Bin\left(n-\gamma_1(j)-2,p\right)$ by Lemma 4.1.1. Then by Lemmas 2.1.2 and 4.3.1 we have

$$\begin{aligned}\mathbb{P}(\mathcal{T}^c) &\leq \mathbb{P}\left(\left\{\gamma_1^*(i),\gamma_1^*(j)\geq np-3\sqrt{np\log n}\right\}^c\right)\\ &\qquad+\mathbb{P}(\psi_1(i)\neq\gamma_1^*(i)\text{ or }\psi_1(j)\neq\gamma_1^*(j))\\ &\leq 2\mathbb{P}\left(Bin\left(n-2np-2,p\right)<np-3\sqrt{np\log n}\right)\\ &\qquad+2\mathbb{P}(\gamma_1(i)>2np)+e^{-(1-o(1))np}.\end{aligned}$$

Then applying the Chernoff bounds, Lemma 2.1.1, and recalling in this instance $np=\omega(\log n)$ gives the following

$$\begin{aligned}\mathbb{P}(\mathcal{T}^c) &\leq 2e^{-(3\sqrt{np\log n}-1)^2/2np}+2e^{-np/2(1+1/3)}+e^{-(1-o(1))np}\\ &=o(1/n^4).\end{aligned}\tag{5.19}$$

Now for large $n$, conditional on the event $\mathcal{T}\cap\mathcal{F}$ we have

$$R(i,j)\leq\left(\frac{2}{np}+\frac{2\cdot3\sqrt{\log n}}{(np)^{3/2}}\right)\left(1+\frac{9\log n}{np\log(np)}\right)<\frac{2}{np}+\frac{7\sqrt{\log n}}{(np)^{3/2}}.\tag{5.20}$$

Choose $\lambda(n)=4\log n$, then applying Lemma 3.1.3 conditional on the event $\mathcal{E}(\lambda(n))$ yields

$$R(i,j)\geq\frac{2}{np}-\frac{2\sqrt{12\log n}}{(np)^{3/2}}>\frac{2}{np}-\frac{7\sqrt{\log n}}{(np)^{3/2}},\tag{5.21}$$

for large $n$. By upper and lower bounds on $R(i,j)$, (5.20) and (5.21) respectively, we have

$$\mathbb{P}\left(\left|R(i,j) - \frac{2}{np}\right| > \frac{7\sqrt{\log n}}{(np)^{3/2}}\right) \leq \mathbb{P}(\mathcal{T}^c) + \mathbb{P}(\mathcal{F}^c) + \mathbb{P}(\mathcal{E}(4\log n)^c)\,.$$

Now by (5.19), Lemma 5.0.1 (iii), $\mathbb{P}(\mathcal{E}(4\log n)^c) \leq 4e^{-np/3}$ (Lemma 2.1.1) we have

$$\begin{aligned}\mathbb{P}\left(\left|R(i,j) - \frac{2}{np}\right| > \frac{7\sqrt{\log n}}{(np)^{3/2}}\right) &\leq o(1/n^4) + o\left(e^{-np/4}\right) + o\left(1/n^{7/2}\right) + o(1/n^4) \\ &= o\left(1/n^{7/2}\right).\end{aligned}$$

□

## 5.4 Paths in $\mathcal{G}(n,p)$, Theorem 5.4.2

Let $\kappa(G)$ be the maximum $k$ such that $G$ has more than $k$ vertices and remains connected whenever fewer than $k$ vertices are removed. Let $\delta_1(G)$ be the size of the smallest first neighbourhood of any vertex in $G$. Bollobás & Thomason showed the following theorem.

**Theorem 5.4.1** ([21, Theorem 1])**.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, $0 < p := p(n) < 1$. Then,*

$$\mathbb{P}(\kappa(\mathcal{G}) = \delta_1(\mathcal{G})) = 1 - o(1).$$

Note that this was proven for the case $\delta_1(\mathcal{G}) = k$, for fixed $k$, by Erdős-Rényi [39], Ivčenko [48] and Bollobás [15]. We will show a theorem which is in this spirit and says that there are many edge independent paths between the second neighbourhoods of two vertices in $\mathcal{G}(n,p)$ with high probability.

Let $paths_2(i,j,l)$ be the maximum number of paths of length at most $l$ between vertices $i$ and $j$ of $\mathcal{G}$ that are vertex disjoint on $V \setminus (B_1(i) \cup B_1(j))$. The strong $k$-path property can be used to prove a related "local first neighbourhood relaxation" of the Bollobás & Thomason theorem for two vertices. What we mean by a relaxation in this context is that we allow the paths between two vertices $i,j$ to be non-disjoint within the first neighbourhoods of $i,j$.

**Theorem 5.4.2.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where for any $c > 0$, $c\log n \leq np \leq n^{1/10}$. Let $l := \log n/\log(np) + 9$. Then for $i,j \in V$ where $i \neq j$,*

*(i)* $\mathbb{P}(paths_2(i,j,l) \neq \min\{\gamma_2(i), \gamma_2(j)\}) \leq 5n^3p^4 + o\left(e^{-7\min\{np,\log n\}/2}\right)$,

*(ii)* $\mathbb{P}\left(\left|paths_2(i,j,l) - (np)^2\right| > 3(np)^{3/2}\sqrt{\log np}\right) = o\left(1/np\right)$.

It is of note that unlike Bollobás & Thomason's result, Theorem 5.4.2 (i) is a statement about the paths between two given vertices rather than a global statement. If one wishes to prove a similar relaxed connectivity condition on the whole graph a more sophisticated statement is needed.

*Proof of Theorem 5.4.2. Item* (i): For $i, j \in V$ we define $\mathcal{E}_{i,j}$ to be the following event:

$$\mathcal{E}_{i,j} := \{\text{there is no path from } i \text{ to } j \text{ of length less than 4.}\}.$$

Then by over-counting the number of paths we have

$$\begin{aligned}\mathbb{P}\big(\mathcal{E}_{i,j}^c\big) &\le \sum_{l=1}^{3} \mathbb{P}(\text{there exists a path from } i \text{ to } j \text{ of length } l)\\ &\le p + (n-2)p^2 + \binom{n-2}{2}p^3\\ &\le n^2p^3. \end{aligned} \tag{5.22}$$

Conditional on $\mathcal{E}_{i,j}$ every path between $i$ and $j$ must pass through at least one vertex from each of $\gamma_2(i)$ and $\gamma_2(j)$, though these vertices may not be distinct. So there cannot be more than $\min\{\gamma_2(i), \gamma_2(j)\}$ paths between $i, j \in V$ which are vertex disjoint on $V^* := V \setminus (B_1(i) \cup B_1(j))$ since $\Gamma_2(i) \cup \Gamma_2(j) \subseteq V^*$. Thus conditional on $\mathcal{E}_{i,j}$ for any $l \ge 0$ we have

$$paths_2(i, j, l) \le \min\{\gamma_2(i), \gamma_2(j)\}. \tag{5.23}$$

For a lower bound on $paths_2(i, j, l)$ we construct $\min\{\psi_2(i), \psi_2(j)\}$ vertex disjoint paths between $i$ and $j$ using the strong $k$-path property, Definition 3.3.1. A coupling is then used to relate the pruned second neighbourhoods to the standard second neighbourhoods.

For the path construction condition on the event $\mathcal{A}_{i,j}^n$ and assume without loss of generality $\psi_2(i) \le \psi_2(j)$. Take any subset $\Psi_2(j)^* \subseteq \Psi_2(j)$ with $\psi_2(i)$ elements and any perfect matching $M$ between $\Psi_2(i)$ and $\Psi_2(j)^*$. Given any pair $(x, y)$ in the matching $M$, conditional on $\mathcal{A}_{i,j}^n$, there is some $k$ and some pair $(x_k, y_k) \in \Gamma_k^*(x) \times \Gamma_k^*(y)$ such that $x_k y_k \in E$. We define the path $P_{x,y} := i, i_x, x, x_1, \ldots, x_k, y_k, y_{k-1}, \ldots, y, j_y, j$, where $x, x_1, \ldots, x_k$ is the unique path from $x$ to $x_k$ in the tree $T_k(x) := \cup_{i=0}^k \Gamma_i^*(x)$ and $i_x$ is the unique vertex in $\Gamma_1^*(i)$ connected to $x$. The equivalent descriptions hold for $y, y_1, \ldots, y_k \in T_k(y)$ and $j_y \in \Gamma_1(j)$ with respect to $y$ and $j$. The paths $\{P_{x,y}\}_{(x,y)\in M}$ are all vertex disjoint on $V^*$ since the trees $\{T_K(u)\}_{u\in\Psi_2}$ are all vertex disjoint. Each path in $P_{i,j}$ has length $l := 2k + 5$

where the $k$ is given by the event $\mathcal{A}_{i,j}^n$. Thus conditional on the event $\mathcal{A}_{i,j}^n$ we have

$$paths_2(i,j,l) \geq \left|\{P_{x,y}\}_{(x,y)\in M}\right| = \min\{\psi_2(i),\psi_2(j)\}. \tag{5.24}$$

Exchanging the $\psi_2$ and $\gamma_2$ distributions on the event $\{\psi_2(i) \neq \gamma_2(i)$ or $\psi_2(j) \neq \gamma_2(j)\}$ yields

$$\begin{aligned}
\mathfrak{P} &:= \mathbb{P}(paths_2(i,j,l) \neq \min\{\gamma_2(i),\gamma_2(j)\}) \\
&\leq \mathbb{P}(\psi_2(i) \neq \gamma_2(i) \text{ or } \psi_2(j) \neq \gamma_2(j)) + \mathbb{P}(paths_2(i,j,l) < \min\{\psi_2(i),\psi_2(j)\}) \\
&\quad + \mathbb{P}(\{paths_2(i,j,l) > \min\{\gamma_2(i),\gamma_2(j)\}\}).
\end{aligned}$$

Now by (5.24) and (5.23) we have the following

$$\mathfrak{P} \leq \mathbb{P}(\psi_2(i) \neq \gamma_2(i) \text{ or } \psi_2(j) \neq \gamma_2(j)) + \mathbb{P}\big((\mathcal{A}_{i,j}^n)^c\big) + \mathbb{P}\big(\mathcal{E}_{i,j}^c\big).$$

By the bounds on these probabilities from Lemma 4.3.1 (iv), Lemma 4.2.1 and (5.22) respectively:

$$\begin{aligned}
\mathfrak{P} &\leq 4n^3p^4 + O\big(n^2p^3\big) + o\left(e^{-7\min\{np,\log n\}/2}\right) + n^2p^3 \\
&\leq 5n^3p^4 + o\left(e^{-7\min\{np,\log n\}/2}\right).
\end{aligned}$$

On the event $\mathcal{A}_{i,j}^n$ the strong $k$-path property is satisfied for some $k \leq \lfloor\frac{\log n}{2\log(np)}\rfloor + 2$. We conditioned on the event $\mathcal{A}_{i,j}^n$, thus $l = 2k+5 \leq \frac{\log n}{\log(np)} + 9$.
*Item* (ii): Let $\mathcal{D}_{i,j}$ be the event

$$\left\{|\gamma_1(i) - np| \leq \sqrt{3np\log(np)}\right\} \cap \left\{|\gamma_1(j) - np| \leq \sqrt{3np\log(np)}\right\}.$$

Observe that we have the following by the Chernoff bounds, Lemma 2.1.1,

$$\begin{aligned}
\mathbb{P}\big(\mathcal{D}_{i,j}^c\big) &\leq \exp\left(-\frac{3np\log np}{2(n-1)p}\right) + \exp\left(-\frac{3np\log np}{2\left((n-1)p + \sqrt{3np\log np}/3\right)}\right) \\
&= o\left(\frac{1}{np}\right).
\end{aligned}$$

Now by Lemma 4.1.1 (iv) $\gamma_2(u) \sim_d Bin\left(n-1-\gamma_1(u), 1-(1-p)^{\gamma_1(u)}\right)$, conditional on $\gamma_1(u)$ for any $u \in V$. Observe that $(1-p)^k \leq 1-kp+(kp)^2$ when $(kp)^i \geq (kp)^{i+1}$ for all $i$ and recall the Bernoulli inequality (2.3). Thus conditional on $\mathcal{D}_{i,j}$ we have the following

$$\gamma_2(i) \succeq_1 Bin\left(n-2np, np^2 - 2p\sqrt{\log(np)np}\right),$$
$$\gamma_2(j) \preceq_1 Bin\left(n, np^2 + p\sqrt{3\log(np)np}\right).$$

Let $\mathcal{R}_{i,j}$ be the event $\left\{\left|\min\{\gamma_2(i),\gamma_2(j)\} - (np)^2\right| \le 3(np)^{3/2}\sqrt{\log np}\right\}$, thus we have

$$\begin{aligned}\mathbb{P}\big(\mathcal{R}_{i,j}^c\big) \le\, & 2\mathbb{P}\Big(\gamma_2(i) > (np)^2 + 3(np)^{3/2}\sqrt{\log np}\Big) \\ & + 2\mathbb{P}\Big(\gamma_2(i) < (np)^2 - 3(np)^{3/2}\sqrt{\log np}\Big).\end{aligned}$$

Then by applying the above stochastic domination for $\gamma_2(i)$ we obtain

$$\begin{aligned}\mathbb{P}\big(\mathcal{R}_{i,j}^c\big) \le\, & 2\mathbb{P}\Big(Bin\left(n, np^2 + p\sqrt{3\log(np)np}\right) > (np)^2 + 3(np)^{3/2}\sqrt{\log np}\Big) \\ & + 2\mathbb{P}\Big(Bin\left(n-2np, np^2 - 2p\sqrt{\log(np)np}\right) < (np)^2 - 3(np)^{3/2}\sqrt{\log np}\Big) \\ & + 4\mathbb{P}\big(\mathcal{D}_{i,j}^c\big).\end{aligned}$$

Finally the Chernoff bounds, Lemma 2.1.1, and the above bound on $\mathbb{P}\Big(\mathcal{D}_{i,j}^c\Big)$ yield

$$\begin{aligned}\mathbb{P}\big(\mathcal{R}_{i,j}^c\big) \le\, & 2\exp\left(\frac{-(3-\sqrt{3})^2(np)^3\log(np)}{2\left((np)^2 + (np)^{3/2}\sqrt{3\log(np)}\right)}\right) \\ & + 2\exp\left(\frac{-(3-5/2)^2(np)^3\log(np)}{2\left((np)^2 + 3(np)^{3/2}\sqrt{\log np}/3\right)}\right) + 4\cdot o\left(\frac{1}{np}\right) \\ =\, & o\left(\frac{1}{np}\right).\end{aligned}$$

The result now follows from Item (i) and the bound on $\mathbb{P}\Big(\mathcal{R}_{i,j}^c\Big)$ directly above since

$$\begin{aligned}& \mathbb{P}\Big(\left|paths_2(i,j,l) - (np)^2\right| > 3(np)^{3/2}\sqrt{\log np}\Big) \\ & \quad \le \mathbb{P}(paths_2(i,j,l) \neq \min\{\gamma_2(i),\gamma_2(j)\}) + \mathbb{P}\big(\mathcal{R}_{i,j}^c\big) \\ & \quad \le 5n^3p^4 + o\left(e^{-7\min\{np,\log n\}/2}\right) + o\left(1/np\right) \\ & \quad = o\left(1/np\right).\end{aligned}$$

□

# Chapter 6

# The distribution of the sizes of $r$-neighbourhoods in $\mathcal{G}(n,p)$

The aim of this chapter is to have a better understanding of the probability that the $r$-neighbourhood of a vertex in $\mathcal{G}(n,p)$ has size $k$, namely

$$\mathbb{P}(\gamma_r(u) = k)\,.$$

If we consider the first neighbourhood of a vertex $v \in V$, the case $r = 1$, then this is simple to understand since $\gamma_1(v) \sim_d Bin(n-1, p)$. More generally, Lemma 4.1.1 (iv) states that for any $r \geq 1$ the distribution of $\gamma_r(u)$, conditional on $\gamma_i(u)$ for $i = 1, \ldots, r-1$, is given by

$$\gamma_r(v) \sim_d Bin\left(n - \sum_{i=0}^{r-1} \gamma_i(v), 1 - (1-p)^{\gamma_{r-1}(v)}\right). \tag{6.1}$$

For $r \geq 2$ this is a far more complex distribution to analyse as it depends on the sizes of the previous neighbourhoods $\gamma_i(v)$ for $0 \leq i \leq r-1$. However, we would expect it to resemble a binomial distribution with some parameters. One naïve guess might be $Bin\left(n, n^{r-1}p^r\right)$, for an appropriate range of $p$ and $r$. This is motivated by (6.1) where we have simply plugged $\mathbb{E}[\gamma_{r-1}(u)] \approx (np)^{r-1}$ in for the value of $\gamma_{r-1}(u)$ and considered the contribution from $\sum_{i=0}^{r-1} \gamma_i(v)$ to be negligible.

Recall that $\log(\cdot)$ denotes the natural logarithm base $e$. The following theorem, which we prove in Section 6.4, shows that the density function of $\gamma_r(u)$ resembles that of the Gaussian and Binomial distributions.

**Theorem 6.0.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, where $np \to \infty$, and $u \in V$. Let $r := r(n)$, $r \geq 1$*

*and $k := k(n)$ be such that $(np)^{2r} = o(n)$ and $k = \Theta((np)^r)$. Let $\alpha = k/(np)^r$. Then there exists $C := C(\alpha) < \infty$ such that*

$$\mathbb{P}(\gamma_r(u) = k) \leq C \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{\sqrt{(np)^{2r-1}}}.$$

*If in addition $\alpha > 1/2\pi$ then there exists $c := c(\alpha) > 0$ such that*

$$\mathbb{P}(\gamma_r(u) = k) \geq c \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{\sqrt{(np)^{2r-1}}}.$$

To see why Theorem 6.0.1 suggests the distribution of $\gamma_r(u)$ is binomial in nature outside of the extreme values we will make a comparison with $\mathbb{P}(\gamma_1(u) = k)$ when $np = o\left(n^{1/2}\right)$ and $k = \Theta(np)$. Let $\alpha := k/np$. By Stirling's approximation we can derive the following asymptotic approximation for $\mathbb{P}(\gamma_1(u) = k)$ in this regime:

$$\begin{aligned}\mathbb{P}(\gamma_1(u) = k) &= \binom{n-1}{k} p^k (1-p)^{n-k-1} \\ &\sim \frac{1}{\sqrt{2\pi k}} \left(\frac{npe}{k}\right)^k e^{-np} \\ &\sim \frac{e^{(\alpha + \alpha\log(\alpha) - 1)np}}{\sqrt{2\pi\alpha(np)}}.\end{aligned}$$

Up to constants this has exactly the same form as the function in Theorem 6.0.1 for $r = 1$. Theorem 6.0.1 is almost a local limit theorem and it suggests that $\gamma_r(u)$ satisfies a central limit theorem with mean $(np)^r$ and variance $(np)^{2r-1}$. Indeed we show this in our next Theorem.

**Theorem 6.0.2** (Central Limit Theorem for $\gamma_r(u)$)**.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, where $np \to \infty$. Let $r := r(n)$ be such that $(np)^{r+1/2} = o(n)$ and let $u \in V$. Then*

$$\left(\frac{\gamma_r(u) - (np)^r}{(np)^{(2r-1)/2}}\right) \xrightarrow{d} \mathcal{N}(0,1).$$

These two theorems are consistent with our understanding that the distribution of $\gamma_r(u)$ is binomial in nature however, the naïve guess of $\gamma_r(u)$ having a distribution close to that of $Bin(n, (np)^{r-1}p)$ is incorrect. This is because a random variable distributed according to $Bin(n, (np)^{r-1}p)$ has variance of order $(np)^r$ whereas Theorem 6.0.2 and Theorem 6.0.1 show the variance of $\gamma_r(u)$ is actually significantly greater since it has order $(np)^{2r-1}$.

We prove Theorem 6.0.1 using a classical method attributed to Pierre-Simon Laplace. We will introduce this in a heuristic manner in Section 6.2 before applying the same idea rigorously in the remainder of the Chapter to prove Lemma 6.2.3. Firstly however we will first prove Theorem 6.0.2 using another tool associated with Laplace, the Laplace transform.

## 6.1 Proof of the CLT for neighbourhood size

We wish to show the distribution of the size of the $r$-neighbourhoods, suitably scaled, converges in distribution to a Normal distribution. If $\mathbb{E}[\cdot]$ is expectation with respect to some probability measure $\mu$ then for a random variable $X$ with law $\mu$ we define the Laplace transform [41, Ch. XIII] to be

$$\mathcal{L}(\mu)(s) = \mathbb{E}\left[e^{-sX}\right].$$

The following classical theorem shows to prove convergence in distribution of $\gamma_r(u)$, suitably scaled, to a normal random variable it suffices to show convergence of the Laplace transform associated to the sequence of random variables to the Laplace transform of a Gaussian random variable.

**Theorem 6.1.1** (Extended continuity theorem [41, Theorem 2(a), Ch. XIII]). *For $n = 1, 2, \ldots$ let $\mu_n$ be a measure with Laplace transform $\omega_n$. If $\omega_n(\lambda) \to \omega(\lambda)$ for $\lambda > a$, then $\omega$ is the Laplace transform of a measure $\mu$ and $\mu_n \to \mu$. Conversely, if $\mu_n \to \mu$ and the sequence $\{\omega_n(a)\}$ is bounded, then $\omega_n(\lambda) \to \omega(\lambda)$ for $\lambda > a$.*

This theorem also appears in Kallenberg's book [52, Theorem 4.22] where it is attributed to Lévy and Bochner. We can now prove Theorem 6.0.2.

*Proof of Theorem 6.0.2.* To begin we define the filtration $\widetilde{\mathfrak{F}}_r$ in the following way

$$\widetilde{\mathfrak{F}}_r := \widetilde{\mathfrak{F}}(\mathcal{G}, u, r) = \sigma\left(\Gamma_i(u) : 0 \le i \le r\right).$$

This is the filtration generated by the vertices at distance at most $r$ from $u$. Observe that by the tower property we have

$$\mathbb{E}\left[e^{-\lambda\gamma_{r+1}(u)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{-\lambda\gamma_{r+1}(u)}\middle|\widetilde{\mathfrak{F}}_r\right]\right].$$

Recall Lemma 4.1.1 (iv) states that conditional on $|B_r(u)|, \gamma_r(u) \in \widetilde{\mathfrak{F}}_r$ we have

$$\gamma_{r+1}(u) \sim_d Bin\left(n - |B_r(u)|, 1 - (1-p)^{\gamma_r(u)}\right).$$

Observe that $\mathbb{E}\left[e^{-\lambda Bin(n,p)}\right] = (1 - p + e^{-\lambda}p)^n$ since the binomial random variable is a sum of independent Bernoulli random variables. Thus we have the following

$$\mathbb{E}\Big[e^{-\lambda\gamma_{r+1}(u)}\Big] = \mathbb{E}\left[\left((1-p)^{\gamma_r(u)} + e^{-\lambda}\left(1-(1-p)^{\gamma_r(u)}\right)\right)^{n-|B_r(u)|}\right].$$

Let $f(\gamma_r(u))$ denote the bracketed term in the expectation above. Observe

$$f(\gamma_r(u)) := (1-p)^{\gamma_r(u)} + e^{-\lambda}\left(1-(1-p)^{\gamma_r(u)}\right) = e^{-\lambda} + \left(1-e^{-\lambda}\right)(1-p)^{\gamma_r(u)}.$$

We will consider $\lambda = o(1)$ and thus $e^{-\lambda} = 1-\lambda+O\big(\lambda^2\big)$. By the Bernoulli inequality (2.3) and (2.4) whenever $\gamma_r(u)p = o(1)$ we have $(1-p)^{\gamma_r(u)} = 1-\gamma_r(u)p + O\big(\gamma_r(u)^2p^2\big)$ . Thus if $\gamma_r(u)p = o(1)$ then the following holds

$$\begin{aligned} f(\gamma_r(u)) &= 1-\lambda+\lambda^2/2-\cdots+\left(\lambda-\lambda^2/2+\cdots\right)\left(1-\gamma_r(u)p+O\big(\gamma_r(u)^2p^2\big)\right) \\ &= 1-\lambda\gamma_1(u)p+O\big(\lambda\gamma_r(u)^2p^2\big)-O\big(\lambda^2\gamma_r(u)p\big). \end{aligned} \tag{6.2}$$

Otherwise by the Taylor series for $\log(1-p)$ we have

$$\begin{aligned} f(\gamma_r(u)) &= 1-\lambda+\lambda^2/2-\cdots+\left(\lambda-\lambda^2/2+\cdots\right)\exp\left(\log(1-p)\gamma_r(u)\right) \\ &= 1-\left(\lambda+\lambda^2/2-\cdots\right)\left(1-e^{-\gamma_r(u)p-O\left(\gamma_r(u)p^2\right)}\right). \end{aligned} \tag{6.3}$$

Thus $f(\gamma_r(u)) \le 1$ by (6.2) and (6.3) since $\sup_{x\in\mathbb{R}_+} e^{-x} = 1$. Recall that for $\beta > 3$:

$$\mathbb{P}\big(|B_r(u)| > (2\beta^2+1)(np)^r\big) = o\left(\exp\left(-\frac{3(\beta-3)np}{2}\right)\right),$$

by Lemma 4.1.2. Let $\mathcal{E}_r := \big\{|B_r(u)| \le (2\beta^2+1)(np)^r\big\}$. Hence for $\beta > 3$ we have

$$\mathbb{E}\Big[e^{-\lambda\gamma_{r+1}(u)}\mathbf{1}_{(\mathcal{E}_r)^c}\Big] \le 1\cdot\mathbb{P}((\mathcal{E}_r)^c) = o\Big(e^{-3(\beta-3)np/2}\Big). \tag{6.4}$$

Now for the remaining part of the partition we have $\gamma_r(u)p\mathbf{1}_{\mathcal{E}_r} \le (\beta^2+1)(np)^{r+1}/n$. We will now reduce the Laplace transform of the centred and $\lambda$-scaled $\gamma_{r+1}(u)$ random variable to that of the centred and $\lambda$-scaled $\gamma_r(u)$ random variable. As we are dealing with $r+1$ for convenience the assumption of the theorem becomes

$(np)^{r+3/2} = o(n)$. We have the following by (6.2)

$$\begin{aligned}
\mathbb{E}\Big[e^{-\lambda\gamma_{r+1}(u)}\mathbf{1}_{\mathcal{E}_r}\Big] &= \mathbb{E}\Big[(1-\lambda\gamma_r(u)p + O(\lambda\gamma_r(u)p\,(\lambda+\gamma_r(u)p)))^{n-|B_r(u)|}\,\mathbf{1}_{\mathcal{E}_r}\Big] \\
&\leq \frac{\mathbb{E}\big[\big(1-\lambda\gamma_r(u)p + O\big(\lambda n^r p^{r+1}\left(\lambda + n^r p^{r+1}\right)\big)\big)^n\,\mathbf{1}_{\mathcal{E}_r}\big]}{(1-O(\lambda n^r p^{r+1}))^{(\beta^2+1)(np)^r}} \\
&= \frac{\mathbb{E}\big[\exp\left(-\lambda\gamma_r(u)np + O\big(\lambda(np)^{r+1}\left(\lambda+n^r p^{r+1}\right)\big)\right)\mathbf{1}_{\mathcal{E}_r}\big]}{(1-O(\lambda n^r p^{r+1}))^{(\beta^2+1)(np)^r}} \\
&= \frac{\mathbb{E}[\exp\left(-\lambda\gamma_r(u)np\right)\mathbf{1}_{\mathcal{E}_r}]\exp\left(O\big(\lambda(np)^{r+1}\left(\lambda+n^r p^{r+1}\right)\big)\right)}{(1-O(\lambda n^r p^{r+1}))^{(\beta^2+1)(np)^r}}. \qquad (6.5)
\end{aligned}$$

Let $\lambda = t/(np)^{(2r+1)/2}$, where $t \in \mathbb{R}_+$. Then for this choice of $\lambda$ we have

$$\lambda(np)^{r+1}\left(\lambda + n^r p^{r+1}\right) = \frac{t^2}{(np)^r} + \frac{t(np)^{r+3/2}}{n},$$

and also applying Bernoulli's inequality to the denominator of (6.5) yields

$$\big(1-O\big(\lambda n^r p^{r+1}\big)\big)^{(\beta^2+1)(np)^r} \geq 1 - O\big(\lambda n^{2r}p^{2r+1}\big) = 1 - O\left(\frac{(np)^{r+1/2}}{n}\right).$$

Thus combining (6.4) with (6.5) for this choice of $\lambda$ we obtain

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\frac{t\gamma_{r+1}(u)}{(np)^{(2r+1)/2}}\right)\right] \\
&\quad\leq \mathbb{E}\left[\exp\left(-\frac{t\gamma_r(u)}{(np)^{(2r-1)/2}}\right)\right]\left(1 + O\left(\frac{1}{(np)^r}\right) + O\left(\frac{(np)^{r+3/2}}{n}\right)\right) \qquad (6.6)\\
&\qquad + o\left(e^{-3(\beta-3)np/2}\right).
\end{aligned}$$

Similarly by (6.2) we have

$$\begin{aligned}
\mathbb{E}\Big[e^{-\lambda\gamma_{r+1}(u)}\Big] &\geq \mathbb{E}[(1-\lambda\gamma_r(u)p + O(\lambda\gamma_r(u)p\,(\lambda+\gamma_r(u)p)))^n\,\mathbf{1}_{\mathcal{E}_r}] \\
&= \mathbb{E}\big[\exp\left(-\lambda\gamma_r(u)np + O\big(\lambda n^2p^2\left(\lambda+np^2\right)\big)\right)\mathbf{1}_{\mathcal{E}_r}\big] \\
&\geq \mathbb{E}\big[\exp\left(-\lambda\gamma_r(u)np\right)\left(1-\mathbf{1}_{(\mathcal{E}_r)^c}\right)\big] \\
&\geq \mathbb{E}[\exp\left(-\lambda\gamma_r(u)np\right)] - \mathbb{P}((\mathcal{E}_r)^c), \qquad (6.7)
\end{aligned}$$

since $\gamma_r(u) \geq 0$. Thus if we let $\lambda = t/(np)^{(2r+1)/2}$ then (6.4) and (6.7) yield

$$\mathbb{E}\left[\exp\left(-\frac{t\gamma_{r+1}(u)}{(np)^{(2r+1)/2}}\right)\right] \geq \mathbb{E}\left[\exp\left(-\frac{t\gamma_r(u)}{(np)^{(2r-1)/2}}\right)\right] - o\left(e^{-3(\beta-3)np/2}\right). \qquad (6.8)$$

Observe that $\mathbb{E}\Big[\exp\Big(\frac{t(np)^{r+1}}{(np)^{(2r+1)/2}}\Big)\Big] = e^{t\sqrt{np}}$. Thus by (6.6) and (6.8) we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\frac{t(\gamma_{r+1}(u)-(np)^{r+1})}{(np)^{(2r+1)/2}}\right)\right] \\
&\quad = e^{t\sqrt{np}}\mathbb{E}\left[\exp\left(-\frac{t\gamma_r(u)}{(np)^{(2r-1)/2}}\right)\right]\left(1+O\left(\frac{1}{(np)^r}\right)+O\left(\frac{(np)^{r+3/2}}{n}\right)\right) \\
&\qquad \pm e^{t\sqrt{np}}o\left(e^{-3(\beta-3)np/2}\right) \\
&\quad = \mathbb{E}\left[\exp\left(-\frac{t\left(\gamma_r(u)-(np)^r\right)}{(np)^{(2r-1)/2}}\right)\right]\left(1+O\left(\frac{1}{(np)^r}\right)+O\left(\frac{(np)^{r+3/2}}{n}\right)\right) \\
&\qquad \pm e^{t\sqrt{np}}o\left(e^{-3(\beta-3)np/2}\right). \qquad (6.9)
\end{aligned}$$

For simplicity we shall denote $a_r := \mathbb{E}\Big[\exp\Big(-\frac{t(\gamma_r(u)-(np)^r)}{(np)^{(2r-1)/2}}\Big)\Big]$, $b_r := \frac{1}{(np)^r} + \frac{(np)^{r+3/2}}{n}$ and $c := e^{t\sqrt{np}}o\big(e^{-3(\beta-3)np/2}\big)$. Then ignoring the fact there may be constant factors in front of the $b_i$ terms we see that (6.9) above is a recurrence of the following form

$$a_{r+1} = a_r(1+b_r) \pm c.$$

Iterating this we obtain the following expression for $a_{r+1}$ in terms of $a_1, b_i$ and $c$

$$\begin{aligned}
a_{r+1} &= a_1 \cdot \prod_{i=1}^{r}(1+b_i) \pm \sum_{i=1}^{r} c \cdot \prod_{j=i+1}^{r}(1+b_j) \\
&= a_1 \cdot (1+b_1+\cdots+b_r+b_1b_2+\cdots+b_ib_j+\cdots+b_1b_2b_3+\cdots) \\
&\qquad \pm c \cdot \left[\sum_{i=1}^{r}(1+b_{i+1}+\cdots b_r+b_{i+1}b_{i+2}+\cdots)\right].
\end{aligned}$$

If we consider the relative sizes of $b_i$ for various $1 \le i \le r$ we see that the largest terms are $b_1$ and $b_r$ and that $b_i = O(\max\{b_1, b_r\}/(np))$ for all $2 \le i \le r-1$. Thus

$$a_{r+1} = a_1 \cdot (1+(b_1+b_r)(1+o(1))) \pm c \cdot (1+(b_2+rb_r)(1+o(1))).$$

If we now apply the above to the original relation (6.9) we obtain

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\frac{t(\gamma_{r+1}(u)-(np)^{r+1})}{(np)^{(2r+1)/2}}\right)\right]\\
&\quad=\mathbb{E}\left[\exp\left(-\frac{t(\gamma_1(u)-np)}{\sqrt{np}}\right)\right]\left(1+O\left(\frac{1}{np}\right)+O\left(\frac{(np)^{r+3/2}}{n}\right)\right)\\
&\qquad\pm e^{t\sqrt{np}}o\left(e^{-3(\beta-3)np/2}\right)\left(1+O\left(\frac{1}{(np)^2}\right)+O\left(\frac{r\cdot(np)^{r+3/2}}{n}\right)\right).
\end{aligned}\tag{6.10}$$

We now have an expression on the right hand side which we can evaluate easily since $\gamma_1(u)\sim_d Bin(n-1,p)$. We include this standard calculation for completeness:

$$a_1=\mathbb{E}\left[\exp\left(-\frac{t(\gamma_1(u)-np)}{\sqrt{np}}\right)\right]=e^{t\sqrt{np}}\cdot\left(1-p+e^{-t/\sqrt{np}}p\right)^{n-1}.$$

Recall that $e^{-\lambda}=1-\lambda+\lambda^2/2-\lambda^3/3!+\cdots$ and observe

$$\begin{aligned}
a_1&=e^{t\sqrt{np}}\left(1-\left(\frac{t}{\sqrt{np}}-\frac{t^2}{2np}+\frac{t^3}{6(np)^{3/2}}-O\left(\frac{t^4}{(np)^2}\right)\right)p\right)^{n-1}\\
&=e^{t\sqrt{np}}\exp\left(-(n-1)\left(\frac{pt}{\sqrt{np}}-\frac{pt^2}{2np}+O\left(\frac{pt^3}{6(np)^{3/2}}\right)-\frac{t^2p^2}{np}-\cdots\right)\right)\\
&=\exp\left(\frac{t^2}{2}+O\left(\frac{1}{\sqrt{np}}\right)\right).
\end{aligned}$$

Thus by (6.10), since $\beta>3$ was arbitrary, we have

$$\mathbb{E}\left[\exp\left(-\frac{t(\gamma_{r+1}(u)-(np)^{r+1})}{(np)^{(2r+1)/2}}\right)\right]\to\exp\left(\frac{t^2}{2}\right).$$

The result follows by Theorem 6.1.1 since $e^{t^2/2}$ is the Laplace transform of a Normal $\mathcal{N}(0,1)$ random variable. □

## 6.2 Laplaces's method

Laplace's method is a means of evaluating integrals by approximating them using a Taylor series in an interval around a global (or local) maxima. The method originated from a 1774 paper of Laplace [56] and appears in many books, for example [31, Chapter 5] and [34, Section 2.4]. This introduction to the method follows the style of Stefan Wagner's notes [74] from the Athens summer school where I first learnt of the method [1].

The typical situation where Laplace's method is applied is when we have an integral of the form

$$f(t) = \int_a^b e^{g(x,t)}\, \mathrm{d}x$$

and we would like to know its asymptotic behaviour as $t \to \infty$. To apply Laplace's method we take the following steps:

- Identify the maximum of $g(x,t)$ for fixed $t$.
- Approximate $g(x,t)$ in an interval of $x(t)$ around the maximum.
- Estimate the contribution of the tails (remaining intervals).

Here is a rough explanation of how the method works. Suppose the function $g(x,t)$ is at least three times differentiable. The maximum occurs at a point $x^* = x^*(t)$ where

$$g_x(x^*,t) := \frac{dg}{\mathrm{d}x} = 0 \qquad \text{and} \qquad g_{xx}(x^*,t) := \frac{d^2 g}{\mathrm{d}x^2} < 0.$$

Around $x^*$ we have the Taylor expansion

$$g(x,t) = g(x^*,t) + \frac{g_{xx}(x^*,t)}{2!}(x-x^*)^2 + \frac{g_{xxx}(x^*,t)}{3!}(x-x^*)^3 + \cdots$$

If we now consider the integral

$$f(t) = \int_a^b e^{g(x,t)}\, \mathrm{d}x$$

then by the Taylor expansion around $x^*$ we can be approximate the integral by

$$\int_{-\infty}^{\infty} e^{g(x^*,t)+\frac{g_{xx}(x^*,t)}{2}(x-x^*)^2}\, \mathrm{d}x = e^{g(x^*,t)} \int_{-\infty}^{\infty} e^{\frac{g_{xx}(x^*,t)}{2}(x-x^*)^2}\, \mathrm{d}x.$$

By inspection of the integral on the right hand side we see that since $g_{xx}(x^*,t)$ is a negative constant the integral has the form of a Gaussian integral with variance $-g_{xx}(x^*,t)$. This can be evaluated explicitly giving

$$e^{g(x^*,t)} \cdot \sqrt{\frac{2\pi}{-g_{xx}(x^*,t)}}.$$

For this approach to work and be rigorous we need several things to be true. Firstly, the error term in the Taylor approximation must be small in a suitable region $[x^* - A, x^* + A]$ around the maximum. This usually means that $A := A(t)$ needs to be chosen in such a way that

$$g_{xxx}(x^*, t)A(t)^3 \to 0 \qquad \text{as} \qquad t \to \infty.$$

This is so that if we truncate the Taylor series the error from the higher order terms goes to zero.

Secondly, the contribution of the parts outside the central region $[x^* - A, x^* + A]$ needs to be negligible. For this, it is generally required that

$$g_{xx}(x^*, t)A(t)^2 \to \infty \qquad \text{as} \qquad t \to \infty.$$

The reason for this is that the error term

$$\mathcal{E} := \int_{-\infty}^{x^*-A} e^{\frac{g_{xx}(x^*,t)}{2}(x-x^*)^2}\, \mathrm{d}x + \int_{x^*+A}^{\infty} e^{\frac{g_{xx}(x^*,t)}{2}(x-x^*)^2}\, \mathrm{d}x,$$

coming from completing the Gaussian integral resembles the tails of a Gaussian with mean $x^*$ and variance $1/|g_{xx}(x^*, t)|$. Thus by the Mill's ratio Gaussian tail bound (2.5) we have

$$\mathcal{E} = \mathbb{P}\Big(\Big|\mathcal{N}\left(x^*, \tfrac{1}{-g_{xx}(x^*,n)}\right) - x^*\Big| \geq A\Big) \leq \sqrt{\frac{2}{\pi|g_{xx}(x^*, t)|A^2}}\, e^{-|g_{xx}(x^*,t)|A^2/2},$$

which goes to zero for sufficiently fast as $t$ grows. The right choice of the central interval size parameter $A$ is key to the success of the method. What we have not yet mentioned is that error arising from originally restricting the integral to the central interval. This is the contribution from the integrals

$$\int_{a}^{x^*-A} e^{g(x,t)}\, \mathrm{d}x \qquad \text{and} \qquad \int_{x^*+A}^{b} e^{g(x,t)}\, \mathrm{d}x.$$

The success of the approximation also depends on this being negligible compared with the contribution from the interval $[x^* - A, x^* + A]$.

### 6.2.1 Laplace's method for Gamma-like functions

The Laplace method will obviously not work for any integral, but it works very well on uni-modal integrals which decay fast outside of the central region around the maximum. A typical example of a function with an integral representation that works well with the Laplace method is the Gamma function

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\ \mathrm{d}x$$

Indeed, using Laplace's method as described above one can prove Stirling's approximation

$$\Gamma(z+1) \sim \sqrt{2\pi z}\left(\frac{z}{e}\right)^z .$$

If more derivatives are considered in the Taylor expansion and a suitable $A$ is found then the approximation can be improved. Indeed the higher order terms bearing the coefficients from Stirling's sequence can also be recovered if Laplace's method is applied to the Gamma function and more derivatives are utilised.

Fortunately for us we can derive an expression for $\mathbb{P}(\gamma_r(v) = k)$ with an integral form that vaguely resembles the Gamma function. We wish to define a family $\mathcal{F}$ of functions which bear resemblance to the integrand of the Gamma function. Recall that throughout we use $o(\cdot), O(\cdot)$ and $\Theta(\cdot)$ in relation to $n$.

**Definition 6.2.1.** *Let $a, b, c : \mathbb{N} \to \mathbb{R}$ be functions of $n$ such that $a(n) \to 0$, $b(n) \to \infty$ and $c(n) = O(1)$ as $n \to \infty$. In addition let $a, b$ and $c$ satisfy $c\log(b) = o\left(|\log(a)|\right)$ and $\log(a)^2 = O(b)$. The family $\mathcal{F}$ consists of all functions $f : \mathbb{R}\times\mathbb{N} \to \mathbb{R}$ of the form*

$$f(x,n) = a(n)^x x^{b(n)-c(n)x},$$

*for some $a, b$ and $c$ as described.*

For ease of presentation we shall purposely neglect to show the dependence on $n$, denoting $a(n)$ by $a$ for example. We also must define what it means for one function to maximise another.

**Definition 6.2.2** (maximiser)**.** *Let $m : \mathbb{N} \to \mathbb{R}$ and $f : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$. We say that the function $m$ is a maximiser of $f$ if the following holds for every $n \in \mathbb{N}$*

$$f(m(n), n) = \max_{x\in\mathbb{R}} f(x,n).$$

**Lemma 6.2.3.** *Let $f \in \mathcal{F}$ have the form $f(x,n) = a^x x^{b-cx}$ given by Definition 6.2.1. Let $m := m(n)$ be a maximiser of $f$. Then*

$$m = \frac{b}{c + c\log(m) - \log(a)} = -\frac{b}{\log(a)}\left(1 + O\left(\frac{c\log(b)}{|\log(a)|}\right)\right).$$

*Let $A := A(n)$ be monotonically increasing and satisfy $\omega\left(\frac{\sqrt{b}}{|\log(a)|}\right) = A = o\left(\frac{b^{2/3}}{|\log(a)|}\right)$. Then*

$$\int\limits_{m-A}^{m+A} f(x,n)\,\mathrm{d}x = f(m,n)\left[\frac{\sqrt{2\pi b}}{|\log(a)|}\left(1 + O\left(\frac{c\log(b)}{|\log(a)|}\right)\right) - O\left(\frac{\sqrt{b}e^{-(A|\log(a)|)^2/2b}}{A|\log(a)|}\right)\right]$$
$$\cdot\left(1 \pm O\left(\frac{(A|\log(a)|)^3}{b^2}\right)\right).$$

*Proof.* Let $g(x,n) := x\log(a) + (b - cx)\log(x)$ so then we have $f(x,n) = e^{g(x,n)}$. Let us compute the first three derivatives of $g$ with respect to $x$

$$\begin{aligned}
g_x(x,n) &= \log(a) + b/x - c - c\log(x)\\
g_{xx}(x,n) &= -b/x^2 - c/x\\
g_{xxx}(x,n) &= 2b/x^3 + c/x^2.
\end{aligned}$$

Thus for each $n$ there is a stationary point at the value $m(n)$ which is the solution to $\log(a(n)) + b(n)/m(n) - c(n) - c(n)\log(m(n)) = 0$. Thus we can write $m$ implicitly as $m = b/(c + c\log(m) - \log(a))$. Recall that $-\log(a) \to \infty$ and $m \geq 0$, this yields

$$\log(m) = \log(b) - \log\left(c + c\log(m) - \log(a)\right) \leq \log(b),$$

and then since $c\log(b) = o\left(|\log(a)|\right)$ by assumption we have

$$m = -\frac{b}{\log(a)}\left(1 + O\left(\frac{c\log(b)}{|\log(a)|}\right)\right).$$

We evaluate the second and third derivatives at the stationary point $m$ to obtain

$$\begin{aligned}
g_{xx}(m,n) &= -\frac{(c + c\log(m) - \log(a))^2}{b} - \frac{c(c + c\log(m) - \log(a))}{b}\\
&= -\frac{\log(a)^2}{b}\left(1 - O\left(\frac{c\log(b)}{|\log(a)|}\right)\right),\\
g_{xxx}(m,n) &= +\frac{(c + c\log(m) - \log(a))^3}{b^2} + \frac{c(c + c\log(m) - \log(a))^2}{b^2},\\
&= -\frac{\log(a)^3}{b^2}\left(1 - O\left(\frac{c\log(b)}{|\log(a)|}\right)\right),
\end{aligned}$$

thus $m$ is a maxima. To begin, any monotonic function $A := A(n)$ satisfying

$$\omega\left(\frac{\sqrt{b}}{|\log(a)|}\right) = A = o\left(\frac{b^{2/3}}{|\log(a)|}\right),$$

gives the following

$$\xi_2 := g_{xx}(m,n)\cdot A^2 = -\frac{\log(a)^2}{b}\left(1 - O\left(\frac{c\log(b)}{|\log(a)|}\right)\right)\cdot A^2 = \omega(1), \tag{6.11}$$

$$\xi_3 := g_{xxx}(m,n)\cdot A^3 = -\frac{\log(a)^3}{b^2}\left(1 - O\left(\frac{c\log(b)}{|\log(a)|}\right)\right)\cdot A^3 = o(1). \tag{6.12}$$

Applying Taylor's theorem we deduce that for $x \in [m-A, m+A]$ we have

$$g(x,n) = g(m,n) + \frac{g_{xx}(m,n)}{2}(x-m)^2 \pm O(\xi_3)\,.$$

Now using this approximation for $g$ in the integral we obtain

$$\int_{m-A}^{m+A} f(x)\,\mathrm{d}x = e^{g(m,n)\pm O(\xi_3)}\int_{m-A}^{m+A}\exp\left(\frac{g_{xx}(m,n)}{2}(x-m)^2\right)\,\mathrm{d}x. \tag{6.13}$$

Observe that if $x = o(1)$ then by the Taylor series for $e$ we have $e^x = 1 + \Theta(x)$. Applying this to the first term of (6.13) and then completing the integral yields

$$\begin{aligned}\int_{m-A}^{m+A} f(x)\,\mathrm{d}x = (1\pm O(\xi_3))\,e^{g(m,n)}\Bigg[&\int_{-\infty}^{\infty}\exp\left(\frac{g_{xx}(m,n)}{2}(x-m)^2\right)\,\mathrm{d}x\\ &-\int_{m+A}^{\infty}\exp\left(\frac{g_{xx}(m,n)}{2}(x-m)^2\right)\,\mathrm{d}x\\ &-\int_{-\infty}^{m-A}\exp\left(\frac{g_{xx}(m,n)}{2}(x-m)^2\right)\,\mathrm{d}x\Bigg].\end{aligned}$$

Label the above integrals $i_1, i_2$ and $i_3$ respectively. We will compute them by recognition with standard Gaussian integrals. To begin observe that we have

$$i_1 = \sqrt{\frac{-2\pi}{g_{xx}(m,n)}} = \frac{\sqrt{2\pi b}}{|\log(a)|}\left(1 + O\left(\frac{c\log(b)}{|\log(a)|}\right)\right).$$

Notice that the second and third integrals, $i_2$ and $i_3$, together give the probability that a Gaussian random variable with mean $m$ and variance $1/|g_{xx}(m,n)|$ takes a value more that $A$ away from its mean. Thus the Mill's ratio (2.5) yields

$$i_1 + i_2 \leq \sqrt{\frac{2}{\pi|g_{xx}(m,n)|}}\cdot\frac{\exp\left(\frac{g_{xx}(m,n)A^2}{2}\right)}{A} = O\left(\frac{\exp(-\xi_2/2)}{\sqrt{\xi_2}}\right).$$

Thus combining the above estimates for $i_1, i_2$ and $i_3$ we have

$$\int_{m-A}^{m+A} f(x)\,\mathrm{d}x = f(m)\left(\frac{\sqrt{2\pi b}}{|\log(a)|}\left(1+O\left(\frac{c\log(b)}{|\log(a)|}\right)\right) - O\left(\frac{e^{-\xi_2/2}}{\sqrt{\xi_2}}\right)\right)(1 \pm O(\xi_3))\,.$$

The result follows by (6.11) and (6.12). □

## 6.3 The Euler-Maclaurin summation formula for Gamma-like sums

The Euler-Maclaurin summation formula is a classical theorem which is useful for approximating sums by integrals. We will now state the first derivative form of this as we shall use it in the proof of Lemma 6.3.2.

**Theorem 6.3.1** (First derivative form of the Euler-Maclaurin summation formula [7])**.** *For any function $f$ with a continuous derivative on the interval $[1, n]$ we have*

$$\sum_{i=1}^{n} f(i) = \int_1^n f(x)\,\mathrm{d}x + \int_1^n (x-[x])f'(x)\,\mathrm{d}x + f(1).$$

Note that in the general version of the Euler-Maclaurin summation formula there are some more terms involving the Bernoulli numbers present and the remainder is a integral involving $k^{th}$ derivatives, where the choice of $k \geq 1$ is left to the user. I have tried using the more general version with higher order derivatives however this appeared to gave no significant improvement of the result so we use the first derivative version of the Theorem for simplicity.

Recall the family $\mathcal{F}$ from Definition 6.2.1 and what it means to be a maximiser of a function, Definition 6.2.2.

**Lemma 6.3.2.** *Let $f \in \mathcal{F}$ have the form $f(x,n) = a^x x^{b-cx}$ given in Definition 6.2.1. Let $m$ be a maximiser of $f$. Let $A := A(n)$ be monotonically increasing in $n$ and satisfy $\omega\left(\log\left(b/|\log(a)|^2+2\right)\frac{\sqrt{b}}{|\log(a)|}\right) = A = o\left(\frac{b^{2/3}}{|\log(a)|}\right)$ and $A^2 = o\left(\frac{b}{c|\log(a)|}\right)$. Then*

*(i)* $$\sum_{x=1}^{m-A} f(x,n) + \sum_{x=m+A}^{\infty} f(x,n) = O\left(f(m,n)\frac{b}{|\log(a)|^2}\exp\left(-\frac{(A\log(a))^2}{b}\right)\right).$$

*(ii)* $$\left|\sum_{x=m-A}^{m+A} f(x,n) - \int_{m-A}^{m+A} f(x,n)\,\mathrm{d}x\right|$$

$$\leq f(m,n)\Bigg(1+O\bigg(\frac{1}{|\log(a)|}\bigg)+O\bigg(\frac{A|\log(a)|}{b}\bigg)+O\bigg(\frac{(A|\log(a)|)^3}{b^2}\bigg)+O\bigg(\frac{cA^2}{m}\bigg)\Bigg).$$

*Proof.* For ease of presentation we will shall abuse notation and write $f(x)$ instead of $f(x,n)$ and use $a$ instead of $a(n)$, hiding the dependence on $n$. To begin we calculate the first derivative of $f$. For each fixed $n$ we have

$$f'(x) := \frac{df(x)}{\mathrm{d}x} = f(x)\cdot\left(\log(a)+\frac{b}{x}-c-c\log(x)\right). \tag{6.14}$$

We will factor $f(m)$ from $f(m+h)$,

$$f(m+h) = a^{m+h}(m+h)^{b-c(m+h)} = f(m)a^h m^{-ch}(1+h/m)^{b-c(m+h)}. \tag{6.15}$$

When $h = o(m)$ the following holds by Taylor's approximation for $\log(1+h/m)$

$$f(m+h) = f(m)a^h m^{-ch} e^{\left(h/m-h^2/2m^2+h^3/3m^3+\cdots\right)(b-c(m+h))}.$$

Recall that $m = b/(c+c\log(m)-\log(a))$, inserting this yields

$$\begin{aligned} f(m+h) &= f(m)a^h m^{-ch} e^{h(c+c\log(m)-\log(a))-ch-h^2b/2m^2+\left(ch^2/2m+h^3b/3m^3\right)(1+o(1))} \\ &= f(m)e^{-h^2b/2m^2+\left(ch^2/2m+h^3b/3m^3\right)(1+o(1))}. \end{aligned} \tag{6.16}$$

*Item* ((i)): If $A+i = o(m)$ then for $j = o(m)$ by (6.16) we have

$$\begin{aligned} \frac{f(m+A+i+j)}{f(m+A+i)} &= \frac{e^{-(A+i+j)^2b/2m^2+\left(c(A+i+j)^2/2m+(A+i+j)^3b/3m^3\right)(1+o(1))}}{e^{-(A+i)^2b/2m^2+(c(A+i)^2/2m+(A+i)^3b/3m^3)(1+o(1))}} \\ &= e^{-\left(2j(A+i)+j^2\right)\left(b/2m^2(1-O((a+i+j)/m))-c/m\right)} \\ &= e^{-\left(j(A+i)+j^2\right)b/m^2(1-o(1))}. \end{aligned} \tag{6.17}$$

We choose $j$ so that $\max\{Aj, j^2\}\cdot(b/m^2) = \Omega(1)$, this is so that the ratio (6.17) is strictly less than 1. Since $m$ is the maximum we know that $f(m+i)$ is decreasing in $i$ thus for such a choice of $j$ we have the following by (6.17)

$$\sum_{i=0}^{m} f(m+A+i) \leq \sum_{k=0}^{\lceil m/j\rceil} j\cdot f(m+A+kj) = O\left(j\cdot f(m+A)\right). \tag{6.18}$$

Otherwise $(A+i)/m > 0$ since $A$ is monotone and we have the following by (6.15)

$$\begin{aligned}\frac{f(m+A+i+1)}{f(m+A+i)} &= \frac{a(1+(A+i+1)/m)^{b-c(m+A+i+1)}}{m^c(1+(A+i)/m)^{b-c(m+A+i)}} \\ &= \frac{a\left[(1+(A+i)/m)\left(1+1/(m+A+i)\right)\right]^{b-c(m+A+i)}}{m^c(1+(A+i)/m)^{b-c(m+A+i)}(1+(A+i+1)/m)^c} \\ &= \frac{a\left(1+1/(m+A+i)\right)^{b-c(m+A+i)}}{m^c(1+(A+i+1)/m)^c},\end{aligned}$$

applying Taylor's approximation for $\log(1+1/(m+a+i))$ yields

$$\frac{f(m+A+i+1)}{f(m+A+i)} = \frac{a\exp\left(\frac{b}{m}\left(1-\frac{A+i}{m+A+i}\right)-c+\frac{b}{2(m+A+i)^2}+o(1)\right)}{m^c(1+(A+i+1)/m)^c}.$$

Recall that $m = b/(c + c\log(m) - \log(a))$, $c\log(m) < -\log(a)$ and $\log(a)^2 = O(b)$ by Lemma 6.2.3 and Definition 6.2.1. Then since $(A+i)/(m+a+i) > \beta$ for some constant $\beta > 0$ we have

$$\begin{aligned}\frac{f(m+A+i+1)}{f(m+A+i)} &= \frac{a\exp\left(c\log(m)-\log(a)+\beta\log(a)+O(1)\right)}{m^c(1+(A+i+1)/m)^c} \\ &= \frac{O\left(a^\beta\right)}{(1+(A+i+1)/m)^c} \\ &= o\left(1\right).\end{aligned}$$

Thus for any $j$ such that $\max\{Aj, j^2\}\cdot(b/m^2) = \Omega\left(1\right)$ we have the following by (6.16), (6.18) and comparison with a geometric series

$$\begin{aligned}\sum_{i=m+A}^{\infty} f(i) &= \sum_{i=m+A}^{2m} f(i) + \sum_{i=2m}^{\infty} f(i) \\ &= O\left(j\cdot f(m+A)\right) + O(f(m+A)) \\ &= O\left(j\cdot f(m)e^{-A^2b/m^2}\right).\end{aligned}$$

The bound on the lower part of the sum is the same. For an upper bound on $j$ take $j = \sqrt{m^2/b} \sim \sqrt{b/|\log(a)|^2}$ as this satisfies $\max\{Aj, j^2\}\cdot(b/m^2) = \Omega\left(1\right)$. Thus

$$\sum_{i=m+A}^{\infty} f(i) = O\left(\sqrt{b/|\log(a)|^2}\cdot f(m)e^{-A^2b/m^2}\right).$$

Provided our choice of $A$ satisfies $A = \omega\left(\log\left(b/|\log(a)|^2+2\right)\frac{\sqrt{b}}{|\log(a)|}\right)$ the term on

the right hand side is $o\left(f(m)\right)$, the extra $+2$ in the log is because it may be the case that $b/|\log(a)|^2 = 1$ and we do not want the expression to be zero.
*Item* ((ii)): By Theorem 6.3.1 and the triangle inequality we have

$$\begin{aligned} \left| \sum_{m-A}^{m+A} f(x) - \int\limits_{m-A}^{m+A} f(x)\,\mathrm{d}x \right| &\leq \int\limits_{m-A}^{m+A} |(x-[x])f'(x)|\,\mathrm{d}x + f(1) \\ &\leq \int\limits_{m-A}^{m+A} |f'(x)|\,\mathrm{d}x + a, \end{aligned} \tag{6.19}$$

since $0 \leq (x-[x]) \leq 1$. For $h \leq A$ equations (6.14) and (6.16) yield

$$\begin{aligned} f'(m+h) &= f(m+h)\left(\log(a) + \frac{b}{m+h} - c - c\log(m+h)\right) \\ &= f(m)e^{-h^2b/2m^2 + \left(h^3b/3m^3 + ch^2/2m\right)(1+o(1))}\left(-\frac{bh}{m^2+mh} - c\log\left(1+\frac{h}{m}\right)\right) \end{aligned}$$

applying Taylor's approximation for $\log(1+h/m)$, where $h = o(m)$ yields

$$\begin{aligned} f'(m+h) &= f(m)e^{-(h^2b/m^2)(1-O(h/m)-O(cm/b))}\left(-\frac{bh}{m^2+mh} - \frac{ch}{m} + O\left(\frac{h^2}{m^2}\right)\right) \\ &= -f(m)e^{-(h^2b/m^2)(1-O(A/m)-O(cm/b))}\frac{bh}{m^2+mh}\left(1+O\left(\frac{m}{b}\right)\right). \end{aligned}$$

Thus we have the following

$$\begin{aligned} \int\limits_{m-A}^{m+A} |f'(x)|\,\mathrm{d}x &= \int\limits_{-A}^{A} |f'(m+h)|\,\mathrm{d}h \\ &= \frac{bf(m)\left(1+O\left(\frac{m}{b}\right)\right)}{m}\int\limits_{-A}^{A}\left|\frac{he^{-(h^2b/m^2)(1-O(A/m)-O(cm/b))}}{m+h}\right|\,\mathrm{d}h. \\ &\leq \frac{bf(m)\left(1+O\left(\frac{m}{b}\right)\right)}{m(m-A)}\int\limits_{-A}^{A}\left|he^{-(h^2b/m^2)}e^{O\left(bA^3/m^3\right)+O\left(cA^2/m\right)}\right|\,\mathrm{d}h. \\ &\leq \frac{2bf(m)\left(1+O\left(\frac{m}{b}\right)+O\left(\frac{A}{m}\right)\right)}{m^2}e^{O\left(bA^3/m^3\right)+O\left(cA^2/m\right)}\int\limits_{0}^{A} he^{-(h^2b/m^2)}\,\mathrm{d}h. \\ &= f(m)\left(1+O\left(\frac{m}{b}\right)+O\left(\frac{A}{m}\right)+O\left(\frac{bA^3}{m^3}\right)+O\left(\frac{cA^2}{m}\right)\right) \\ &\quad \cdot\left[1-e^{-(A^2b/m^2)}\right]. \end{aligned}$$

The result follows by (6.19) since $m \sim -b/\log(a)$ by Lemma 6.2.3 and because of the conditions on $A$ from the assumptions of the Lemma. $\square$

Before we begin with the proof of Theorem (6.0.1) we need the following Lemma.

**Lemma 6.3.3.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$. Let $v \in V$, $r \geq 2$ and $1 \leq k = o(n)$. Then*

$$\mathbb{P}\big(\{\gamma_r(v) = k\} \cap \{\gamma_{r-1}(v) \leq k/e^2np\}\big) \leq e^{-k}.$$

*Proof.* Let $h = \sum_{i=1}^{r-2} \gamma_i(v)$ and we have the following by Lemma 4.1.1

$$\begin{aligned}\mathfrak{P} &:= \mathbb{P}\Big(\gamma_r(v) = k \,\Big|\, \{\gamma_{r-1}(v) \leq k/e^2np\} \cap \sigma(h)\Big) \\ &\leq \mathbb{P}\Big(Bin\left(n-h, 1-(1-p)^{k/e^2np}\right) = k\Big)\end{aligned}$$

Now since $1-(1-p)^{k/e^2np} \leq k/e^2n$ by the Bernoulli inequality (2.4) and $k = o(n)$ we have

$$\begin{aligned}\mathfrak{P} &\leq \mathbb{P}\big(Bin\left(n-h, k/e^2n\right) = k\big) \\ &\leq \binom{n-h}{k}\left(\frac{k}{e^2n}\right)^k\left(1-\frac{k}{e^2n}\right)^{n-h} \\ &\leq \left(\frac{ne}{k}\right)^k\left(\frac{k}{e^2n}\right)^k \\ &\leq e^{-k}.\end{aligned}$$

Since this bound is independent of $h$ we have

$$\mathbb{P}\big(\{\gamma_r(v) = k\} \cap \{\gamma_{r-1}(v) \leq k/e^2np\}\big) \leq e^{-k}.$$

$\square$

## 6.4 Proof of Theorem 6.0.1

We are now ready to prove the main theorem of this chapter.

*Proof of Theorem 6.0.1.* We will first consider the distribution of $\gamma_2(u)$. Observe the following

$$\mathbb{P}(\gamma_2(u)=k) = \sum_{l=1}^{n-k-1} \mathbb{P}\big(\gamma_2(u)=k\big|\gamma_1(u)=l\big)\,\mathbb{P}(\gamma_1(u)=l)$$
$$= \sum_{l=1}^{\lfloor 4np\rfloor} \mathbb{P}\big(\gamma_2(u)=k\big|\gamma_1(u)=l\big)\,\mathbb{P}(\gamma_1(u)=l) + o\left(e^{-2np}\right), \qquad (6.20)$$

since $\mathbb{P}(\gamma_1(u) > \lfloor 4np\rfloor) \leq \exp\left(-\frac{9np}{2(1+(3/3))}\right) = o\left(e^{-2np}\right)$ by Lemma 2.1.1. Let $\mathfrak{P}_k$ denote the sum in (6.20) above so that we can ignore the $o\left(e^{-2np}\right)$ error term for the time being. Recall that conditional on $\gamma_1(u)$ we have $\gamma_2(u) \sim_d Bin(n-\gamma_1(u)-1, 1-(1-p)^{\gamma_1(u)})$ by Lemma 4.1.1, thus we have

$$\mathfrak{P}_k = \sum_{l=1}^{\lfloor 4np\rfloor} \binom{n-1-l}{k}\left(1-(1-p)^l\right)^k (1-p)^{l(n-1-l-k)}\binom{n-1}{l}p^l(1-p)^{n-1-l}.$$

Applying the identity $\binom{n-1}{l}\binom{n-1-l}{k} = \binom{n-1}{k}\binom{n-k-1}{l}$ and collecting some lower order terms we obtain

$$\mathfrak{P}_k = \binom{n-1}{k}(1-p)^n \sum_{l=1}^{\lfloor 4np\rfloor} \binom{n-k-1}{l}\left(1-(1-p)^l\right)^k (1-p)^{ln+O(lk)}p^l.$$

By inequalities (2.3) and (2.4) and since $k = O\big((np)^2\big)$ and $l = O(np)$ we have

$$\left(1-(1-p)^l\right)^k = (lp)^k\left(1-\frac{lp}{2}+O\left(\frac{(lp)^2}{3!}\right)\right)^k = (lp)^k e^{-O(lkp)}. \qquad (6.21)$$

the Taylor approximation to the logarithm yields $(1-p)^a = e^{-pa+O(ap^2)}$, thus

$$\mathfrak{P}_k = \binom{n-1}{k}e^{-np+O(n^3p^4)}p^k \sum_{l=1}^{\lfloor 4np\rfloor}\binom{n-k-1}{l} l^k e^{-lnp}p^l,$$

where we bounded and collected all errors in the second term. Observe

$$\binom{n}{k} = \frac{n\cdot(n-1)\cdots(n-k+1)}{k!} = \frac{1}{\sqrt{2\pi k}}\left(\frac{en}{k}\right)^k\left(1-O\left(\tfrac{k^2}{n}\right)\right), \qquad (6.22)$$

by Stirling's approximation, provided $k = o\left(\sqrt{n}\right)$. Thus

$$\mathfrak{P}_k = \frac{e^{-np}\left(1\pm O\big(n^3p^4\big)\right)}{2\pi\sqrt{k}}\left(\frac{enp}{k}\right)^k \sum_{l=1}^{\lfloor 4np\rfloor}\left(e^{1-np}np\right)^l l^{k-1/2-l}. \qquad (6.23)$$

Our plan is to recognise that the summand above has the form of a function from the family $\mathcal{F}$ for some suitable $a := a(n)$, $b := b(n)$ and $c := c(n)$. We shall then bound the sum from above using Lemmas 6.3.2 and 6.2.3.

To this end let $a = e^{1-np}np$, $b = k - 1/2$, $c = 1$ and $f(l,n) := a^l l^{b-l}$. Thus $f(l,n)$ is the summand in (6.23) and we notice that $f \in \mathcal{F}$ (see Definition 6.2.1) and $m \sim k/np$ by Lemma 6.2.3. Also observe that $b/|\log(a)|^2 = \Theta(1)$. Choose $A := A(n) = \log(np)$ and we see that $\omega(1) = A = o\left((np)^{1/3}\right)$. Thus we can apply Lemma 6.3.2 to give the following upper and lower bounds

$$\sum_{l=1}^{\lfloor 4np \rfloor} f(l,n) \leq \int\limits_{m-A}^{m+A} f(l,n) + f(m,n)\left(1 + O\left(\frac{(\log(np))^2}{np}\right) + O\left(e^{-\Theta\left(\log(np)^2\right)}\right)\right), \quad (6.24)$$

$$\sum_{l=1}^{\lfloor 4np \rfloor} f(l,n) \geq \int\limits_{m-A}^{m+A} f(l,n) - f(m,n)\left(1 + O\left(\frac{(\log(np))^2}{np}\right) + O\left(e^{-\Theta\left(\log(np)^2\right)}\right)\right). \quad (6.25)$$

Again since $\omega(1) = A = o\left((np)^{1/3}\right)$ we can apply Lemma 6.2.3 to give

$$\begin{aligned}\int_{l=m-A}^{m+A} f(l,n) = f(m,n) &\left[\frac{\sqrt{2\pi(k-1/2)}}{np - \log(np) - 1}\left(1 + O\left(\frac{\log(np)}{np}\right)\right)\right. \\ &\left. - O\left(\frac{e^{-\Theta\left(\log(np)^2\right)}}{\sqrt{\log(np)}}\right)\right]\left(1 \pm O\left(\frac{(\log(np))^3}{np}\right)\right). \end{aligned} \quad (6.26)$$

Combining (6.24) and (6.26) then collecting error terms yields the upper bound

$$\sum_{l=1}^{\lfloor 4np \rfloor} f(l,n) \leq f(m,n)\left(\frac{\sqrt{2\pi k}}{np} + 1 + O\left(\frac{(\log(np))^3}{np}\right)\right). \quad (6.27)$$

Likewise (6.25), (6.26) and collecting error terms provide the lower bound

$$\sum_{l=1}^{\lfloor 4np \rfloor} f(l,n) \geq f(m,n)\left(\frac{\sqrt{2\pi k}}{np} - 1 - O\left(\frac{(\log(np))^3}{np}\right)\right). \quad (6.28)$$

All that remains it to compute the value of $f(m,n)$, for large $n$.

Observe that $f(l,n) = e^{g(l,n)}$ where $g(l,n) := (1 - np + \log(np)) \cdot l + \log(l) \cdot (k - 1/2 - l)$. Let us compute the first derivative of $g$ with respect to $l$,

$$\frac{dg}{\mathrm{d}l} := g_l(m,n) = -np + \frac{k - 1/2}{l} + \log(np) - \log(l).$$

We can find the following implicit form for the solution to $g_l(m, n) = 0$

$$m = \frac{k - 1/2}{np + \log(m) - \log(np)}. \tag{6.29}$$

We need to estimate $\log(m)$. Let $\alpha = k/(np)^2$, and observe that by (6.29):

$$\log(m) = \log(\alpha) + \log(np) + \log(1 - 1/2k) - \log(1 + (\log(m) - \log(np))/np).$$

Applying the Taylor expansion for $\log(1 + x)$ we obtain

$$\begin{aligned}\log(m) = \log(\alpha) + \log(np) - \tfrac{1}{2k} - O\big(\tfrac{1}{k^2}\big) - \tfrac{\log(m) - \log(np)}{np} \\ + \tfrac{(\log(m) - \log(np))^2}{2(np)^2} - O\Big(\tfrac{(\log(m) - \log(np))^3}{(np)^3}\Big).\end{aligned} \tag{6.30}$$

Let $t_m = \log(m) - \log(np)$. Inserting the first few terms from (6.30) yields

$$\begin{aligned}t_m &= \log(\alpha) - \tfrac{1}{2k} - O\big(\tfrac{1}{k^2}\big) - \tfrac{t_m}{np} + \tfrac{t_m^2}{(np)^2} - O\Big(\tfrac{t_m^3}{(np)^3}\Big) \\ &= \log(\alpha) - \log(\alpha)/np \pm O\left(1/(np)^2\right).\end{aligned} \tag{6.31}$$

Now using the estimate (6.31) for $t_m$ in the expression (6.30) for $\log(m)$ we have

$$\log(m) = \log(\alpha) + \log(np) - \tfrac{\log(\alpha)}{np} + \tfrac{2\log(\alpha) + \log(\alpha)^2 - 1/\alpha}{2(np)^2} \pm O\Big(\tfrac{1}{(np)^3}\Big). \tag{6.32}$$

We now apply (6.32) to (6.29) to obtain an explicit asymptotic expression for $m$

$$m = \frac{k - 1/2}{np + \log(\alpha) - \frac{\log(\alpha)}{np} + \frac{2\log(\alpha) + \log(\alpha)^2 - 1/\alpha}{2(np)^2} \pm O\Big(\frac{1}{(np)^3}\Big)}. \tag{6.33}$$

Observe that for $a, b, c \in \mathbb{R}$ and $x \to \infty$ we have the following Laurent series

$$\frac{1}{x + a + b/x + c/x^2} = \frac{1}{x} - \frac{a}{x^2} + \frac{a^2 - b}{x^3} \pm O\left(\frac{1}{x^4}\right).$$

Using the Laurent series to expand the denominator of (6.33) we have

$$m = \alpha(np) - \alpha\log(\alpha) + \frac{\alpha\log(\alpha)^2 + \alpha\log(\alpha) - 1/2}{(np)} \pm O\left(\frac{1}{(np)^2}\right). \tag{6.34}$$

Let $a_g := (1 - np + \log(np)) \cdot m$ and $b_g := \log(m) \cdot (k - 1/2 - m)$. We can now calculate $a_g$ up to terms of lower order using (6.32) and (6.34):

$$\begin{aligned}
a_g &= (1 - np + \log(np))\Big(\alpha(np) - \alpha\log(\alpha) + \tfrac{\alpha\log(\alpha)^2 + \alpha\log(\alpha) - 1/2}{(np)} \pm O\Big(\tfrac{1}{(np)^2}\Big)\Big) \\
&= \alpha(np) - \alpha\log(\alpha) - \alpha(np)^2 + \alpha\log(\alpha)np - \alpha\log(\alpha)^2 - \alpha\log(\alpha) + 1/2 \\
&\qquad + \alpha(np)\log(np) - \alpha\log(\alpha)\log(np) \pm O\Big(\tfrac{\log(np)}{np}\Big) \\
&= -\alpha(np)^2 + \alpha\log(np)np + \alpha\left(1 + \log(\alpha)\right)np - \alpha\log(\alpha)\log(np) \\
&\qquad - \alpha\log(\alpha)^2 - 2\alpha\log(\alpha) + \tfrac{1}{2} \pm O\Big(\tfrac{\log(np)}{np}\Big).
\end{aligned}$$

Again using (6.32) and (6.34) we obtain the following estimate for $b_g$

$$\begin{aligned}
b_g &= \Big(\log(\alpha) + \log(np) - \tfrac{\log(\alpha)}{np} + \tfrac{2\log(\alpha) + \log(\alpha)^2 - 1/\alpha}{2(np)^2} \pm O\Big(\tfrac{1}{(np)^3}\Big)\Big) \\
&\qquad \cdot \Big(\alpha(np)^2 - \tfrac{1}{2} - \alpha(np) + \alpha\log(\alpha) + O\Big(\tfrac{1}{np}\Big)\Big) \\
&= \alpha\log(\alpha)(np)^2 + \alpha\log(np)(np)^2 - \alpha\log(\alpha)np + \alpha\log(\alpha) + \tfrac{\alpha\log(\alpha)^2}{2} - \tfrac{1}{2} \\
&\qquad - \tfrac{\log(\alpha)}{2} - \tfrac{\log(np)}{2} - \alpha\log(\alpha)np - \alpha\log(np)np + \alpha\log(\alpha) \\
&\qquad + \alpha\log(\alpha)^2 + \alpha\log(\alpha)\log(np) + O\Big(\tfrac{\log(np)}{np}\Big) \\
&= \alpha\log(np)(np)^2 + \alpha\log(\alpha)(np)^2 - \alpha\log(np)np - 2\alpha\log(\alpha)np \\
&\qquad + \left(\alpha\log(\alpha) - \tfrac{1}{2}\right)\log(np) + \tfrac{3\alpha\log(\alpha)^2 + 4\alpha\log(\alpha) - \log(\alpha) - 1}{2} + O\Big(\tfrac{\log(np)}{np}\Big).
\end{aligned}$$

We then combine to give $g(m,n) = a_g + b_g$ and thankfully some terms cancel:

$$\begin{aligned}
g(m,n) = \alpha\log(np)(np)^2 &+ \alpha\left(\log(\alpha) - 1\right)(np)^2 + \alpha\left(1 - \log(\alpha)\right)np \\
&- \tfrac{\log(np)}{2} + \tfrac{\alpha\log(\alpha)^2 - \log(\alpha)}{2} + O\Big(\tfrac{\log(np)}{np}\Big).
\end{aligned}$$

Finally, since $f(m,n) = e^{g(m,n)}$ and $e^x = 1 + \Theta(x)$ whenever $x = o(1)$, we have

$$f(m) = \left(\frac{k}{npe}\right)^k \frac{e^{(\alpha - \alpha\log(\alpha))np}}{\sqrt{np}} \cdot \frac{\alpha^{\frac{\alpha\log(\alpha)}{2}}}{\sqrt{\alpha}} \left(1 + O\Big(\tfrac{\log(np)}{np}\Big)\right).$$

Thus by (6.20), (6.23) and (6.27) we have the following the upper bound

$$
\begin{aligned}
\mathbb{P}(\gamma_2(u) = k) &\leq \frac{e^{-np}\left(1 \pm O\big(n^3p^4\big)\right)}{2\pi\sqrt{k}} \left(\frac{enp}{k}\right)^k f(m) \left(\frac{\sqrt{2\pi k}}{np} + 1 + O\Big(\frac{(\log(np))^3}{np}\Big)\right) \\
&\quad + o\left(e^{-2np}\right) \\
&= \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{(np)^{3/2}} \left(\frac{\alpha^{\frac{\alpha\log(\alpha)}{2}}}{\alpha\sqrt{2\pi}} \left(\sqrt{2\pi\alpha} + 1 + o(1)\right)\right) \\
&\leq C \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{(np)^{3/2}}.
\end{aligned}
$$

for some constant $C$ since $\alpha = k/(np)^2 = \Theta(1)$. Similarly, provided that $\alpha > 1/2\pi$, we have the following lower bound by (6.20), (6.23) and (6.28)

$$
\begin{aligned}
\mathbb{P}(\gamma_2(u) = k) &\geq \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{(np)^{3/2}} \left(\frac{\alpha^{\frac{\alpha\log(\alpha)}{2}}}{\alpha\sqrt{2\pi}} \left(\sqrt{2\pi\alpha} - 1 - o(1)\right)\right) \\
&\geq c \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{(np)^{3/2}}.
\end{aligned}
$$

for some constant $c > 0$ since $\alpha > 1/2\pi$.

We now consider $\mathbb{P}(\gamma_r(u) = k)$ when $r \geq 3$. We shall use our upper and lower bounds on $\mathbb{P}(\gamma_2(u) = k)$ as the basis for an induction on $r$. Observe the following for some $\beta > 3$ to be chosen later.

$$
\begin{aligned}
\mathbb{P}(\gamma_{r+1}(u) = k) &= \sum_{l=1}^{n-k} \mathbb{P}(\{\gamma_{r+1}(u) = k\} \cap \{\gamma_r(u) = l\}) \\
&= \sum_{l=k/e^2np}^{\beta^2(np)^r} \mathbb{P}(\{\gamma_{r+1}(u) = k\} \cap \{\gamma_r(u) = l\}) + \mathbb{P}\big(\gamma_r(u) \geq \beta^2(np)^r\big) \\
&\qquad + \mathbb{P}\big(\{\gamma_{r+1}(u) = k\} \cap \big\{\gamma_r(u) \leq k/e^2np\big\}\big) \\
&= \sum_{l=k/e^2np}^{\beta^2(np)^r} \mathbb{P}(\{\gamma_{r+1}(u) = k\} \cap \{\gamma_r(u) = l\}) \qquad (6.35) \\
&\quad + o\Big(e^{-\frac{3(\beta-3)np}{2}}\Big) + O\Big(e^{-k}\Big),
\end{aligned}
$$

where we bounded the last two terms using Lemmas 4.1.2 and 6.3.3 respectively. Let $\mathfrak{B}$ denote the sum in line (6.35) and $\mathcal{I}_r$ be the event $\big\{|B_r(u)| \leq (2\beta^2 + 1)(np)^r\big\}$.

Then

$$\mathfrak{B} \leq \sum_{l=k/e^2np}^{\beta^2(np)^r} \mathbb{P}(\{\gamma_{r+1}(u) = k\} \cap \mathcal{I}_{r-1} \cap \{\gamma_r(u) = l\}) + \mathbb{P}((\mathcal{I}_{r-1})^c)$$

Now since $\mathbb{P}((\mathcal{I}_{r-1})^c) = o\Big(e^{-\frac{3(\beta-3)np}{2}}\Big)$ by Lemma 4.1.2 we have

$$\mathbb{P}(\mathcal{I}_{r-1} \cap \{\gamma_r(u) = l\}) = \mathbb{P}(\gamma_r(u) = l) - o\Big(e^{-\frac{3(\beta-3)np}{2}}\Big).$$

Since $k = \omega(np)$ we have $e^{-k} = o\Big(e^{-\frac{3(\beta-3)np}{2}}\Big)$ for any $\beta > 0$, thus

$$\mathbb{P}(\gamma_{r+1}(u) = k) = \sum_{l=k/e^2np}^{\beta^2(np)^r} \mathbb{P}\Big(\gamma_{r+1}(u) = k \Big| \mathcal{I}_{r-1} \cap \{\gamma_r(u) = l\}\Big) \, \mathbb{P}(\gamma_r(u) = l) \pm o\Big(e^{-\frac{3(\beta-3)np}{2}}\Big) \tag{6.36}$$

Let $\mathfrak{F}_r$ be the filtration $\sigma(\gamma_i(v); 0 \leq i \leq r)$ and let $h = \sum_{i=0}^{r-1} \gamma_i(u)$. Observe that we have the following by Lemma 4.1.1

$$\begin{aligned}
\mathfrak{K} &:= \mathbb{P}\Big(\gamma_{r+1}(u) = k \Big| \mathfrak{F}_{r-1} \cap \{\gamma_r(u) = l\}\Big) \\
&= \mathbb{P}\Big(Bin\left(n - l - h, 1 - (1-p)^l\right) = k\Big) \\
&= \binom{n-l-h}{k} (1 - (1-p)^l)^k (1-p)^{l(n-l-h-k)}
\end{aligned}$$

Similarly to (6.21) we have $\left(1 - (1-p)^l\right)^k = (lp)^k e^{-O(lkp)}$, thus

$$\mathfrak{K} = \binom{n-l-h}{k} (lp)^k e^{-pl(n-h) \pm O(lkp)}$$

Now notice that on the event $\mathcal{I}_{r-1}$ we have the following by (6.22)

$$\begin{aligned}
\mathfrak{K} \cdot \mathbf{1}_{\mathcal{I}_{r-1}} &= \frac{1}{\sqrt{2\pi k}} \left(\frac{ne}{k}\left(1 - \frac{O(l)}{n}\right)\right)^k \left(1 - O\Big(\frac{k^2}{n}\Big)\right) (lp)^k e^{-npl \pm O(lkp)} \\
&= \frac{\left(1 \pm O\big(k^2/n\big)\right)}{\sqrt{2\pi k}} \left(\frac{npe}{k}\right)^k l^k e^{-npl},
\end{aligned}$$

since conditional on $\mathcal{I}_{r-1}$ we have $(l+h)k = O(lk)$. Let $\mathfrak{P}_u$ denote the sum in Equation (6.36) so we can temporally ignore the error term. Notice that the summand of the sum in (6.36) is precisely $\left(\mathbb{E}\left[\mathfrak{K} \cdot \mathbf{1}_{\mathcal{I}_{r-1}}\right] / \mathbb{P}(\mathcal{I}_{r-1})\right) \mathbb{P}(\gamma_r(u) = l)$, thus

$$\mathfrak{P}_u = \frac{(npe)^k \left(1 \pm O\left(k^2/n\right)\right)}{\sqrt{2\pi}\,(k)^{k+1/2}} \sum_{l=k/e^2np}^{\beta^2(np)^r} l^k e^{-pln} \mathbb{P}(\gamma_r(u) = l)\,. \tag{6.37}$$

Recall $\alpha = (l/(np)^r)$. We shall assume the following inductive hypothesis: for $\alpha = \Theta(1)$ there exists a constant $0 < C < \infty$ such that

$$\mathbb{P}(\gamma_r(u) = l) \leq C \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{\sqrt{(np)^{2r-1}}} \left(1 + O\left(\frac{l^2}{n}\right)\right),$$

and if in addition $\alpha > 1/2\pi$ then there exists a constant $0 < c < \infty$ such that

$$\mathbb{P}(\gamma_r(u) = l) \geq c \cdot \frac{e^{(\alpha - \alpha\log(\alpha) - 1)np}}{\sqrt{(np)^{2r-1}}} \left(1 - O\left(\frac{l^2}{n}\right)\right).$$

Applying the inductive hypothesis to (6.37) yields the upper bound

$$\mathfrak{P}_u \leq \frac{C(npe)^k \left(1 + O\left(k^2/n\right)\right)}{(k)^{k+1/2}\sqrt{2\pi(np)^{2r-1}}} \sum_{l=k/e^2np}^{\beta^2(np)^r} l^k e^{-pln} \cdot e^{(\alpha - \alpha\log(\alpha) - 1)np}. \tag{6.38}$$

To bound $\mathfrak{P}_u$ from below, in the case that $k/(np)^{r+1} > 1/2\pi$, we discard the lower tail of the sum (6.37) and apply the inductive hypothesis, this yields

$$\mathfrak{P}_u \geq \frac{c(npe)^k \left(1 - O\left(k^2/n\right)\right)}{(k)^{k+1/2}\sqrt{2\pi(np)^{2r-1}}} \sum_{l=\lfloor k/np - (np)^r/\log(np)\rfloor}^{\beta^2(np)^r} l^k e^{-pln} \cdot e^{(\alpha - \alpha\log(\alpha) - 1)np}. \tag{6.39}$$

This is valid since $l/(np)^r \geq k/(np)^{r+1} - O(1/\log(np))$, thus for large enough $n$ by conditions on $k$ we have $l/(np)^r > 1/2\pi$.

The sums in (6.38) and (6.39), with summand $f(l,n) := e^{-pln} {\cdot} e^{(\alpha - \alpha\log(\alpha) - 1)np}$, above are identical and by expressing $\alpha = l/(np)^r$ in full we have

$$\begin{aligned} f(l,n) &= l^k e^{-pln} \left(\frac{e(np)^r}{l}\right)^{l/(np)^{r-1}} \frac{e^{-np}}{\sqrt{(np)^{2r-1}}} \\ &= l^{k - l/(np)^{r-1}} \left(e^{(1 + r\log(np))/(np)^{r-1} - np}\right)^l. \end{aligned}$$

Once again we wish to bound a sum from above using Lemmas 6.3.2 and 6.2.3. Let $a = e^{(1 + r\log(np))/(np)^{r-1} - np}$, $b = k$ and $c = (np)^{-r+1}$. Then letting $f(l,n) := a^l l^{b-cl}$ denote the summand in (6.23) we notice that $f \in \mathcal{F}$ and $m \sim k/np$. This time however we have $b/|\log(a)|^2 = \Theta\left((np)^{r-1}\right)$. Let $A = (np)^{3(r-1)/5}$ so that $\omega\left(r \cdot \log(np) \cdot (np)^{(r-1)/2}\right) = A = o\left((np)^{(2r-1)/3}\right)$ and also $A = o\left((np)^{2r+1}\right)$. Thus Lemma 6.3.2 (i) yields

$$\sum_{l=k/e^2np}^{\beta^2(np)^r} f(l,n) = \sum_{l=m-A}^{m+A} f(l,n) + O\Big((np)^{r-1} f(m,n) e^{-np^{(r-1)/5}}\Big).$$

Now we can apply Lemma 6.3.2 (ii) to the sum term above to obtain

$$\sum_{l=m-A}^{m+A} f(l,n) = \int_{l=m-A}^{m+A} f(l,n) \pm f(m,n)\left(1 + O\left(\frac{1}{np}\right) + O\left(\frac{1}{(np)^{(2r+1)/5}}\right)\right.$$
$$\left. + O\left(\frac{1}{(np)^{(r+4)/5}}\right) + O\left(\frac{1}{(np)^{(4r+11)/5}}\right)\right).$$

The integral above can be expressed in terms of $f(m,n)$ by Lemma 6.2.3 giving

$$\int_{m-A}^{m+A} f(l,n)\,\mathrm{d}l = f(m,n)\left[\frac{\sqrt{2\pi k}}{np - (1 + r\log(np))/(np)^{r-1}}\left(1 + O\left(\frac{r\log(np)}{(np)^r}\right)\right)\right.$$
$$\left. - O\left(\frac{e^{-(np)^{(r-1)/5}}}{(np)^{(r-1)/5}}\right)\right]\left(1 \pm O\left(\frac{1}{(np)^{(r+4)/5}}\right)\right).$$

Hence, combining the three estimates above yields

$$\sum_{l=k/e^2np}^{\beta^2(np)^r} f(l,n) = f(m,n)\frac{\sqrt{2\pi k}}{np}\left(1 \pm O\left(\frac{1}{(np)^{(r-1)/2}}\right)\right) \tag{6.40}$$

Notice that the term $\sqrt{2\pi b}/|\log(a)|$ from Lemma 6.2.3 is $\Theta\left((np)^{(r-1)/2}\right)$. Hence any error terms not multiplied by this term will be relatively very small. When we were estimating $\mathbb{P}(\gamma_2(u) = k)$ the parameters gave $\frac{\sqrt{2\pi k}}{np} = \Theta(1)$ and so the error term from Lemma 6.3.2 (ii) had the same order as the integral.

Let $g(l,n) = \left((1 + r\log(np))/(np)^{r-1} - np\right) l + \log(l)\left(k - l/(np)^{r-1}\right)$. Thus

$$g_l(l,n) := \frac{dg(l,n)}{\mathrm{d}l} = \frac{r\log(np)}{(np)^{r-1}} - np + \frac{k}{l} - \frac{\log(l)}{(np)^{r-1}}.$$

Hence we can find the following implicit form for the solution to $g_l(m,n) = 0$

$$m = \frac{k}{np + (\log(m) - r\log(np))/(np)^{r-1}}. \tag{6.41}$$

We need to estimate $\log(m)$. Let $\alpha = k/(np)^{r+1}$ and observe that by (6.41):

$$\log(m) = \log(\alpha) + r\log(np) - \log\left(1 + (\log(m) - r\log(np))/(np)^r\right). \tag{6.42}$$

Then using the Taylor expansion for $\log(1+x)$ we have

$$\begin{aligned}\log(m) = \log(\alpha) + r\log(np) - \frac{\log(m) - r\log(np)}{(np)^r} \\ + \frac{(\log(m) - r\log(np))^2}{(np)^{2r}} - O\left(\frac{(\log(m) - r\log(np))^3}{(np)^{3r}}\right).\end{aligned} \tag{6.43}$$

Let $t_m = \log(m) - r\log(np)$. Inserting the first few terms of (6.42) yields

$$\begin{aligned}t_m &= \log(\alpha) - \frac{t_m}{(np)^r} + \frac{t_m^2}{(np)^{2r}} - O\left(\frac{t_m^3}{(np)^{3r}}\right) \\ &= \log(\alpha) - \frac{\log(\alpha)}{(np)^r} \pm O\left(\frac{1}{(np)^{2r}}\right).\end{aligned} \tag{6.44}$$

Then using the estimate (6.44) for $t_m$ in the estimate (6.43) for $\log(m)$ we have

$$\begin{aligned}\log(m) = \log(\alpha) + r\log(np) - \frac{\log(\alpha)}{(np)^r} \\ + \frac{2\log(\alpha) + \log(\alpha)^2}{2(np)^{2r}} \pm O\left(\frac{1}{(np)^{3r}}\right).\end{aligned} \tag{6.45}$$

We can use this to get the following expression for $m$

$$m = \frac{k}{np + \frac{\log(\alpha)}{(np)^{r-1}} - \frac{\log(\alpha)}{(np)^{2r-1}} + \frac{2\log(\alpha)+\log(\alpha)^2}{2(np)^{3r-1}} \pm O\left(\frac{1}{(np)^{4r-1}}\right)}. \tag{6.46}$$

Observe that for $a, b, c \in \mathbb{R}$ and $x \to \infty$ we have the following Laurent series

$$\frac{1}{x + a/x^{r-1} + b/x^{2r-1} + c/x^{3r-1}} = \frac{1}{x} - \frac{a}{x^{r+1}} + \frac{a^2 - b}{x^{2r+1}} \pm O\left(\frac{1}{x^{3r+1}}\right).$$

By expanding the denominator of (6.46) using the Laurent series we obtain

$$m = \alpha(np)^r - \alpha\log(\alpha) + \frac{\alpha\log(\alpha)^2 + \alpha\log(\alpha)}{(np)^r} \pm O\left(\frac{1}{(np)^{2r}}\right). \tag{6.47}$$

Let $a_g := \big((1 + r\log(np))/(np)^{r-1} - np\big)\, m$ and $b_g := \log(m)\,\big(k - m/(np)^{r-1}\big)$, thus $g(m,n) = a_g + b_g$. We will estimate $a_g$ and $b_g$ using (6.45) and (6.47):

$$\begin{aligned}
a_g &= \left(\frac{1 + r\log(np)}{(np)^{r-1}} - np\right)\left(\alpha(np)^r - \alpha\log(\alpha) \pm O\left(\frac{1}{(np)^r}\right)\right) \\
&= \alpha np + \alpha r\log(np)np - \alpha(np)^{r+1} + \alpha\log(\alpha)np \pm O\left(\frac{1}{(np)^{r-1}}\right), \\
b_g &= \left(\log(\alpha) + r\log(np) - \frac{\log(\alpha)}{(np)^r} \pm O\left(\frac{1}{(np)^{2r}}\right)\right) \\
&\quad \cdot \left(\alpha(np)^{r+1} - \alpha np + O\left(\frac{1}{(np)^{r-1}}\right)\right) \\
&= \alpha r\log(np)(np)^{r+1} + \alpha\log(\alpha)(np)^{r+1} - 2\alpha\log(\alpha)np \\
&\quad - \alpha r\log(np)np \pm O\left(\frac{1}{(np)^{r-1}}\right).
\end{aligned}$$

Hence $g(m,n)$ is equal to

$$g(m,n) = \alpha\log(\alpha(np)^r)(np)^{r+1} - \alpha(np)^{r+1} + \alpha\,(1 - \log(\alpha))\, np \pm O\left(\frac{1}{(np)^{r-1}}\right).$$

Thus we have

$$f(m,n) = \left(\frac{k}{enp}\right)^k e^{\alpha(1-\log(\alpha))np}.$$

Therefore by (6.36), (6.38) and (6.40) we have the following for any $\beta > 3$

$$\begin{aligned}
\mathbb{P}(\gamma_{r+1}(u) = k) &\le \frac{C(npe)^k\,\big(1 + O(k^2/n)\big)}{\sqrt{2\pi}\,(k)^{k+1/2}\sqrt{(np)^{2r-1}}} f(m,n)\frac{\sqrt{2\pi k}}{np}\left(1 + O\left(\frac{1}{(np)^{(r-1)/2}}\right)\right) \\
&\quad + o\left(e^{-\frac{3(\beta-3)np}{2}}\right) \\
&= C \cdot \frac{e^{(\alpha-\alpha\log(\alpha)-1)np}}{\sqrt{(np)^{2r+1}}}\left(1 + O\left(\frac{k^2}{n}\right)\right).
\end{aligned}$$

Similarly by (6.36), (6.39) and (6.40) we obtain the corresponding lower bound when $\alpha > 1/2\pi$

$$\mathbb{P}(\gamma_{r+1}(u) = k) \ge c \cdot \frac{e^{(\alpha-\alpha\log(\alpha)-1)np}}{\sqrt{(np)^{2r+1}}}\left(1 - O\left(\frac{k^2}{n}\right)\right).$$

□

# Chapter 7

# Vertices with the smallest $r$-neighbourhoods in $\mathcal{G}(n, p)$

Recall that for a vertex $u \in V$ and some $r \geq 0$ the $r$-neighbourhood of a vertex $u$, denoted $\Gamma_r(u)$, is the collection of vertices at distance precisely $r$ from $u$, so once again

$$\Gamma_r(u) := \{v \in V \,:\, d(u, v) = r\}\,,$$

and we use $\gamma_r(u)$ to denote $|\Gamma_r(u)|$, the size of the $r$-neighbourhood of $u$. Let

$$\delta_r := \min_{u \in V} \gamma_r(u),$$

be the size of the smallest $r$-neighbourhood in the graph.

The main result of this chapter, Theorem 7.3.4, is an upper bound on the number of vertices which attain an $r$-neighbourhood of minimum size, i.e. the number of $u \in V$ such that $\gamma_r(u) = \delta_r(\mathcal{G})$. Let $X_{i,k}$ be the collection of vertices with an $i^{th}$-neighbourhood of size $k$, more formally

$$X_{i,k} := \{v \in V : \gamma_i(v) = k\}\,.$$

The difficulty with controlling the sizes of the sets $X_{r,k}$ for $r \geq 2$ is that despite the results of Theorem 6.0.1 we still know relatively little about the distribution of the sizes of the $r$-neighbourhoods in $\mathcal{G}(n, p)$. In particular we do not know the distribution of $\delta_r(\mathcal{G})$ or the joint distributions of the sizes of $r$-neighbourhoods of different vertices.

## 7.1 Relating $r$ -neighbourhood to first neighbourhoods

In this section we present a theorem which goes some way towards addressing the problem of not knowing the distribution of $\delta_r(\mathcal{G})$ by relating the $r$-neighbourhoods of vertices to their first neighbourhoods. The theorem is a concentration result for $\gamma_r(u)$. It differs slightly from standard $r$-neighbourhood growth estimates, such as those of Lemma 4.1.2, as we have concentration of $\gamma_r(u)$ around $\gamma_1(u)(np)^{r-1}$ as opposed to around $\mathbb{E}[\gamma_r(u)] \approx (np)^r$. By asking for concentration around $\gamma_1(u)(np)^{r-1}$, which is essentially the expected value conditional on the size of the first neighbourhood, rather then just the expected value of $\gamma_r(u)$ we can obtain good concentration with a small exceptional probability. The price we pay for this is that the value $\gamma_1(u)(np)^{r-1}$, which the random variable $\gamma_r(u)$ is concentrated around, is itself random. We wish to apply this to estimate the minimum $r$-neighbourhood $\delta_r(\mathcal{G})$ and so concentration around $\gamma_1(u)(np)^{r-1}$ is more useful than concentration around $\mathbb{E}[\gamma_r(u)]$ as $\delta_r(\mathcal{G})$ is by definition as far as possible from $\mathbb{E}[\gamma_r(u)]$.

The proof of Theorem 7.1.1 below is similar to those of Lemma 4.1.2 and 4.1.3 and the idea of establishing a recurrence relation for the coefficients $a_i, b_i$ is inspired by [25].

**Theorem 7.1.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, $u \in V$ and $r := r(n) \geq 1$ be such that $(np)^{r+1} = o(n)$. Let $\lambda^* := \sqrt{\min\{10\gamma_1(u)\log(np), 2\log(n)\}}$ and define the event*

$$\mathcal{E}_{u,r} := \bigcap_{i=1}^{r} \left\{ \left|\gamma_i(u) - \gamma_1(u)(np)^{i-1}\right| \leq \lambda^*(np)^{i-1}\sqrt{\frac{\gamma_1(u)}{np}} \right\}.$$

*For any $c > 0$ if $np \geq c\log n$ then $\mathbb{P}((\mathcal{E}_{u,r})^c) = o\left(\frac{1}{n}\right) + o(e^{-np})$.*

Notice in the statement above that $\lambda * = o(np)$, thus for each $\gamma_i(u)$ we have concentration in an interval smaller than the value of $\gamma_i(u)$. Also the exceptional probability in the statement above is $o(1/n)$ when $np \geq \log(n)$ so a union bound can be taken and the statement holds for all vertices.

*Proof of Theorem 7.1.1.* To begin we will introduce the filtration $\widetilde{\mathfrak{F}}_r$ given by

$$\widetilde{\mathfrak{F}}_r := \widetilde{\mathfrak{F}}(\mathcal{G}, u, r) = \sigma\left(\Gamma_i(u) : 0 \leq i \leq r\right).$$

This is the filtration generated by the vertices at distance at most $r$ from $u$. Let

$$n_r := n - B_r(u), \quad p_r := \left(1 - (1-p)^{\gamma_r(u)}\right)/\gamma_r(u), \quad t_r := \prod_{i=1}^{r} n_i p_i,$$

for $r \geq 0$ and $t_0 = 1$. Observe that $\gamma_{r+1} \sim_d Bin(n_r, \gamma_r(u)p_r)$ conditional on $\widetilde{\mathfrak{F}}_r$ by Lemma 4.1.1. Also note that $n_r, p_r, t_r \in \widetilde{\mathfrak{F}}_r$. Let $i \geq 2$ and define $\mathcal{E}_u^i$ to be the event

$$\mathcal{E}_u^i := \{a_i t_{i-1} \leq \gamma_i(u) \leq b_i t_{i-1}\}, \tag{7.1}$$

where $a_i, b_i$ are given by the following recurrences for $i \geq 1$,

$$\begin{aligned} a_{i+1} &= a_i - \lambda\sqrt{a_i/t_i}, \\ b_{i+1} &= b_i + \lambda\sqrt{b_i/t_i}, \end{aligned} \tag{7.2}$$

with initial value $a_1 = b_1 = \gamma_1(u)$ and $\lambda = \sqrt{\min\{9\gamma_1(u)\log(np), 3\log(n)/2\}}$.

Let $\mathcal{H}_r$ be the event $\bigcap_{i=1}^r \left\{\gamma_i(u) \leq 2C^2(np)^i\right\}$ for some $C > 3$ to be chosen later. Notice that by the Bernoulli inequality we have

$$\gamma_i(u)p(1 - \gamma_i(u)p/2) \leq 1 - (1-p)^{\gamma_i(u)} \leq \gamma_i(u)p.$$

Thus conditional on $\mathcal{H}_r$ we have the following for any $k \leq r$

$$\begin{aligned} p \geq p_k &\geq p(1 - C^2 n^k p^{k+1}) \\ n \geq n_k &\geq n - (2C^2 + o(1))(np)^k \\ (np)^k \geq t_k &\geq \prod_{i=1}^k np\left(1 - (2C^2 + o(1))n^{i-1}p^i\right)\left(1 - C^2 n^i p^{i+1}\right) \\ &= (np)^k \left(1 - (C^2 + o(1))n^k p^{k+1}\right). \end{aligned} \tag{7.3}$$

We wish to show by induction that conditional on $\mathcal{H}_r$ the following holds for all $2 \leq i \leq r$

$$\begin{aligned} a_i &\geq \gamma_1(u) - \lambda\sqrt{\gamma_1(u)/np}(1 + o(1)), \\ b_i &\leq \gamma_1(u) + \lambda\sqrt{\gamma_1(u)/np}(1 + o(1)). \end{aligned} \tag{7.4}$$

Recall $\lambda = \sqrt{\min\{9\gamma_1(u)\log(np), 3\log(n)/2\}}$ and $a_1 = \gamma_1(u)$. We will establish (7.4) for $a_i$, the proof for $b_i$ is identical. By the recurrence relation (7.2) for $a_i$ we have

$$a_{k+1} = a_1 - \sum_{i=1}^k \lambda\sqrt{\frac{a_i}{t_i}} = \gamma_1(u) - \lambda\sqrt{\frac{\gamma_1(u)}{t_1}} - \sum_{i=2}^k \lambda\sqrt{\frac{a_i}{t_i}}.$$

Assuming the inductive hypothesis that $a_i$ satisfies (7.4) for all $i \leq k$ yields

$$a_{k+1} \geq \gamma_1(u) - \lambda\sqrt{\frac{\gamma_1(u)}{t_1}} - \sum_{i=2}^{k} \lambda\sqrt{\frac{\gamma_1(u) - \lambda\sqrt{\gamma_1(u)/np}(1+o(1))}{t_i}}.$$

Recall that $(np)^{r+1} = o(n)$. Since $t_k = (np)^k \left(1 - O\big(n^k p^{k+1}\big)\right)$ conditional on $\mathcal{H}_r$ we have

$$\begin{aligned} a_{k+1} &= \gamma_1(u) - \lambda\sqrt{\frac{\gamma_1(u)}{np(1-o(1))}} - \sum_{i=2}^{k} \lambda\sqrt{\frac{\gamma_1(u)}{(np)^i}}(1+o(1)) \\ &\geq \gamma_1(u) - \lambda\sqrt{\frac{\gamma_1(u)}{np}}\,(1+o(1)). \end{aligned}$$

Recall the event $\mathcal{E}_u^i := \{a_i t_{i-1} \leq \gamma_i(u) \leq b_i t_{i-1}\}$ for $i \geq 2$ from (7.1). We extend $\mathcal{E}_u^i$ to $i = 1$ by defining $\mathcal{E}_u^i = \Omega$. Then we have $\mathbb{P}\Big(\big(\mathcal{E}_u^2\big)^c \,\Big|\, \widetilde{\mathfrak{F}}_1 \cap \mathcal{E}_u^1\Big) = \mathbb{P}\Big(\big(\mathcal{E}_u^2\big)^c \,\Big|\, \widetilde{\mathfrak{F}}_1\Big)$. Let $\mathcal{D}_u^r$ be the event $\bigcap_{i=1}^{r} \mathcal{E}_u^i$ and observe that for $i \geq 1$ the following holds

$$\begin{aligned} \mathbb{P}\Big(\big(\mathcal{E}_u^{i+1}\big)^c \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big) &= \mathbb{P}\Big(Bin\left(n_i, \gamma_i(u) p_i\right) < a_{i+1} t_i \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big) \\ &\quad + \mathbb{P}\Big(Bin\left(n_i, \gamma_i(u) p_i\right) > b_{i+1} t_i \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big). \end{aligned}$$

Now by the recurrence relations for $a_i$ and $b_i$, given by (7.2), we have

$$\begin{aligned} \mathbb{P}\Big(\big(\mathcal{E}_u^{i+1}\big)^c \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big) &= \mathbb{P}\Big(Bin\left(n_i, \gamma_i(u) p_i\right) < a_i t_i - \lambda\sqrt{a_i t_i} \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big) \\ &\quad + \mathbb{P}\Big(Bin\left(n_i, \gamma_i(u) p_i\right) > b_i t_i + \lambda\sqrt{b_i t_i} \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big). \end{aligned}$$

An application of the Chernoff bounds, Lemma 2.1.1, yields

$$\mathbb{P}\Big(\big(\mathcal{E}_u^{i+1}\big)^c \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big) \leq \exp\left(-\frac{\lambda^2}{2}\right) + \exp\left(-\frac{\lambda^2}{2 + \lambda/3\sqrt{b_i t_i}}\right).$$

Recall that conditional on $\mathcal{H}_r$ we have $t_i \sim (np)^i$ by (7.3) and $b_i \sim \gamma_1$ by (7.4), thus

$$\frac{\lambda}{3\sqrt{b_i t_i}} = \sqrt{\frac{\min\left\{9\gamma_1(u)\log(np), 3\log(n)/2\right\}}{9\gamma_1(u)(np)^i\left(1-o(1)\right)}} = O\left(\sqrt{\frac{\log(np)}{(np)^i}}\right).$$

This gives us the following upper bound

$$\mathbb{P}\Big(\big(\mathcal{E}_u^{i+1}\big)^c \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big)\mathbf{1}_{\mathcal{H}_r} \leq \exp\left(-\frac{\lambda^2}{3}\right)\mathbf{1}_{\mathcal{H}_r} = \max\left\{\frac{1}{(np)^{3\gamma_1(u)}}, \frac{1}{n^{3/2}}\right\}.$$

Recall that $\gamma_1(u) \in \widetilde{\mathfrak{F}}_1$ and so $\exp\left(-\lambda^2/3\right) \in \widetilde{\mathfrak{F}}_1$. Observe that

$$\begin{aligned} 1 &= \left(\mathbf{1}_{\{\gamma_1(u)=0\}} + \mathbf{1}_{\{\gamma_1(u)>0\}}\right)\left(\mathbf{1}_{\mathcal{H}_r} + \mathbf{1}_{(\mathcal{H}_r)^c}\right) \\ &\leq \mathbf{1}_{\{\gamma_1(u)=0\}} + \mathbf{1}_{\{\gamma_1(u)>0\}}\mathbf{1}_{\mathcal{H}_r} + \mathbf{1}_{(\mathcal{H}_r)^c}. \end{aligned}$$

Also notice that $\mathbb{P}\Big(\left(\mathcal{E}_u^{i+1}\right)^c \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big)\,\mathbf{1}_{\{\gamma_1(u)=0\}} = 0$ because $\gamma_1(u) = 0$ implies that $\gamma_i(u) = 0$. Combining these observations yields

$$\begin{aligned} \mathbb{P}\Big(\left(\mathcal{E}_u^{i+1}\right)^c \,\Big|\, \mathcal{D}_u^i\Big) &= \mathbb{E}\Big[\mathbb{P}\Big(\left(\mathcal{E}_u^{i+1}\right)^c \,\Big|\, \widetilde{\mathfrak{F}}_i \cap \mathcal{D}_u^i\Big)\Big] \\ &\leq 0 + \mathbb{E}\big[\exp\left(-\lambda^2/3\right)\mathbf{1}_{\{\gamma_1(u)>0\}}\big] + \mathbb{P}((\mathcal{H}_r)^c)\,. \end{aligned} \tag{7.5}$$

If $9\gamma_1(u)\log(np) > 3\log(n)/2$ then $\exp\left(-\lambda^2/3\right) \leq 1/n^{3/2}$. Thus assume that $9\gamma_1(u)\log(np) \leq 3\log(n)/2$ and we have the following for $\mathbb{E}\big[\exp\left(-\lambda^2/3\right)\mathbf{1}_{\{\gamma_1(u)>0\}}\big]$

$$\begin{aligned} \mathbb{E}\left[\frac{\mathbf{1}_{\{\gamma_1(u)>0\}}}{(np)^{3\gamma_1(u)}}\right] &= \sum_{k=1}^{n-1}\frac{1}{(np)^{3k}}\mathbb{P}(\gamma_1(u) = k) \\ &\leq \sum_{k=1}^{np}\frac{1}{(np)^{3k}}\cdot\binom{n-1}{k}p^k(1-p)^{n-1-k} + \frac{n}{(np)^{np}}. \end{aligned}$$

Applying the bound $\binom{n}{k} \leq n^k/k!$ and Bernoulli's inequality (2.3) we have

$$\mathbb{E}\left[\frac{1}{(np)^{3\gamma_1(u)}}\right] \leq (1-p)^n\sum_{k=1}^{np}\frac{(np)^k}{(np)^{3k}k!(1-(k+1)p)} + \frac{n}{(np)^{np}} \leq \frac{2e^{-np}}{(np)^2}.$$

Recall that $\mathbb{P}((\mathcal{H}_r)^c) = o\left(e^{-3(C-3)np/2}\right)$, by Lemma 4.1.2 where $C > 3$ was arbitrary. Thus we have the following by (7.5)

$$\begin{aligned} \mathbb{P}\Big(\mathcal{E}_u^{i+1} \,\Big|\, \mathcal{D}_u^i\Big) &\geq 1 - \frac{2e^{-np}}{(np)^2} - \frac{1}{n^{3/2}} - o\left(e^{-3(C-3)np/2}\right) \\ &\geq 1 - \frac{3e^{-np}}{(np)^2} - \frac{2}{n^{3/2}}. \end{aligned} \tag{7.6}$$

Recall that $\mathcal{D}_u^r := \bigcap_{i=1}^{r}\mathcal{E}_u^i$. Since $\bigcap_{i=1}^{r}\mathcal{E}_u^i \subseteq \bigcap_{i=1}^{r-1}\mathcal{E}_u^i$ the event $\mathcal{D}_u^{r-1}$ is the disjoint union of the events $\mathcal{D}_u^r$ and $(\mathcal{E}_u^r)^c \cap \mathcal{D}_u^{r-1}$. Thus we have the following

$$\mathbb{P}(\mathcal{D}_u^r) = \mathbb{P}\big(\mathcal{D}_u^{r-1}\big)\left(1 - \mathbb{P}\Big((\mathcal{E}_u^r)^c \,\Big|\, \mathcal{D}_u^{r-1}\Big)\right).$$

By applying this decomposition repeatedly we obtain

$$\mathbb{P}(\mathcal{D}_u^r) = \mathbb{P}\big(\mathcal{E}_u^1\big) \prod_{j=2}^{r} \Big(1 - \mathbb{P}\Big(\big(\mathcal{E}_u^j\big)^c \,\Big|\, \mathcal{D}_u^{j-1}\Big)\Big) .$$

The upper bound on $\mathbb{P}\Big(\big(\mathcal{E}_u^r\big)^c \,\Big|\, \mathcal{D}_u^{r-1}\Big)$ from (7.6) holds for all $j \leq r$, hence

$$\mathbb{P}(\mathcal{D}_u^r) \geq \mathbb{P}\big(\mathcal{E}_u^1\big) \left(1 - \frac{3e^{-np}}{(np)^2} - \frac{2}{n^{3/2}}\right)^r .$$

Recall that $\mathbb{P}\big(\mathcal{E}_u^1\big) = 1$, thus by the Bernoulli inequality (2.3) we have

$$\mathbb{P}(\mathcal{D}_u^r) \geq \left(1 - r\left(\frac{3e^{-np}}{(np)^2} + \frac{2}{n^{3/2}}\right)\right) = 1 - O\left(\frac{e^{-np}}{np}\right) - O\left(\frac{1}{n^{3/2}}\right), \tag{7.7}$$

where the final equality holds since $r \leq \log(n)/\log(np) \leq np$. Consider the event

$$\mathcal{R} := \bigcap_{i}^{r} \left\{ \left|\gamma_i(u) - \gamma_1(u)(np)^{i-1}\right| \leq (np)^{i-1} \cdot \lambda^* \sqrt{\frac{\gamma_1(u)}{np}} \right\},$$

from the statement of the theorem. Observe that conditional on the event $\mathcal{H}_1$

$$\frac{100\lambda}{99}\sqrt{\frac{\gamma_1(u)}{np}} + \gamma_1(u)2C^2p \leq \lambda^* \sqrt{\frac{\gamma_1(u)}{np}},$$

where we recall $\mathcal{H}_1 = \big\{\gamma_1(u) \leq C^2 np\big\}$. Hence by the triangle inequality

$$\begin{aligned}\mathcal{R} \cap \mathcal{H}_1 \supseteq &\left( \bigcap_{i}^{r} \left\{ |\gamma_i(u) - \gamma_1(u)t_{i-1}| \leq (np)^{i-1} \cdot \frac{100\lambda}{99}\sqrt{\frac{\gamma_1(u)}{np}} \right\} \right) \\ &\bigcap \left( \bigcap_{i}^{r} \left\{ \left|\gamma_1(u)t_{i-1} - \gamma_1(u)(np)^{i-1}\right| \leq \gamma_1(u) \cdot 2C^2 n^{i-1}p^i \right\} \right).\end{aligned}$$

The first bracketed event in the intersection above is a subset of $\mathcal{D}_u^r$ by (7.1) and (7.4) and second is a subset of $\mathcal{H}_u^r = \bigcap_{i=1}^r \big\{\gamma_i(u) \leq 2C^2(np)^i\big\}$ by (7.3). Thus by (7.7) and Lemma 4.1.2 we have

$$\mathcal{R}^c \leq \mathbb{P}((\mathcal{D}_u^r)^c) + \mathbb{P}((\mathcal{H}_r)^c) = O\left(\frac{e^{-np}}{np}\right) + O\left(\frac{1}{n^{3/2}}\right) + o\left(e^{-3(C-3)np/2}\right).$$

The result follows since $C > 3$ was arbitrary. $\square$

## 7.2 Second moment estimates for sets of vertices with given degree

As mentioned before we do not have so much control over the joint distribution of the $r$-neighbourhoods in $\mathcal{G}(n,p)$. This is an issue if we wish to try to calculate the number of vertices which attain a minimum size $r$-neighbourhood using the second moment method. The joint distributions are very complicated to calculate even for second neighbourhoods so instead we control the second moments of the sizes of sets of vertices with first neighbourhoods of a certain size.

The following lemma is essentially a more detailed version of the second moment estimates used to estimate the number of vertices of a certain degree by Bollobás [15]. The proof is based around standard calculations found in [20, 43]. We state and prove it here as I cannot find it stated or proved explicitly in the literature for sets of vertices with degrees taking values in a set $R$ as opposed to just a single value.

**Lemma 7.2.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, $R \subseteq [0,1,\ldots,n-1]$ and let $r^* := \max\{r : r \in R\}$. Then* $\mathrm{Var}\left[\left|\bigcup_{i\in R} X_{1,i}\right|\right]$ *is bounded from above by*

$$\mathbb{E}\left[\left|\bigcup\nolimits_{i\in R} X_{1,i}\right|\right]\left(1+O\left(\frac{r^*(r^*+np)+(np)^2}{np}\mathbb{P}(\gamma_1(v)\in R)\right)\right).$$

*Moreover, if $k = O(np)$, $p = o(1)$ and $\mathbb{E}[|X_{1,k}|] \to \infty$, then*

$$\mathbb{P}(|X_{1,k}| > 0) = 1 - o(1).$$

*Proof.* Notice that $\mathbb{P}\big(uv \in E \big| \gamma_1(v) = r\big) = r/(n-1)$ since each edge is independent of all other edges. Thus letting $r^* = \max_{r\in R} r$ we have $\mathbb{P}\big(uv \in E\big|\gamma_1(v) \in R\big) \le r^*/(n-1)$. Therefore we have the following by the law of total probability

$$\begin{aligned}
&\mathbb{P}\big(\gamma_1(u) = i\big|\gamma_1(v) \in R\big)\\
&\quad\le \binom{n-2}{i-1}p^{i-1}(1-p)^{n-1-i}\frac{r^*}{n-1} + \binom{n-2}{i}p^i(1-p)^{n-2-i}\\
&\quad= \binom{n-1}{i}p^i(1-p)^{n-1-i}\left(\frac{ir^*}{(n-1)^2p} + \frac{n-1-i}{(n-1)(1-p)}\right).
\end{aligned}$$

Then since $i \le r^*$ and $\mathbb{P}(\gamma_1(u) = i) = \binom{n-1}{i}p^i(1-p)^{n-1-i}$ we have the following

$$\mathbb{P}\big(\gamma_1(u) = i\big|\gamma_1(v) \in R\big) \le \mathbb{P}(\gamma_1(u) = i)\left(1 + O\left(\frac{r^*(r^*+np)+(np)^2}{n^2p}\right)\right). \tag{7.8}$$

We can then apply the bound in line (7.8) above to obtain

$$\begin{aligned}
&\mathbb{P}(\gamma_1(u) \in R, \gamma_1(v) \in R) \\
&\quad = \sum_{i \in R} \mathbb{P}\big(\gamma_1(u) = i \big| \gamma_1(v) \in R\big)\, \mathbb{P}(\gamma_1(v) \in R) \\
&\quad \leq \sum_{i \in R} \mathbb{P}(\gamma_1(u) = i)\, \mathbb{P}(\gamma_1(v) \in R) \left(1 + O\left(\frac{r^*\,(r^* + np) + (np)^2}{n^2 p}\right)\right) \\
&\quad = \mathbb{P}(\gamma_1(u) \in R)\, \mathbb{P}(\gamma_1(v) \in R) \left(1 + O\left(\frac{r^*\,(r^* + np) + (np)^2}{n^2 p}\right)\right).
\end{aligned}$$

Let $X := \bigcup_{i \in R} X_{1,i}$. We have the following expression for $\mathrm{Var}\,[|X|]$

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum\nolimits_{u \in V} \mathbf{1}_{\gamma_1(u) \in R} - \sum\nolimits_{u \in V} \mathbb{P}(\gamma_1(u) \in R)\right)^2\right] \\
&\quad = \sum_{u \in V} \sum_{v \in V} \left(\mathbb{P}(\gamma_1(u) \in R, \gamma_1(v) \in R) - \mathbb{P}(\gamma_1(u) \in R)\, \mathbb{P}(\gamma_1(v) \in R)\right) \\
&\quad \leq \mathbb{E}[|X|] + \sum_{\substack{u,v \in V,\\ u \neq v}} O\left(\frac{r^*\,(r^* + np) + (np)^2}{n^2 p}\right) \mathbb{P}(\gamma_1(u) \in R)\, \mathbb{P}(\gamma_1(v) \in R) \\
&\quad \leq \mathbb{E}[|X|] + \mathbb{E}[|X|] \sum_{v \in V} O\left(\frac{r^*\,(r^* + np) + (np)^2}{n^2 p}\right) \mathbb{P}(\gamma_1(v) \in R) \\
&\quad \leq \mathbb{E}[|X|] \left(1 + O\left(\frac{r^*\,(r^* + np) + (np)^2}{np} \mathbb{P}(\gamma_1(v) \in R)\right)\right).
\end{aligned}$$

For the second part if $R = \{k\}$ where $k = O(np)$ then $\max\{(r^*)^2/np, np\} = O(np)$. Therefore by the main statement of the theorem we have

$$\mathrm{Var}\,[|X_{1,k}|] \leq \mathbb{E}[|X_{1,k}|]\,(1 + O(p\mathbb{E}[|X_{1,k}|]))\,.$$

Thus if $\mathbb{E}[|X_{1,k}|] \to \infty$ then by the second moment method [6, Theorem 4.3.1]:

$$\mathbb{P}(|X_{1,k}| = 0) \leq \frac{\mathrm{Var}\,[|X_{1,k}|]}{\mathbb{E}[|X_{1,k}|]^2} \leq \frac{1}{\mathbb{E}[|X_{1,k}|]} + O(p) = o\,(1)\,.$$

□

Notice $|X_{1,k}| \leq n$ so this estimate is at least as good as the estimate in [43].

## 7.3 A bound on the number of vertices with a minimum $r$-neighbourhood

The following proposition seems to be a well known fact. This proposition is needed to prove Lemma 7.3.2 so I prove it for completeness as I cannot find it stated explicitly in the literature, though it is of note that [20, Theorem 7.2] is a similar result for the related model $\mathcal{G}(n,m)$. The proof is based on the proof of connectivity threshold in the book by Freeze & Karoński [43, Theorem 4.1]. Notice that if $p$ is below the threshold $2np = \log(n)$ then isolated paths of length one start to appear.

**Proposition 7.3.1** (Folklore)**.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ with $2np = \log(n) + \log\log(n) + w(n)$ where $w(n) \to \infty$. Then with high probability $\mathcal{G}$ consists of a giant component and isolated vertices.*

*Proof.* It suffices to prove the Proposition when $\frac{\log(n)+\log\log(n)+w(n)}{2n} \leq p \leq \frac{2\log(n)}{n}$, because by Theorem 2.3.1 above this range we are connected w.h.p.

Let $\mathcal{T}$ be the event that $\mathcal{G}$ consists of a giant component and isolated vertices. Let $C_k := C_{k,n}$ be the number of components with $k$ vertices in $\mathcal{G}(n,p)$.

$$\mathbb{P}(\mathcal{T}^c) = \mathbb{P}\left(\bigcup_{k=2}^{\lfloor n/2 \rfloor} \{C_k > 0\}\right) \leq \sum_{k=2}^{\lfloor n/2 \rfloor} \mathbb{P}(C_k > 0) \leq \sum_{k=2}^{\lfloor n/2 \rfloor} \mathbb{E}[C_k]\,,$$

by the union bound and Markov's inequality. Now observe that

$$\mathbb{E}[C_k] \leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)},$$

as there are $\binom{n}{k}$ ways of choosing the vertices and the component must be connected, contributing $k^{k-2}p^{k-1}$, but not connected to anything else in the graph, which happens with probability $(1-p)^{k(n-k)}$. Thus if $k \leq 10$:

$$\mathbb{E}[C_k] \leq \left(\frac{npe}{k}\right)^k \frac{k^{k-2} e^{-kp(n-10)}}{p} \leq en\,(npe)^{k-1}\, e^{-knp+o(1)},$$

and so we see that the bound on $\mathbb{E}[C_k]$ achieves its maximum at $k = 2$ in which case we have the following since $2np \geq \log(n) + \log\log(n) + w(n)$:

$$\begin{aligned}\mathbb{E}[C_2] &\leq ne\,(npe)\, e^{-2np+o(1)} \\ &\leq ne^2 \log(n) e^{-\log(n)-\log\log(n)-w(n)+o(1)} \\ &= O\Big(e^{-w(n)}\Big)\,.\end{aligned}$$

Now if $10 \leq k \leq \lfloor n/2 \rfloor$ then we have

$$\mathbb{E}[C_k] \leq \left(\frac{npe}{k}\right)^k k^{k-2} p^{k-1} e^{-knp/2} \leq en \frac{(npe)^{k-1}}{k^2} e^{-knp/2} \leq n \left(\frac{2e\log(n)}{n^{1/4}}\right)^{k-1}.$$

Thus we have

$$\mathbb{P}(\mathcal{T}^c) \leq O\Big(e^{-w(n)}\Big) + \sum_{k=10}^{\lfloor n/2 \rfloor} n^{1-k/4+o(1)} = o\left(1\right).$$

□

A key observation is that vertices of degree zero have $r$-degree zero. The following lemma shows that, provided that $r$ is not too large, w.h.p. these are the only vertices of $r$-degree zero.

**Lemma 7.3.2.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where $np = \frac{\log(n)+\log\log(n)+w(n)}{2}$ and $w(n) \to \infty$. Let $r := r(n) \geq 2$ be such that $(np)^r = o\left(n\right)$. Then*

$$\mathbb{P}(X_{r,0} = X_{1,0}) = 1 - o(1).$$

*Proof.* To begin if $\Gamma_1(u)$ is empty then $\Gamma_r(u)$ is empty for all $r \geq 1$, thus $X_{1,0} \subseteq X_{r,0}$. Notice that $\{X_{r,0} \not\subseteq X_{1,0}\} = \cup_{u \in V}\{\gamma_1(u) \geq 1\} \cap \{\gamma_r(u) = 0\}$ and denote this event $\mathcal{E}$. We decompose the event $\mathcal{E}$ into two disjoint events $\mathcal{C}_1(u) := \mathcal{E} \cap \{B_r(u) = V\}$ and $\mathcal{C}_2(u) := \mathcal{E} \cap \{B_r(u) \neq V\}$. By Theorem 4.1.2 we have

$$\mathbb{P}(|B_r(u)| \leq 33(np)^r, \text{ for all } u \in V) \geq 1 - n \cdot e^{-3np/2} = 1 - o(1). \tag{7.9}$$

If $\mathcal{C}_1(u)$ holds then $B_i(u) = V$ for some $1 \leq i < r$, thus $|B_r(u)| = n$. By the assumption $(np)^r = o\left(n\right)$ we have $B_r(u) = o\left(n\right)$ w.h.p. for every $u \in V$ from (7.9). This implies that $\mathbb{P}(\cup_{u \in V}\mathcal{C}_1(u)) = o(1)$.

By Proposition 7.3.1 with high probability every vertex is either isolated or contained in the giant component. We know that on the event $\mathcal{C}_2(u)$ the ball $B_i(u)$ contains at least two vertices and so w.h.p every vertex outside the ball must be isolated. However by (7.9) we know that w.h.p. $|B_i(u)| = o\left(n\right)$, for every $u \in V$, and so w.h.p. the graph must have at least $n/2$ isolated vertices. The probability there are $n/2$ or more isolated vertices is $o(1)$. This is by the second moment method since $\mathbb{E}[|X_{1,0}|] = \lambda_0(n) = e^{\log(n)-np+O(np^2)} = o(\sqrt{n})$ by (7.13) and by Lemma 7.2.1 $\mathrm{Var}(|X_{1,0}|) \leq \mathbb{E}[|X_{1,0}|]\left(1 + o(1)\right)$. This implies that $\mathbb{P}(\cup_{u \in V}\mathcal{C}_2(u)) = o(1)$. Thus $\mathbb{P}(\mathcal{E}) = \mathbb{P}(\cup_{u \in V}\left(\mathcal{C}_1(u) \cup \mathcal{C}_2(u)\right)) = o(1)$ and so $\mathbb{P}(X_{r,0} = X_{1,0}) = 1 - o(1)$. □

The next lemma allows us to control the number of vertices of minimum $r$-degree by showing that w.h.p they must also have a small degree. This is of use because sets of vertices defined by the size of their first neighbourhoods are easier than those defined by the sizes of their $r$-neighbourhoods. In particular we have good control of the joint distribution of sets of vertices with given degree by Lemma 7.2.1.

**Lemma 7.3.3.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where $\liminf_{n\to\infty} np - \log(n) > -\infty$. Let $r := r(n) \geq 2$ be such that $(np)^{r+1} = o(n)$. Let*

$$h = \left\lfloor 3\sqrt{\frac{\min\{10(\delta_1)^2 \log(np), 2\delta_1 \log(n)\}}{np}} \right\rfloor,$$

*and let $X := \bigcup_{i=0}^{h} X_{1,\delta_1+i}$. Then $\mathbb{P}(X_{r,\delta_r} \not\subseteq X) = o(1)$.*

*Proof.* Let $\lambda^* = \sqrt{\min\{10\gamma_1(u)\log(np), 2\log(n)\}}$. For $r \geq 2$ define the event

$$\mathcal{E}_r := \left\{ \left|\gamma_r(u) - \gamma_1(u)(np)^{r-1}\right| \leq \lambda^*(np)^{r-1}\sqrt{\frac{\gamma_1(u)}{np}}, \text{ for all } u \in V \right\}.$$

Let $h = \lfloor 3\lambda^*\sqrt{\delta_1/np} \rfloor$ and $X := \bigcup_{i=0}^{h} X_{1,\delta_1+i}$ be as in the statement of the lemma. If $\{\delta_1 = 0\}$ then by Lemma 7.3.2 we have $X_{r,\delta_r} = X_{r,0} = X_{1,0} = X$ with high probability. Thus we now assume that $\delta_1 \geq 1$. For any $x \in X_{1,\delta_1}$ and $y \in V$ such that $\gamma_1(y) > \delta_1 + h$ we have the following conditional on $\mathcal{E}_r \cap \{\delta_1 \geq 1\}$

$$\begin{aligned}
\gamma_r(y) &\geq \delta_1(np)^{r-1} + \lfloor 3\lambda^*\sqrt{\delta_1/np}\rfloor(np)^{r-1} \\
&\qquad - \lambda^*(np)^{r-1}\sqrt{\left(\delta_1 + \lfloor 3\lambda^*\sqrt{\delta_1/np}\rfloor\right)/np} \\
&> \delta_1(np)^{r-1} + \lambda^*(np)^{r-1}\sqrt{\delta_1/np} \\
&\geq \gamma_r(x),
\end{aligned}$$

provided $n$ is large. Thus conditional on $\mathcal{E}_r \cap \{\delta_1 \geq 1\}$ we have $X_{r,\delta_r} \subseteq X$. Hence collecting these bounds together we have the following

$$\mathbb{P}(X_{r,\delta_r} \not\subseteq X) \leq \mathbb{P}(X_{r,0} \not\subseteq X_{1,0}) + \mathbb{P}\left(X_{r,\delta_r} \not\subseteq X \middle| \mathcal{E}_r \cap \{\delta_1 \geq 1\}\right) + \mathbb{P}(\mathcal{E}_r^c) = o(1),$$

where $\mathbb{P}(\mathcal{E}_r^c) \leq n(o(1/n) + o(e^{-np})) = o(1)$ by Lemma 7.1.1. □

We are now ready to prove an upper bound on the number of vertices attaining an $r$-neighbourhood of minimum size.

**Theorem 7.3.4.** *Let* $\mathcal{G} \sim_d \mathcal{G}(n,p)$ *where* $\liminf_{n\to\infty} np - \log(n) > -\infty$. *Let* $r := r(n) \geq 2$ *be such that* $(np)^{r+1} = o(n)$. *Then with high probability*

$$|X_{r,\delta_r}| = e^{O\left(\sqrt{\log(n)}\right)}.$$

*Proof.* Recall $h$ and $X$ from the statement of Lemma 7.3.3. We wish to bound $\mathbb{E}[|X|]$ from above. By linearity of expectation we have

$$\mathbb{E}\Big[|X|\Big|\delta_1\Big] = n\sum_{i=0}^{h}\binom{n-1}{\delta_1+i}p^{\delta_1+i}(1-p)^{n-1-\delta_1-i}.$$

Now let $j$ be any $i \in [0,h]$ which maximises the function $\binom{n-1}{\delta_1+i}p^{\delta_1+i}(1-p)^{n-1-\delta_1-i}$. Thus

$$\begin{aligned}\mathbb{E}\Big[|X|\Big|\delta_1\Big] &\leq nh\binom{n-1}{\delta_1+j}p^{\delta_1+j}(1-p)^{n-1-\delta_1-j}\\ &\leq nh\frac{(n-\delta_1-1)\cdots(n-\delta_1-j)\cdot p^j}{(\delta_1+1)\cdots(\delta_1+j)\cdot(1-p)^j}\binom{n-1}{\delta_1}p^{\delta_1}(1-p)^{n-1-\delta_1}.\end{aligned}$$

Then since $(1-p)^j \geq (1-p)^h \geq 1-hp = 1-o(1)$ and also $h = o(\delta_1)$ we have the following by the Bernoulli inequality (2.3)

$$\mathbb{E}\Big[|X|\Big|\delta_1\Big] \leq 2h\left(\frac{np}{\delta_1}\right)^j\mathbb{E}\Big[|X_{1,\delta_1}|\Big|\delta_1\Big] \leq 2h\left(\frac{np}{\delta_1}\right)^h\mathbb{E}\Big[|X_{1,\delta_1}|\Big|\delta_1\Big].$$

Recall $h = \lfloor 3\lambda^*\sqrt{\delta_1/np}\rfloor$, where $\lambda^* = \sqrt{\min\{10\delta_1(u)\log(np), 2\log(n)\}}$. Assume we are in the case where $10\delta_1\log(np) < 2\log(n)$ and consider the function $ax\,(b/x)^{ax}$. The first derivative of $ax\,(b/x)^{ax}$ is

$$\frac{d}{\mathrm{d}x}\left(ax\left(\frac{b}{x}\right)^{ax}\right) = a\left(\frac{b}{x}\right)^{ax}\left(ax\log\left(\frac{b}{x}\right) - ax + 1\right). \tag{7.10}$$

This is positive for small enough $x$. Let $a = 3\sqrt{10\log(np)/np}$ and $b = np$, then $\delta_1 a\,(b/\delta_1)^{\delta_1 a}$ is the function $h\left(\frac{np}{\delta_1}\right)^h$ up to constants. Examination of (7.10) reveals that $h\left(\frac{np}{\delta_1}\right)^h$ is increasing in $\delta$ whenever $10\delta_1\log(np) < 2\log(n)$. Thus in this case

$$2h\left(\frac{np}{\delta_1}\right)^h \leq \frac{6\sqrt{10}\log(n)}{5\sqrt{\log(np)np}}\left(\frac{5\log(np)np}{\log(n)}\right)^{\frac{3\sqrt{10}\log(n)}{5\sqrt{\log(np)np}}} = e^{O\left(\sqrt{\log(n)}\right)}.$$

For the second case where $10\delta_1 \log(np) \geq 2\log(n)$ we will consider the function $a\sqrt{x}\,(b/x)^{a\sqrt{x}}$ with first derivative

$$\frac{d}{\mathrm{d}x}\left(a\sqrt{x}\left(\frac{b}{x}\right)^{a\sqrt{x}}\right) = a\left(\frac{b}{x}\right)^{a\sqrt{x}} \frac{\left(a\sqrt{x}\left(\log\left(\frac{b}{x}\right) - 2\right) + 1\right)}{2\sqrt{x}}.$$

Thus the maximiser is close to the point $x = b/e^2$, and if we take $a = 3\sqrt{2\log(n)/np}$ then we have the function $h\left(\frac{np}{\delta_1}\right)^h$ up to constants. Thus in the second case we have

$$2h\left(\frac{np}{\delta_1}\right)^h \leq \sqrt{np\log(n)}(e^2 - o(1))^{\sqrt{np/e^2}\cdot 3\sqrt{2\log(n)/np}} = e^{O\left(\sqrt{\log(n)}\right)}.$$

Thus we have

$$\mathbb{E}[|X|] \leq \mathbb{E}\left[e^{O\left(\sqrt{\log(n)}\right)}\mathbb{E}\left[|X_{1,\delta_1}|\,\middle|\delta_1\right]\right] = e^{O\left(\sqrt{\log(n)}\right)}\mathbb{E}[|X_{1,\delta_1}|]\,. \tag{7.11}$$

Consider the following for any $k \geq 0$

$$\begin{aligned}\mathbb{P}(\gamma_1(u) = k+1) &= \frac{(n-1)!}{k!(n-1-k)!}p^k(1-p)^{n-1-k}\left(\frac{p(n-1-k)}{(1-p)(k+1)}\right)\\ &= \frac{np}{k}(1 \pm o(1))\mathbb{P}(\gamma_1(u) = k)\,.\end{aligned}$$

Now by Lemma 7.2.1 we have that if $\mathbb{E}[|X_{1,k}|] \to \infty$ for some $k = O(np)$ then $\mathbb{P}(|X_{1,k}| > 0) = 1 - o(1)$. When $np \geq 10\log(n)$ we know that $\delta_1 \geq (n-1)p/2$ with high probability since by the Chernoff bound, Lemma 2.1.1 we have

$$\mathbb{P}(\delta_1 < (n-1)p/2) \leq n \cdot \exp\left(-\frac{(n-1)p}{4}\right) \leq o\left(1/n\right).$$

Thus for $np \geq 10\log(n)$ we have $np/\delta_1 \geq 1/2$. This means we can bound $\mathbb{E}[|X_{1,\delta_1}|]$ by some function going to infinity with $n$ fairly slowly, for concreteness shall take $\mathbb{E}[|X_{1,\delta_1}|] = O(\log(n))$, since otherwise we must have some $k < \delta_1$ such that $\mathbb{E}[|X_{1,k}|] = \omega(1)$ in which case $\delta_1 \leq k$ with high probability, a contradiction! When $np \leq 10\log(n)$ and $\delta_1 > 0$ by the same argument we have $\mathbb{E}[|X_{1,\delta_1}|] = O(np/\delta_1) = O(\log(n))$. This argument fails if $\delta_1 = 0$ and so our bound on $\mathbb{E}[|X_{1,\delta_1}|]$ must include the expected number of vertices of degree zero given by $\mathbb{E}[|X_{1,0}|] = n(1-p)^{n-1} \leq ne^{-np} = O(1)$.

Thus combining these observations with (7.11) and noticing that $\log(n) =$

$o\left(e^{\sqrt{\log(n)}}\right)$ yields the following bound

$$\mathbb{E}[|X|] \leq \left(1 + ne^{-np}\right) \cdot e^{O\left(\sqrt{\log(n)}\right)} = e^{O\left(\sqrt{\log(n)}\right)}.$$

Let $t := \max\{\mathbb{E}[|X|], \log(n)\}$ and observe that

$$\mathbb{P}(|X_{r,\delta_r}| \leq \mathbb{E}[|X|] + t) \leq \mathbb{P}(|X| \leq \mathbb{E}[|X|] + t) + \mathbb{P}(X_{r,\delta_r} \not\subseteq X).$$

Now since $\delta_1 + h = O(np)$ w.h.p. and $(np)^{r+1} = o(n)$ we have that $\mathrm{Var}\,|X| \leq \mathbb{E}[|X|](1 + O(p\mathbb{E}[|X|])) = \mathbb{E}[X](1 + o(1))$ by Lemma 7.2.1. Thus by the Cantelli second moment inequality (2.7) and Lemma 7.3.3 we have

$$\mathbb{P}(|X_{r,\delta_r}| \leq \mathbb{E}[|X|] + t) \leq \frac{\mathbb{E}[|X|](1 + o(1))}{\mathbb{E}[|X|](1 + o(1)) + t^2} + o(1) = o(1).$$

The result follows since $\mathbb{E}[|X|] + t = e^{O\left(\sqrt{\log(n)}\right)}$. $\square$

## 7.4 A conjecture regarding the uniqueness of vertices with a minimal $r$-neighbourhood

Let $\mathcal{U}_r$ be the event that there is only one vertex which attains the smallest $r$-neighbourhood. Thus, in keeping with our notation, we would denote

$$\mathcal{U}_r := \{|X_{r,\delta_r}| = 1\}.$$

We make the following conjecture.

**Conjecture 7.4.1.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$, let $r(n) \geq 2$ be such that $(np)^{2r} = o(n)$, let $w = w(n) = np - \log(n)$. Then the following assertions hold.*

*(i) If $\lim_{n\to\infty} w(n) = -\infty$, then $\lim_{n\to\infty} \mathbb{P}(\mathcal{U}_r) = 0$.*

*(ii) If $\lim_{n\to\infty} w(n) = \infty$, then $\lim_{n\to\infty} \mathbb{P}(\mathcal{U}_r) = 1$.*

*(iii) If $-\infty < \liminf_{n\to\infty} w(n) \leq \limsup_{n\to\infty} w(n) < \infty$, then*

$$\mathbb{P}(\mathcal{U}_r) \sim e^{-e^{-w}}(1 + e^{-w}).$$

If instead we consider the first neighbourhoods of vertices in $\mathcal{G}(n,p)$, this is the case $r = 1$, then we have a different situation which is outlined in the following theorem.

**Theorem 7.4.1** (Theorem 4 in [18])**.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ and $p \leq 1/2$. Then*

*(i) If* $\lim_{n\to\infty} \frac{np}{\log(n)} = \infty$, *then* $\lim_{n\to\infty} \mathbb{P}(\mathcal{U}_1) = 1$.

*(ii) If* $\lim_{n\to\infty} \mathbb{P}(\mathcal{U}_1) = 1$, *then* $\lim_{n\to\infty} \frac{np}{\log(n)} = \infty$.

Recall that the threshold for connectedness in $\mathcal{G}(n,p)$ is $np = \log(n)$ by Theorem 2.3.1. To summarise the theorem and the conjecture above we can say that w.h.p. there is a unique vertex with minimum first neighbourhood if and only if $np$ dominates $\log(n)$, whereas for $r \geq 2$ it is conjectured that w.h.p. there is a unique vertex attaining the smallest $r$-neighbourhood whenever we are above the connectedness threshold $np = \log(n)$.

To understand the threshold in Conjecture 7.4.1 consider the following observation: if a vertex has an empty first neighbourhood then it has an empty $r$-neighbourhood, for any $r \geq 1$. If we are well below the connectedness threshold there are many isolated vertices and so we do not have a unique vertex with smallest $r$-neighbourhood as there are many vertices with an empty first neighbourhood. Indeed we will prove Conjecture 7.4.1 Item (i) in a slightly stronger form.

**Proposition 7.4.2.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$. If* $\lim_{n\to\infty} np - \log(n) = -\infty$, *then*

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{r=1}^{\infty} \mathcal{U}_r\right) = 0.$$

*Proof.* Observe that if the first neighbourhood of a vertex is empty then its $r$-neighbourhoods are empty for all $r \geq 1$, thus for any $t \geq 2$ we have

$$\{|X_{1,0}| \geq t\} \subseteq \bigcap_{r=1}^{\infty} \{|X_{r,\delta_r}| \geq t\} \subseteq \bigcap_{r=1}^{\infty} \mathcal{U}_r^c. \tag{7.12}$$

Let $\lambda_0(n) := \mathbb{E}[|X_{1,0}|] = n\mathbb{P}(Bin(n-1,p) = 0)$ be the expected number of vertices of degree 0, as in the statement of Theorem 2.3.2. Observe that by the Taylor expansion for the logarithm we have

$$\lambda_0(n) = n(1-p)^{n-1} = ne^{\log(1-p)(n-1)} = ne^{-np+O\left(np^2\right)} = e^{\log(n)-np+o(1)}. \tag{7.13}$$

Now if $\lim_{n\to\infty} np - \log(n) = -\infty$ then $\lambda_0(n) \to \infty$ and so by Theorem 2.3.2 we have $\mathbb{P}(|X_{1,0}| > t) \to 1$ for any fixed $t$. The result follows by (7.12). $\square$

What is not trivial to establish is that if the smallest first neighbourhood is non zero then the smallest $r$-neighbourhood is attained uniquely. This statement is formulated in the following conjecture.

**Conjecture 7.4.2.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where $\liminf_{n\to\infty} np - \log(n) > -\infty$. Let $r := r(n) \geq 2$ be such that $(np)^{2r} = o(n)$. Then*

$$\mathbb{P}\big(\mathcal{U}_r \big| \delta_1 \geq 1\big) = 1 - o(1).$$

The conjecture above states that conditional on the graph having minimum degree 1 (and thus being connected w.h.p.) we have that the smallest $r$-neighbourhood is attained uniquely. Proposition 7.14 below shows that Conjecture 7.4.1 follows from Conjecture 7.4.2. Hopefully the proof of Proposition 7.14 will explain the conjectured behaviour of $\mathbb{P}(\mathcal{U}_r)$ around $\log(n)$ threshold. In particular we shall see that if Conjecture 7.4.2 holds then $\mathbb{P}(\mathcal{U}_r)$ is asymptotically equivalent to the probability that there are exactly 0 or 1 vertices of degree zero.

**Proposition 7.4.3.** *Let $\mathcal{G} \sim_d \mathcal{G}(n,p)$ where $\liminf_{n\to\infty} np - \log(n) > -\infty$. Let $r := r(n) \geq 2$ be such that $(np)^{2r} = o(n)$. If*

$$\mathbb{P}\big(\mathcal{U}_r \big| \delta_1 \geq 1\big) = 1 - o(1), \tag{7.14}$$

*then Items* (ii) *and* (iii) *of Conjecture 7.4.1 hold.*

*Proof.* To begin recall Lemma 7.3.2 and observe the following decomposition

$$\begin{aligned}\mathbb{P}(\mathcal{U}_r \cap \{\delta_1 = 0\}) &= \mathbb{P}(\mathcal{U}_r \cap \{\delta_1 = 0\} \cap \{X_{1,0} = X_{r,0}\}) \\ &\qquad + \mathbb{P}\big(\mathcal{U}_r \cap \{\delta_1 = 0\} \,\big|\, \{X_{1,0} \neq X_{r,0}\}\big)\, \mathbb{P}(X_{1,0} \neq X_{r,0}) \\ &= \mathbb{P}(|X_{1,0}| = 1) + o(1).\end{aligned}$$

Hence conditioning on the event that at least one vertex has no neighbours yields

$$\begin{aligned}\mathbb{P}(\mathcal{U}_r) &= \mathbb{P}(\mathcal{U}_r \cap \{\delta_1 = 0\}) + \mathbb{P}(\mathcal{U}_r \cap \{\delta_1 \geq 1\}) \\ &= \mathbb{P}(|X_{1,0}| = 1) + \mathbb{P}\big(\mathcal{U}_r \big| \delta_1 \geq 1\big)\, \mathbb{P}(\delta_1 \geq 1) + o(1). \end{aligned}\tag{7.15}$$

Now assuming (7.14) holds we are almost done.

For Item (ii) of Conjecture 7.4.1: If $w(n) \to \infty$ then $\mathbb{P}(\delta_1 \geq 1) = 1 - o(1)$ by Theorem 2.3.2. Thus by (7.14) we have $\mathbb{P}\big(\mathcal{U}_r \big| \delta_1 \geq 1\big)\, \mathbb{P}(\delta_1 \geq 1) = 1 - o(1)$ and so the result follows by (7.15).

For Item (iii) of Conjecture 7.4.1: If $-\infty < \liminf w(n) \leq \limsup w(n) < \infty$

then by Theorem 2.3.2 we have

$$\mathbb{P}(|X_{1,0}| = 1) \sim e^{-\lambda_0}\lambda_0 = e^{-e^{-w}}e^{-w},$$
$$\mathbb{P}(|X_{1,0}| = 0) \sim e^{-\lambda_0} = e^{-e^{-w}},$$

where $\mathbb{P}(|X_{1,0}| = 0) = \mathbb{P}(\delta_1 \geq 1)$. Plugging these expressions for $\mathbb{P}(\delta_1 \geq 1)$ and $\mathbb{P}(|X_{1,0}| = 1)$ into (7.15) yields the result. □

To motivate Conjecture 7.4.2 the overall variance of $\gamma_r(u)$ is large (of order $(np)^{2r-1}$) however, the number of $r$-neighbourhoods is $n$ so these neighbourhoods must take values in a wide range as we consider large $r$. It can also be seen by an argument in the proof of Theorem 7.3.4 that there cannot be many vertices attaining a first neighbourhood of size $\delta_1$. This leads me to believe that the range of sizes for the second neighbourhoods of vertices with degree $\delta_1$ is fairly large and only one of them attains the minimum. In addition (much in the spirit of Theorem 7.1.1 and Lemma 7.3.3) no vertex with a first neighbourhood larger than the smallest attains a minimal sized second neighbourhood. To prove this would require control of the joint distribution of the $r$-neighbourhoods.

# Bibliography


[1] *Techniques in Random Discrete Structures, National and Kapodistrian University of Athens, Greece, May 22-27, 2017.*

[2] David Aldous and James Allen Fill. Reversible Markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014.

[3] Romas Aleliunas, Richard M Karp, Richard J Lipton, Laszlo Lovasz, and Charles Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *FOCS*, volume 79, pages 218–223, 1979.

[4] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. Theory of computing (Singer Island, Fla., 1984).

[5] Noga Alon and Fan Chung. Explicit construction of linear sized tolerant networks. In *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, volume 72, pages 15–19, 1988.

[6] Noga Alon and Joel Spencer. *The probabilistic method.* Third edition, 2008.

[7] Tom M. Apostol. An elementary view of Euler's summation formula. *Amer. Math. Monthly*, 106(5):409–418, 1999.

[8] László Babai, Paul Erdős, and Stanley M. Selkow. Random graph isomorphism. *SIAM J. Comput.*, 9(3):628–635, 1980.

[9] Laurent Bartholdi and Bálint Virág. Amenability via random walks. *Duke Math. J.*, 130(1):39–56, 2005.

[10] Itai Benjamini, Simi Haber, Michael Krivelevich, and Eyal Lubetzky. The isoperimetric constant of the random graph process. *Random Structures Algorithms*, 32(1):101–114, 2008.

[11] Itai Benjamini and Gady Kozma. A resistance bound via an isoperimetric inequality. *Combinatorica*, 25(6):645–650, 2005.

[12] Itai Benjamini and Gady Kozma. A resistance bound via an isoperimetric inequality. arXiv:0212322, 2012.

[13] Itai Benjamini, Gady Kozma, and Nicholas Wormald. The mixing time of the giant component of a random graph. *Random Structures Algorithms*, 45(3):383–407, 2014.

[14] Nathanaël Berestycki, Eyal Lubetzky, Yuval Peres, and Allan Sly. Random walks on the random graph. *Ann. Probab.*, to appear, 2017.

[15] Béla Bollobás. Degree sequences of random graphs. *Discrete Math.*, 33(1):1–19, 1981.

[16] Béla Bollobás. The diameter of random graphs. *Trans. Amer. Math. Soc.*, 267(1):41–52, 1981.

[17] Béla Bollobás. Distinguishing vertices of random graphs. In *Graph theory (Cambridge, 1981)*, volume 62 of *North-Holland Math. Stud.*, pages 33–49. North-Holland, Amsterdam-New York, 1982.

[18] Béla Bollobás. Vertices of given degree in a random graph. *J. Graph Theory*, 6(2):147–155, 1982.

[19] Béla Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 1998.

[20] Béla Bollobás. *Random graphs.* Cambridge University Press, second edition, 2001.

[21] Béla Bollobás and Andrew Thomason. Random graphs of small order. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 47–97. North-Holland, Amsterdam, 1985.

[22] Nicolas Boumal and Xiuyuan Cheng. Concentration of the Kirchhoff index for Erdős–Rényi graphs. *Systems Control Lett.*, 74:74–80, 2014.

[23] Andrei Z. Broder and Anna R. Karlin. Bounds on the cover time. *J. Theoret. Probab.*, 2(1):101–120, 1989.

[24] Ashok K. Chandra, Prabhakar Raghavan, Walter L. Ruzzo, Roman Smolensky, and Prasoon Tiwari. The electrical resistance of a graph captures its commute and cover times. *Comput. Complexity*, 6(4):312–340, 1996/97.

[25] Fan Chung and Linyuan Lu. The diameter of sparse random graphs. *Adv. in Appl. Math.*, 26(4):257–279, 2001.

[26] Fan Chung and Linyuan Lu. *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. 2006.

[27] Amin Coja-Oghlan. On the Laplacian eigenvalues of $G_{n,p}$. *Combin. Probab. Comput.*, 16(6):923–946, 2007.

[28] Colin Cooper and Alan Frieze. The cover time of sparse random graphs. *Random Structures Algorithms*, 30(1-2):1–16, 2007.

[29] Colin Cooper, Alan Frieze, and Tomasz Radzik. Multiple random walks in random regular graphs. *SIAM J. Discrete Math.*, 23(4):1738–1761, 2009/10.

[30] Colin Cooper and Alan M. Frieze. Random walks on random graphs. In *NanoNet - Third International ICST Conference, NanoNet 2008, Boston, MA, USA, September 14-16, 2008, Revised Selected Papers*, pages 95–106, 2008.

[31] Edward T. Copson. *Asymptotic expansions*. Cambridge Tracts in Mathematics and Mathematical Physics, No. 55. Cambridge University Press, New York, 1965.

[32] Persi Diaconis. Some things we've learned (about Markov chain Monte Carlo). *Bernoulli*, 19(4):1294–1305, 2013.

[33] Peter G. Doyle and J. Laurie Snell. *Random walks and electric networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984.

[34] Arthur Erdélyi. *Asymptotic expansions*. Dover Publications, Inc., New York, 1956.

[35] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.*, 314(3):587–640, 2012.

[36] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

[37] Paul Erdős and Alfred Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[38] Paul Erdös and Alfred Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[39] Paul Erdős and Alfred Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12:261–267, 1961.

[40] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[41] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[42] Nikolaos. Fountoulakis and Bruce. A. Reed. The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures Algorithms*, 33(1):68–86, 2008.

[43] Alan Frieze and Michał Karoński. *Introduction to Random Graphs.* Cambridge University Press, 2015.

[44] Agelos Georgakopoulos. A tractable variant of cover time. arXiv:1206.6605, 2012.

[45] Agelos Georgakopoulos and Stephan Wagner. Hitting times, cover cost, and the Wiener index of a tree. *J. Graph Theory*, 84(3):311–326, 2017.

[46] Edgar N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.

[47] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.

[48] Grigori I. Ivčenko. The strength of the connectedness of a random graph. *Theor. Probability Appl.*, 18(2):396403, 1974.

[49] Svante Janson and Malwina J. Luczak. Susceptibility in subcritical random graphs. *J. Math. Phys.*, 49(12):125207, 23, 2008.

[50] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[51] Johan Jonasson. On the cover time for random walks on random graphs. *Combin. Probab. Comput.*, 7(3):265–279, 1998.

[52] Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[53] Jun Kigami. Resistance forms, quasisymmetric maps and heat kernel estimates. *Mem. Amer. Math. Soc.*, 216(1015):vi+132, 2012.

[54] Douglas J. Klein and Milan Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.

[55] Andreas Knoblauch. Closed-form expressions for the moments of the binomial probability distribution. *SIAM J. Appl. Math.*, 69(1):197–204, 2008.

[56] Pierre S. Laplace. Mmoire sur la probabilité des causes par les vnemens. *Memoir on the probability of causes of events. Mmoires de Mathmatique et de Physique, Tome Sixime. (English translation by S. M. Stigler 1986. Statist. Sci., 1(19):364378)*, 1774.

[57] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[58] László Lovász. Random walks on graphs: a survey. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 353–397. János Bolyai Math. Soc., Budapest, 1996.

[59] Matthias Löwe and Felipe Torres. On hitting times for a simple random walk on dense Erdös-Rényi random graphs. *Statist. Probab. Lett.*, 89:81–88, 2014.

[60] Russell Lyons and Yuval Peres. *Probability on Trees and Networks.* Cambridge University Press, 2016.

[61] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[62] Elchanan Mossel and Nathan Ross. Shotgun assembly of labeled graphs. arXiv:1504.07682, 2015.

[63] Asaf Nachmias and Yuval Peres. Critical random graphs: diameter and mixing time. *Ann. Probab.*, 36(4):1267–1286, 2008.

[64] Crispin St. J. A. Nash-Williams. Random walk and electric currents in networks. *Proc. Cambridge Philos. Soc.*, 55:181–194, 1959.

[65] Robin Pemantle. Uniform random spanning trees. In *Topics in contemporary probability and its applications*, Probab. Stochastics Ser., pages 1–54. CRC, Boca Raton, FL, 1995.

[66] Roger Penrose. A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.*, 51:406–413, 1955.

[67] Oliver Riordan and Nicholas Wormald. The diameter of sparse random graphs. *Combin. Probab. Comput.*, 19(5-6):835–926, 2010.

[68] Thomas Sauerwald and Alexandre Stauffer. Rumor spreading and vertex expansion on regular graphs. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 462–475. SIAM, Philadelphia, PA, 2011.

[69] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011.

[70] John Sylvester. Neighbourhood statistics of binomial random graphs. *In preparation.*

[71] John Sylvester. Random walk hitting times and effective resistance in sparsely connected Erdős-Rényi random graphs. *Submitted*, arXiv:1612.00731:41pp.

[72] Prasad Tetali. Random walks and the effective resistance of networks. *J. Theoret. Probab.*, 4(1):101–109, 1991.

[73] Ulrike von Luxburg, Agnes Radl, and Matthias Hein. Hitting and commute times in large random neighborhood graphs. *J. Mach. Learn. Res.*, 15:1751–1798, 2014.

[74] Stefan Wagner. Analytic combinatorics  part I: Laplaces method.

[75] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996)*, pages 296–303. ACM, New York, 1996.