2022

Randomly Perturbed Graphs and Rainbow Connectivity

John Arthur Finlay

Randomly Perturbed Graphs and Rainbow Connectivity

By

John Arthur Finlay

B.A., University of Montana, Missoula, USA, 2013

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Mathematics

University of Montana
Missoula, MT

May 2022

Approved by:

Scott Whittenburg, Dean of The Graduate School
Graduate School

P. Mark Kayll, Co-Chair
Mathematical Sciences

Cory Palmer, Co-Chair
Mathematical Sciences

Eric Chesebro
Mathematical Sciences

Nikolaus Vonessen
Mathematical Sciences

Richard Bridges
Biomedical And Pharmaceutical Sciences

Finlay, John, Ph.D., May 2022 Mathematical Sciences

Randomly perturbed graphs and rainbow connectivity

Co-Chairperson: P. Mark Kayll

Co-Chairperson: Cory Palmer

In this work we explore randomly perturbed graphs; that is, for an arbitrarily dense graph H we add a set R consisting of m edges randomly to create a graph G. We then randomly color the edges of G with r colors. We prove, for $r \geq 5$ and m a large enough constant, that between any two vertices in G there exists a rainbow path and thus G is rainbow connected. This result confirms a conjecture of Anastos and Frieze [How many randomly colored edges make a randomly colored dense graph rainbow Hamiltonian or rainbow connected?, *J. Graph Theory* **92** (2019), no. 4, 405–414] which resolved the case when $r \geq 7$ and m is a function of n (that tends to infinity arbitrarily slowly). We also explore concepts and results related to this result.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without countless people whose names have either escaped me or whom I never knew their name; note-takers, scribes, people moving the robot I used to attend class from room to room, and everyone who helped a physically disabled me get around. My hope is that if any of them read this, they will know how grateful I am.

In particular, I would like to thank my family for all of their support, patience, and help through the long process of getting my degree. A special thanks to my wife, Abby, my mother, Jane, and, of course, Grandpa Gerry Morgan.

I have spent over a third of my life in the University of Montana Mathematics Department and have nothing but kind words for everyone involved. Their guidance is so appreciated.

I am grateful for the efforts of my entire committee. I would especially like to thank my advisors, Mark Kayll and Cory Palmer, without whom this dream would have never been realized. I am incredibly grateful to know them and am truly a better person for it.

To Abby.

Of all of the joys in my life, by far the greatest is getting to spend my life with you.

TABLE OF CONTENTS

CHAPTER 1: Foundations

The main mathematical object(s) under study in this dissertation are combinatorial graphs, particularly randomly generated ones. In this chapter, we introduce the basic notion of a graph, along with notation and other definitions that will be used when discussing graph theory. Diestel's book [18] provides a suitable reference for any omitted terminology. Additionally, we prove two tools from probability theory, namely Boole's Inequality and the Chernoff Inequality, which will be used in the proof of the main result (Theorem 5.2). The books by Alon and Spencer [3] and Ross [33] contain reference material for any probability theory needed herein.

1.1 Basic Definitions and Notation

When we informally describe a graph, we think of a set of points, which we call vertices, and lines between these points, which we call edges. Consider almost any network, such as the calls between members of a crime ring or flights between airports in the world. These networks can be modeled using graphs.

Formally, a *graph* G consists of a pair $(V(G), E(G))$ where $V(G)$ is a set of *vertices* of G and $E(G)$ is a set of *edges* made up of 2-element subsets of $V(G)$. The cardinality of the set of vertices is called the *order* of G and is denoted by n. The cardinality of the set of edges is called the *size* of G and is denoted by m. A *subgraph* of G is a graph whose vertex and edge sets are, respectively, subsets of $V(G)$ and $E(G)$. If a subgraph is obtained from G by deleting only edges, i.e., it has the same vertex set, that subgraph is a *spanning subgraph*. An *induced subgraph* H begins with some subset $V(H)$ of the vertices of G and contains all of the edges of G joining pairs within $V(H)$.

An edge $\{u, v\}$ of G is typically denoted uv or vu, and u is said to be *adjacent* to v and vice versa. An edge that connects to a vertex is said to be *incident* to that vertex. Two edges that share a vertex are *incident*. The set of vertices adjacent to a particular vertex u is called the *neighborhood* of u and is denoted by $N(u)$. All the graphs in this dissertation are *simple*, meaning that there is only one edge between adjacent vertices and no edge joins a vertex to itself. The number of edges incident to a vertex u is the *degree* of u. The minimum degree of G is denoted by $\delta(G)$ and the maximum degree of G is denoted by $\Delta(G)$. Loosely speaking a graph with a 'high' δ, and thus a high ratio of size to order, is said to be *dense*. A graph with a 'low' Δ, and thus a low ratio of size to order, is said to be *sparse*.

In a graph G, a *walk* is a sequence of vertices where each vertex is adjacent to the next vertex in the sequence. If all the vertices in a walk are distinct, it is a *path*. When referring to a path, we may denote the sequence of vertices as, for example, $wxyz$; we might also denote the path only by its endvertices such as $w - z$, and in this case we call it a wz-path. A walk in which the initial and final vertices are the same vertex is called a *circuit*. A circuit where all of the vertices are distinct is a *cycle*; if a cycle contains every vertex in G, it is called a *Hamilton cycle*. The length (in edges) of a shortest cycle in a graph is the *girth*, and the length of a longest cycle is the *circumference*. If any two vertices in G have a path between them, then G is *connected*; otherwise, G is *disconnected*. The diameter of G is the length of a longest path in G and is denoted by $diam(G)$. A maximal connected subgraph of a disconnected graph is a *component* of G. If there exists an edge in a connected graph, such that removing that edge causes the graph to become disconnected, that edge is called a *bridge*. A graph is said to be *k-connected* if upon removal of any set of $k - 1$ or fewer vertices, the graph remains connected.

A graph is a *tree* if any two of its vertices are connected by exactly one path. An *independent set* S is a subset of $V(G)$ such that no two vertices in S are adjacent.

1.2 Boole's Inequality

Throughout this work we will utilize the 'probabilistic method' in trying to understand randomly created objects (see, e.g., [3]). This method invokes many tools from probability theory. One of the most common tools used, in fact, a tool used in the first use of the probabilistic method (see [3, p. 16]), is Boole's Inequality or the 'union bound'. This says that the probability that at least one event happens (among a collection of events) is never greater than the sum of the probabilities of the events in the collection.

Theorem 1.1 (Boole's Inequality)**.** *For any events* $E_1, E_2, \ldots, E_n$ *we have*

$$Pr\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} Pr(E_i).$$

Proof. We proceed by induction on n, for which the basis case is trivial. So suppose that for some n and collection of events $E_1, E_2, \ldots, E_n$ we have

$$Pr\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} Pr(E_i).$$

Now recall the basic probability result:

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

for any events A and B.

We now set $A = \bigcup_{i=1}^{n} E_i$ and $B = E_{n+1}$, and observing that $\bigcup_{i=1}^{n+1} E_i = A \cup B$, we obtain

$$Pr\left(\bigcup_{i=1}^{n+1} E_i\right) = Pr\left(\bigcup_{i=1}^{n} E_i\right) + Pr(E_{n+1}) - Pr\left(\left(\bigcup_{i=1}^{n} E_i\right) \cap E_{n+1}\right).$$

As the probability of an event can never be negative

$$Pr\left(\bigcup_{i=1}^{n} E_i \cap E_{n+1}\right) \geq 0;$$

therefore,

$$Pr\left(\bigcup_{i=1}^{n+1} E_i\right) \leq Pr\left(\bigcup_{i=1}^{n} E_i\right) + Pr(E_{n+1}).$$

Now by induction,

$$Pr\left(\bigcup_{i=1}^{n+1} E_i\right) \leq \sum_{i=1}^{n} Pr(E_i) + Pr(E_{n+1}) = \sum_{i=1}^{n+1} Pr(E_i).$$

□

1.3 Chernoff Inequality

Another tool taken from probability theory and used in the exploration of random objects through the probabilistic method, including in the main theorem of this dissertation, is the Chernoff Inequality. The Chernoff Inequality gives us a good bound, an exponentially decreasing bound, on the sum of random variables.

We begin with Markov's Inequality, which measures the spread from the mean of a distribution. Although Markov's Inequality is generally weak in practice, it is useful in deriving better bounds, including the Chernoff Inequality.

Theorem 1.2 (Markov's Inequality)**.** *If X is a random variable that assumes only nonnegative values, then, for all $a > 0$, we have*

$$Pr(X \geq a) \leq \frac{E[X]}{a}.$$

See [17] for a proof.

We now give a statement and proof of the Chernoff Inequality (which are adapted from [31, p. 66]). There are many ways that the Chernoff Inequality is written; below we give two. The first is stronger, but the second one is generally sufficient and easier to use in practice.

Theorem 1.3 (Chernoff Inequality)**.** *Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials and let $X = \sum_{i=1}^{n} X_i$. Then, for $0 < \alpha < 1$, we have*

1. $$Pr(X \leq (1-\alpha)E[X]) \leq \left(\frac{e^{-\alpha}}{(1-\alpha)^{(1-\alpha)}}\right)^{E[X]};$$

2. $$Pr(X \leq (1-\alpha)E[X]) \leq \exp\left(-\frac{1}{2}\alpha^2 E[X]\right).$$

Proof. (from [31]) By employing Markov's Inequality (Theorem 1.2), for any $t < 0$ we observe the following:

$$\begin{aligned} Pr(X \leq (1-\alpha)E[X]) &= Pr(e^{tX} \geq e^{t(1-\alpha)E[X]}) \\ &\leq \frac{E[e^{tX}]}{e^{t(1-\alpha)E[X]}} \\ &\leq \frac{e^{(e^t-1)E[X]}}{e^{t(1-\alpha)E[X]}}. \end{aligned}$$

Now for $0 < \alpha < 1$ and by setting $t = \ln(1-\alpha) < 0$ we get the first result,

$$Pr(X \leq (1-\alpha)E[X]) \leq \left(\frac{e^{-\alpha}}{(1-\alpha)^{(1-\alpha)}}\right)^{E[X]}.$$

To finish the proof for the second result, we have to show that for $0 < \alpha < 1$, the following holds:

$$\frac{e^{-\alpha}}{(1-\alpha)^{(1-\alpha)}} \leq e^{-\frac{1}{2}\alpha^2}.$$

Equivalently, we need to show that

$$f(\alpha) := -\alpha - (1-\alpha)\ln(1-\alpha) + \frac{1}{2}\alpha^2 \leq 0$$

for $0 < \alpha < 1$. One easily checks that f is nondecreasing for $\alpha \in (0,1)$, and, as $f(0) = 0$, it follows that $f(\alpha) \leq 0$ for $0 < \alpha < 1$. □

Another form of the Chernoff Inequality that is used often is the following consequence.

Corollary 1.4. *If $X_1, X_2, \ldots, X_n$ are independent Poisson trials and $X = \sum_{i=1}^{n} X_i$, then, for $0 < \alpha < 1$, we have*

$$Pr(|X - E[X]| \geq \alpha E[X]) \leq 2e^{-\alpha^2 E[X]/3}.$$

1.4 Chebyshev's Inequality

Another useful estimate in probabilistic combinatorics is Chebyshev's Inequality. For example, it forms the basis for the so-called 'Second Moment Method' (see, e.g., [3], which devotes an entire chapter to the method). As for the Chernoff Inequality, this estimate is again derived from Markov's Inequality. The proof presented here is from Ross [33].

Theorem 1.5 (Chebyshev's Inequality)**.** *Let X be a random variable with mean μ and variance σ^2. If $c > 0$, then the following holds:*

$$Pr(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}.$$

Proof. As $(X - \mu)^2$ is a positive random variable, we can apply Markov's Inequality (Theorem 1.2) to get

$$Pr((X - \mu)^2 \geq c^2) \leq \frac{E[(X - \mu)^2]}{c^2}.$$

Note that $(X - \mu)^2 \geq c^2$ if and only if $|X - \mu| \geq c$, and thus

$$Pr(|X - \mu| \geq c) \leq \frac{E[(X - \mu)^2]}{c^2}.$$

Recalling that $E[(X - \mu)^2] = \text{Var}[X] = \sigma^2$ gives us our result. □

CHAPTER 2: Rainbow Connectivity

The concept of a rainbow connected graph was first introduced by Chartrand et al. [14]. The topic was motivated by a problem articulated by Erickson [21]. The following is a paraphrasing of that motivation: In the wake of the terrorist attacks on September 11, 2001, it became apparent that government agencies were not able to communicate through regular channels — differences in technology and databases prohibited shared access. The result was that law enforcement officers and intelligence agents were unable to cross-check information across various agencies.

The classified nature of the national security information, along with the desire to be able to share access between appropriate parties, introduced an issue. How could transfer paths be created between any two agencies using other agencies as intermediaries that were large enough to have adequate numbers of passwords and firewalls to prevent hackers, but not large enough to become unmanageable? That is, there had to exist at least one path between any two agencies where no password was repeated.

The question arises, what is the minimum number of passwords and firewalls required to create at least one secure path between any two agencies where all the passwords are distinct?

In this chapter, we introduce the topic of rainbow connected graphs, which will be integral to the understanding of our main result (Theorem 5.2). We will introduce the definitions and notation, along with a survey of results that will give context to this theorem. We prove some of these theorems as an illustration of constructive proofs in graph theory. In subsequent chapters, we shall primarily focus on non-constructive (i.e. probabilistic) proofs.

2.1 Definitions and Notation Regarding Rainbow Connectivity

Consider a set of k colors represented by $[k] = \{1, 2, \ldots, k\}$. An *edge-coloring* of G is a function $c : E(G) \to [k]$ where incident edges are allowed to be the same color. A *rainbow path* is a path in which no two edges in the path are the same color. A k-edge-colored graph G is *rainbow connected* if there exists a rainbow path between any two vertices in G. Naturally, it follows that if G is rainbow connected then it is also connected. The converse of this is that if G is connected then there is a trivial edge-coloring that makes it rainbow connected, simply by coloring every edge of G a different color.

For a connected graph G, the *rainbow connectivity* of G, denoted $rc(G)$, is the minimum number of colors required to make G rainbow connected. A rainbow coloring of G using $rc(G)$ colors is called a *minimum rainbow coloring*. If H is a connected spanning subgraph of G, then $rc(G) \leq rc(H)$.

2.2 Bounding Rainbow Connectivity

The primary focus of our exploration of rainbow connected graphs will be attempting to find a bound on $rc(G)$. To that end, first we provide a few bounds that are more or less trivial.

Theorem 2.1. *Let G be a connected graph on n vertices. The following are true:*

1. $rc(G) = 1$ *if and only if* G *is the complete graph on* n *vertices;*

2. $rc(G) = n - 1$ *if and only if* G *is a tree;*

3. $rc(G) \geq diam(G)$*;*

4. *if* G *is a cycle of length* $n > 3$*, then* $rc(G) = \lceil n/2 \rceil$.

In [13], Caro et al. prove the following lemma and two propositions. To give an idea of the flavor of these proofs, we present two arguments from [13] below.

Lemma 2.2. *If* G *is a connected graph, and* $H_1, H_2, \ldots, H_k$ *are connected subgraphs with* $\{V(H_1), V(H_2), \ldots, V(H_k)\}$ *forming a partition of* $V(G)$*, then*

$$rc(G) \leq k - 1 + \sum_{i=1}^{k} rc(H_i).$$

Proposition 2.3. *If* G *is a 2-connected graph on* n *vertices, then*

$$rc(G) \leq \frac{2n}{3}.$$

Proof. We first consider the case where G is a cycle of length 5. Then clearly $rc(G) = 3$; thus the proposition holds.

So we may assume that G is not a 5-cycle. To obtain our desired bound, it will help to consider a maximal connected subgraph H of G—of order h—such that

$$rc(H) \leq \frac{2h}{3} - \frac{2}{3}. \tag{2.1}$$

To see that H exists, we consider a few cases depending on the girth of G.

If G contains a triangle T, then

$$rc(T) = 1 \leq \frac{2|V(T)|}{3} - \frac{2}{3},$$

so at minimum, we could take $H = T$. If G contains a cycle C of length $k \geq 4$ (but $k \neq 5$), then Theorem 2.1 shows that

$$rc(C) = \left\lceil \frac{k}{2} \right\rceil \leq \frac{2k}{3} - \frac{2}{3}, \tag{2.2}$$

and again we see that H exists. (Note that we needed $k \neq 5$ to support the inequality in (2.2).) Now we may suppose that each cycle in G is a C_5, but since $G \neq C_5$, we know that G contains a C_5 together with an incident edge that is not a chord; call this subgraph C^+. Then

$$rc(C^+) = 3 < \frac{2 \cdot 6}{3} - \frac{2}{3} = \frac{2|V(C^+)|}{3} - \frac{2}{3},$$

and now at a minimum, we could take $H = C^+$. We have now considered all possible cases because the girth of a 2-connected graph must be finite.

Now that we know a maximal H satisfying (2.1) exists, we shall prove that $h \geq n - 2$. For a contradiction, assume that there are three distinct vertices—say x_1, x_2, x_3—outside of H such that each of these vertices has two neighbors in H. The neighbors of x_i in H do not have to be distinct from the neighbors of x_j in H. We can add x_1, x_2, x_3 to H to form a larger subgraph H' which has $h + 3$ vertices. Suppose that e_i, f_i are two edges connecting x_i to H. By using only two new colors to color the six designated edges (e_1, e_2, e_3 get one color and f_1, f_2, f_3 get the other color) we get

$$rc(H') \leq rc(H) + 2 \leq \frac{2h}{3} - \frac{2}{3} + 2 = \frac{2(h+3)}{3} - \frac{2}{3},$$

which contradicts the maximality of H. It follows that for at least one of the vertices, say x, the shortest path from H to H through x has length at least 3; such a path has to exist because the graph is 2-connected.

Let $ax_1 \cdots x_t b$ be a path where $a, b \in H$ and $x_1, \ldots, x_t \notin H$ with $t \geq 2$. We can add $x_1, \ldots, x_t$ to H to form a larger subgraph H' which has $h + t$ vertices. If t is odd, then we can color the $t + 1$ edges with $(t + 1)/2$ new colors. We color the first half of the path with each edge being a distinct color and repeat that coloring for the second half of the path. Clearly, H' is rainbow connected.

If t is even, then we color the $t + 1$ edges with $t/2$ colors. The middle edge of the path, $x_{t/2}x_{t/2+1}$, we color the same color as one of the colors already in H. We color each of the first $t/2$ edges with a distinct color and repeat that coloring for the second $t/2$ edges. Again, H' is rainbow connected.

We observe that

$$rc(H') \leq rc(H) + \left\lceil \frac{t}{2} \right\rceil \leq \frac{2h}{3} - \frac{2}{3} + \left\lceil \frac{t}{2} \right\rceil \leq \frac{2(h+t)}{3} - \frac{2}{3},$$

again contradicting the maximality of H.

Now we have proven $h \geq n - 2$. Therefore, $rc(G) \leq 2(n-2)/3 - 2/3 + 2 = 2n/3$ as was to be shown. □

Caro et al. [13] also proved the following result about bridgeless graphs. This result helps us to build up to a more general class of graphs.

Proposition 2.4 ([13]). *If G is a connected bridgeless graph with n vertices, then*

$$rc(G) \leq \frac{4n}{5} - 1.$$

Proof. We begin first by considering the case where G is 2-connected. It follows from Proposition 2.3 that $rc(G) \leq \lfloor 2n/3 \rfloor \leq 4n/5 - 1$ for all $n \geq 7$. Since all 2-connected graphs contain a cycle, it follows from Lemma 2.2 (where we take H_1 to be a cycle and all other H_i's to be singletons) that $rc(G) \leq n - 2$. We also observe that $n - 2 \leq 4n/5 - 1$ when $n = 3, 4, 5$.

Therefore, the only remaining case is when $n = 6$. It remains to be shown that, in this case, $rc(G) \leq 3$. If the longest cycle in G has length 6, then $rc(G) \leq 3$. If the longest cycle in G has length 5, then G is C_5 with an additional vertex of degree 2 whose neighbors are two non-adjacent vertices within the C_5. It is then easy to see that this graph has $rc(G) = 3$. If the longest cycle in G has length 4, then G is $K_{2,4}$. This implies that $rc(G) = 2$. A 2-connected graph on n vertices necessarily contains a cycle of length of at least 4. Therefore, we have proven the proposition for all 2-connected graphs.

We now want to expand the proof to include all bridgeless graphs. In order to do this, we will use induction on the number of 2-connected components. We set X to be the set of vertices of a 2-connected component of G such that X contains only one cut vertex — call it x. We consider the subgraph H of G induced by $(V(G)\backslash X) \cup \{x\}$. It follows that H has $n - |X| + 1$ vertices and has the following properties; it is connected, bridgeless, and with one fewer 2-connected components. By the induction hypothesis,

$$rc(H) \leq \frac{4(n - |X| + 1)}{5} - 1.$$

As X induces a 2-connected graph, it follows from the observations in the first paragraph that $rc(X) \leq 4|X|/5 - 1$. Therefore,

$$rc(G) \leq \frac{4n}{5} - 2 + \frac{4}{5} < \frac{4n}{5} - 1,$$

as was to be shown. □

One area of interest we will encounter in our main result (Theorem 5.2) is that of the relationship between $rc(G)$ and the minimum degree of a graph. Caro et al. [13] (see also [30]) proved the following bound for connected graphs of minimum degree at least 3.

Theorem 2.5. *If G is a connected graph on n vertices with $\delta(G) \geq 3$, then*

$$rc(G) \leq \frac{5n}{6}.$$

We finish this chapter on rainbow connected graphs and bounding the rainbow connectivity with the following result, also from Caro et al. [13] (see also [30]). This theorem relates the rainbow connectivity to minimum degree and shows, not surprisingly, that as the minimum degree increases, the rainbow connectivity decreases.

Theorem 2.6. *If G is a graph on n vertices with minimum degree δ, then*

$$rc(G) \leq \min\left\{n\frac{\ln\delta}{\delta}(1 + o_\delta(1)), n\frac{4\ln\delta + 3}{\delta}\right\}.$$

CHAPTER 3: Random Graph Theory

In [19], Erdős introduced the notion of a random graph to prove the existence of graphs with high girth and high chromatic number. From this beginning, Erdős and Rényi published a series of papers, highlighted by [20], where the two fleshed out the field.

The following is a paraphrased analogy Bollobás explained in [9]: It can sometimes be helpful to think of a random graph as a living organism going through evolution. We think of the organism starting out as a set of vertices with no edges. As the number of edges increases through some random process, additional properties "evolve" that were not there before. For example, small components might show up, followed by cliques, and eventually the graph becomes complete.

In this chapter, we introduce definitions and notation related to the study of random graphs. We will also introduce the two most commonly used random graph models and prove that they are highly linked. We will then introduce the reader to a tool that is integral to the study of random graphs: *the probabilistic method*. We then develop the idea of a threshold and establish some thresholds for various graph properties. Finally, we will explore some theorems related to connectivity in random graphs as well as a threshold for rainbow connectivity.

3.1 Random Graph Definitions, Notation, and Basic Models

Following [20], we define $\mathcal{G}_{n,m}$ to be the collection of graphs with vertex set $[n]$ and m edges where $0 \leq m \leq \binom{n}{2}$. To every graph $G \in \mathcal{G}_{n,m}$, we assign the following probability:

$$Pr(G) = \binom{\binom{n}{2}}{m}^{-1}.$$

Another way to think of this model is to begin with a set of n vertices and add distinct m edges uniformly at random. We call a graph of this type a *uniform random graph* and denote it as $G_{n,m} = ([n], E_{n,m})$. It is important to note that $\mathcal{G}_{n,m}$ is not a particular graph but rather the collection of all graphs on n vertices with m edges placed uniformly at random.

In [25], Gilbert introduced a similar model. Here, we define $\mathcal{G}_{n,p}$ to be the set of graphs on n vertices with the following edges: Fix $0 \leq p \leq 1$; then, for $0 \leq m \leq \binom{n}{2}$, assign each graph G with vertex set $[n]$ and m edges a probability of

$$Pr(G) = p^m(1-p)^{\binom{n}{2}-m}.$$

Another way to think about this model is to consider an empty graph on vertex set $[n]$ and then perform $\binom{n}{2}$ Bernoulli trials[1] — one for each pair of vertices — and add an edge between them with probability p. We call this graph a *binomial random graph* and denote it as $G_{n,p} = ([n], E_{n,p})$. Here again, it is important to remember that $\mathcal{G}_{n,p}$ is a collection of graphs rather than a specific graph.

The following basic lemma (see, e.g., Frieze and Karoński [22]) shows just how closely related the above two models are:

[1] A Bernoulli trial is an experiment with exactly two outcomes.

Lemma 3.1. *A random graph $G_{n,p} \in \mathcal{G}_{n,p}$, given that its number of edges is m, is equally likely to be any one of the $\binom{\binom{n}{2}}{m}$ graphs that have m edges.*

Proof. Suppose that G_0 is any labeled graph with m edges. Observe that

$$\begin{aligned} Pr(G_{n,p} = G_0 : |E_{n,p}| = m) &= \frac{Pr(G_{n,p} = G_0, |E_{n,p}| = m)}{Pr(|E_{n,p}| = m)} \\ &= \frac{Pr(G_{n,p} = G_0)}{Pr(|E_{n,p}| = m)} \\ &= \frac{p^m (1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}} \\ &= \frac{1}{\binom{\binom{n}{2}}{m}}. \end{aligned}$$

□

Additionally, it makes intuitive sense that the two models behave in a similar fashion when n is large enough and the number m of edges in $G_{n,m}$ is close to the expected number of edges in $G_{n,p}$. This occurs when

$$m = \binom{n}{2} p \approx \frac{n^2 p}{2},$$

i.e., when

$$p \approx \frac{2m}{n^2}.$$

We define a *graph property*, $\mathcal{P}$, as a subset of all labeled graphs on n vertices. Some examples of graph properties are connected graphs, planar graphs, and Hamiltonian graphs. We define a property $\mathcal{P}$ to be *monotone increasing* if adding an edge e to G does not destroy that property; that is, $G \in \mathcal{P}$ implies that $G + e \in \mathcal{P}$ for $e \notin E(G)$. Some examples of monotone increasing properties are graphs with clique number at least 5 and Hamiltonian graphs. We define a property $\mathcal{P}$ to be *monotone decreasing* if the removal of an edge e does not destroy that property; that is, $G \in \mathcal{P}$ implies that $G - e \in \mathcal{P}$ for $e \in E(G)$. Some examples of monotone decreasing properties are disconnected graphs and planar graphs.

3.2 The Probabilistic Method

One strategy for understanding random graphs is through the probabilistic method. In the probabilistic method, we use tools taken from probability theory to prove statements in combinatorics. For example, if we intend to show that a structure with certain desired properties exists, we construct an appropriate probability space and show that a structure with desired properties has positive probability in the space. Equivalently, if we can show the probability that a structure with the desired properties does not exist is less than one, then we have shown that the desired structure must exist. Proofs involving the use of the probabilistic method are what we call non-constructive proofs as they (usually) show that something exists but do not construct it explicitly. This is in contrast to the constructive proofs seen in the last chapter. To understand the difference between these two methods of proof, think of two cities; a non-constructive proof would argue that there exists a route from one city to the other, whereas a constructive proof would draw a map.

Perhaps the best way to understand the probabilistic method is through examples of its use. With that in mind, we examine so-called Ramsey numbers. Introduced by Ramsey [32] in 1929, the *Ramsey number* $R(k,l)$ is the smallest integer n such that any given two-coloring, say in red and blue, of the edges of the complete graph K_n must have either an all-blue K_k or an all-red K_l. In [32], Ramsey showed that $R(k,l)$ is finite for any positive integers k and l. We present a proof of a lower bound on the diagonal Ramsey numbers due to Erdős and adapted from [3].

Theorem 3.2. *$R(k,k) > \lfloor 2^{k/2} \rfloor$ for all $k \geq 3$.*

Proof. Suppose we color the edges of K_n with blue and red independently and uniformly at random, i.e., with probability $p = 1/2$. For any fixed set K of k vertices, let A_K be the event that the induced subgraph on K is monochromatic; i.e., all edges in K are the same color. It follows that

$$Pr(A_K) = 2^{1-\binom{k}{2}}.$$

As there are $\binom{n}{k}$ choices for K, we have, by Boole's Inequality (Theorem 1.1), that the probability that at least one of the events A_K occurs is at most

$$\binom{n}{k} 2^{1-\binom{k}{2}}.$$

If the above expression is less than 1, then with positive probability, no event A_K occurs. That is, there exists a two-coloring of $E(K_n)$ which contains no monochromatic K_k. To conclude, set $n = \lfloor 2^{k/2} \rfloor$, and observe that when $k \geq 3$ we have

$$\binom{n}{k} 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \cdot \frac{n^k}{2^{\frac{k^2}{2}}} < 1.$$

Therefore,

$$R(k,k) > \lfloor 2^{k/2} \rfloor$$

for all $k \geq 3$. □

A treatment of the probabilistic method and random graphs would not be complete without mention of a result of Erdős [19], where he used random graphs and the probabilistic method to prove an incredibly surprising result.

Recall that the girth $g(G)$ of a graph G is the length of a shortest cycle in G and $\chi(G)$ is the chromatic number of G.

Theorem 3.3 ([19])**.** *For all k, l there exists a graph G with $g(G) > l$ and $\chi(G) > k$.*

I personally would be remiss if I finished this section without including the following result, one of my all-time favorite proofs. It also presents how clean and clever a proof using the probabilistic method can be. The result comes from Ajtai et al. [2]. We present a proof adapted from [1].

We begin by defining a few terms. If a graph can be drawn in the plane so that none of the edges cross, then we say that the graph is *planar*. Otherwise we say G is *non-planar*. When G is non-planar, the minimum number of crossings required to draw G on a plane is the *crossing number* $cr(G)$ of G, which we denote by $cr(G)$.

Theorem 3.4 ([2])**.** *If G is a graph on n vertices with m edges, where $4n \leq m$, then*

$$cr(G) \geq \frac{1}{64}\frac{m^3}{n^2}.$$

Proof. Fix $p = 4n/m$ and select vertices randomly with probability p from G. Call the induced subgraph created by this process G_p. Now let n_p, m_p, X_p be random variables that count, respectively, the vertices, edges, and crossings in G_p. Euler's formula implies the following inequality:

$$cr(G) - m + 3n \geq 0.$$

Thus, we have

$$E[X_p - m_p + 3n_p] \geq 0.$$

We now compute the expected values of the random variables n_p, m_p, and X_p. Observe that $E[n_p] = pn$, $E[m_p] = p^2m$ as an edge exists if and only if both endpoints are present in G_p, and $E[X_p] = p^4cr(G)$ as a crossing exists in G_p if and only if all four distinct vertices are present.

Therefore, by linearity of expectation we have

$$0 \leq E[X_p] - E[m_p] + 3E[n_p] = p^4cr(G) - p^2m + 3pn.$$

Isolating $cr(G)$ gives

$$\begin{aligned} cr(G) &\geq \frac{p^2m - 3pn}{p^4} \\ &= \frac{m}{p^2} - \frac{3n}{p^3} \\ &\geq \frac{1}{64}\left(\frac{4m}{(n/m)^2} - \frac{3n}{(n/m)^3}\right) \\ &= \frac{1}{64}\frac{m^3}{n^2}. \end{aligned}$$

□

3.3 Thresholds

One of the most interesting and unexpected aspects of random graphs is the presence of thresholds: the sudden appearance of graph properties during the "evolution" of a random graph.

Some authors use the term "phase transition" instead of threshold. This term is borrowed from physics and is a particularly useful metaphor. We think of 0 Celsius as a threshold for the state of water. Water freezes below 0 Celsius, and water is liquid above 0 Celsius. Similar to thresholds for random graphs, this change is abrupt.

We begin by defining a threshold for a monotone increasing property $\mathcal{P}$.

Definition 3.5. *A function $p(n)$ is a* threshold *for a monotone increasing property $\mathcal{P}$ in the random graph $G_{n,p}$ if*

$$\lim_{n\to\infty} Pr(G_{n,p} \in \mathcal{P}) = \begin{cases} 0 & \textit{if } p/p(n) \to 0 \\ 1 & \textit{if } p/p(n) \to \infty. \end{cases}$$

It is straightforward to adjust the above definition for monotone decreasing properties.

As a concrete example, we give a threshold for the appearance of a triangle in the random graph $G_{n,p}$. This follows from a more general result of Erdős and Rényi [20]. The proof is an application of the second moment method (and so uses Chebyshev's inequality, Theorem 1.5).

Theorem 3.6. *A threshold for a random graph $G_{n,p}$ to contain a triangle is $1/n$.*

The search for and the refinement of thresholds continues to be an area of significant research. The following result of Bollobás and Thomason [10] shows that thresholds always exist for monotone graph properties.

Theorem 3.7 ([10])**.** *Every non-trivial monotone graph property has a threshold.*

3.4 Random Graphs and Connectivity

Some of the most important results in the theory of random graphs deal with connectivity. Indeed, most of the results from Erdős and Rényi in [20] have to do with this very topic. As we begin to narrow our focus to rainbow connectivity, it is important to highlight some of the results associated with ordinary connectedness. The following result is from [20] and the proof is adapted from [22]. We omit several of the details.

Theorem 3.8 ([20]). *A threshold function for connectivity in the random graph $G_{n,p}$ is* $\log n/n$.

Proof. Let X_k denote the number of components with k vertices in $G_{n,p}$. We now consider the event that $G_{n,p}$ is **not** connected. Its probability is

$$\begin{aligned} Pr(G_{n,p} \text{ is not connected}) &= Pr\left(\bigcup_{k=1}^{n/2}(G_{n,p} \text{ has a component of order } k)\right) \\ &= Pr\left(\bigcup_{k=1}^{n/2}(X_k > 0)\right). \end{aligned}$$

Applying Boole's inequality (Theorem 1.1) and using the fact that X_1 counts isolated vertices, we obtain

$$Pr(X_1 > 0) \leq Pr(G_{n,p} \text{ is not connected}) \leq Pr(X_1 > 0) + \sum_{k=2}^{n/2} Pr(X_k > 0).$$

Applying Markov's inequality (Theorem 1.2) and Cayley's formula for the number of spanning trees gives

$$\sum_{k=2}^{n/2} Pr(X_k > 0) \leq \sum_{k=2}^{n/2} E[X_k]$$
$$\leq \sum_{k=2}^{n/2} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)}.$$

Now we can select a constant c such that the next two bounds hold. For $2 \leq k \leq 10$, we have

$$\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \leq e^k n^k \left(\frac{\log n + c}{n}\right)^{k-1} e^{-k(n-10)\frac{\log n + c}{n}}$$
$$\leq (1+o(1)) e^{k(1-c)} \left(\frac{\log n}{n}\right)^{k-1}.$$

On the other hand, when $k > 10$ we have

$$\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \leq \left(\frac{ne}{k}\right)^k k^{k-2} \left(\frac{\log n + c}{n}\right)^{k-1} e^{-k(\log n + c)/2}$$
$$\leq n \left(\frac{e^{1-c/2+o(1)} \log n}{n^{1/2}}\right)^k.$$

Therefore,

$$\sum_{k=2}^{n/2} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \leq (1+o(1)) \frac{e^{-c} \log n}{n} + \sum_{k=10}^{n/2} n^{1+o(1)-k/2}$$
$$= O(n^{o(1)-1})$$
$$= o(1).$$

Hence, it follows that

$$Pr(G_{n,p} \text{ is connected}) = Pr(X_1 = 0) + o(1).$$

That is, the probability that $G_{n,p}$ is connected is asymptotic to the probability that it has no isolated vertex.

The remainder of the proof is to show that a threshold for the existence of isolated vertices is $\log n/n$. We set aside the details as the proof involves the use of the second moment method which is beyond the scope of this section. □

3.5 Random Graphs and Rainbow Connectivity

In the preceding section, we explored ordinary connectivity in random graphs. We now move one step beyond that to rainbow connectivity in random graphs. We begin with a threshold for rainbow connectivity being at least two in random graphs from Caro et al. [13].

At this point we need another definition in our repertoire: For a sequence (E_n) of events we say that the sequence occurs *with high probability* (w.h.p.) if $\lim_{n\to\infty} Pr(E_n) = 1$.

Recall that $rc(G)$ denotes the rainbow connectivity of G, i.e., the minimum number of colors required to make G rainbow connected.

Theorem 3.9 ([13]). *Let $\mathcal{P}$ be the property that a graph G has $rc(G) \leq 2$. A threshold for $\mathcal{P}$ is $\sqrt{\log n/n}$.*

Proof. We begin by showing that for a sufficiently large constant c, the graph $G_{n,p}$, where $p = c\sqrt{\log n/n}$, almost always has $rc(G) \leq 2$.

We first show that with high probability any two vertices x and y of G have at least $2\log n$ common neighbors. A vertex z is a common neighbor of x and y with probability $c^2 \log n/n$. Let X be the number of common neighbors of x and y. Then

$$E[X] = \frac{n-2}{n}(c^2 \log n) > \frac{1}{2}c^2 \log n.$$

As X is the sum of $n-2$ indicator variables of independent events we may apply the Chernoff Inequality (Corollary 1.4) to conclude that

$$Pr\left(X < \frac{1}{2}E[X]\right) = o(1/n^2)$$

for sufficiently large c. There are $\binom{n}{2}$ pairs x, y, so applying Boole's Inequality (Theorem 1.1) gives that for c large enough, with high probability, every pair of vertices has at least $\frac{1}{2}E[X] > \frac{1}{4}c^2 \log n \geq 2 \log n$ common neighbors.

Now, color the edges of G with two colors uniformly at random. For vertices x, y and a common neighbor z, the path xzy is rainbow with probability $1/2$. All such paths are pairwise edge-disjoint, so the probability that all these paths are not rainbow is at most $(1/2)^{2\log n} = 1/n^2$. As there are at most $\binom{n}{2}$ choices for x, y we may again apply Boole's Inequality to conclude that with positive probability each pair of vertices is joined by a rainbow path of length at most 2. Therefore, $rc(G) \leq 2$.

To prove the other direction, we need to show that for a sufficiently small constant k, the random graph $G_{n,p}$, where $p = k\sqrt{\log n/n}$, almost always has diameter greater than two.

We begin by fixing a set U with $n^{1/5}$ vertices. For simplicity, assume that $n^{1/5}$ is an even integer. We let W be the set of $n - n^{1/5}$ vertices not in U.

Let A be the event that U induces an independent set and let B be the event that there exists a pair of vertices in U with no common neighbor in W. Observe that if both A and B occur, then the diameter of G is at least 3. Therefore, it is sufficient to show that A and B each occur with high probability for small enough k.

For k sufficiently small we have

$$
\begin{aligned}
Pr(A) &= (1-p)^{\binom{n^{1/5}}{2}} \\
&= \left(1 - k\sqrt{\log n/n}\right)^{\binom{n^{1/5}}{2}} \\
&= 1 - o(1).
\end{aligned}
$$

We now bound $Pr(B)$. To this end, partition U into $|U|/2 = n^{1/5}/2$ pairs. For such a pair x, y, the probability that they have a common neighbor in W is

$$1 - \left(1 - \frac{c^2 \log n}{n}\right)^{n-n^{1/5}}.$$

It follows that for a sufficiently small constant k, the probability that all $|U|/2$ pairs have a common neighbor in W is

$$\left(1 - \left(1 - \frac{c^2 \log n}{n}\right)^{n-n^{1/5}}\right)^{n^{1/5}/2} = o(1).$$

Therefore, $Pr(B) = 1 - o(1)$. □

We conclude this chapter with a result of Frieze and Tsourakakis [23] that estimates the rainbow connectivity of a random graph with edge probability approximately $\log n/n$.

Theorem 3.10 ([23]). *Let $G \in G_{n,p}$, where $p = \frac{\log n + \omega}{n}$, $\omega \to \infty$, and $\omega = o(\log n)$. Then with high probability,*

$$rc(G) = (1 + o(1)) \max\left\{X_1, \frac{\log n}{\log \log n}\right\}$$

(with X_1 counting the degree-1 vertices).

CHAPTER 4: Randomly Perturbed Graphs

In this chapter, we explore the randomly perturbed graph model. This model was introduced in 2003 by Bohman et al. [8]. The model is particularly suited to situations that are largely deterministic but have an element of randomness.

As this model is relatively new, there is a lot of work left to be done to fill out its theory. We shall present some results to give the reader an understanding of the types of questions being asked and answered on the topic. We finish by showing a result related to rainbow connectivity in randomly perturbed graphs. This will lead us to the main result of this work (Theorem 5.2).

4.1 Definition and Notation

A *randomly perturbed graph* is a model created in the following way. We first fix a positive constant $d > 0$. We let $\mathcal{G}(n, d)$ be the set of graphs on n vertices with minimum degree $\delta \geq dn$. A graph $H = (V, E)$ is chosen arbitrarily from this set. To H we add a set R of m edges chosen uniformly at random from all of the $\binom{n}{2}$ possible edges, ignoring duplicate edges. This creates the graph $G_{H,m} = (V, E \cup R)$.

4.2 Results Related to Hamiltonicity

In their original paper on randomly perturbed graphs, Bohman et al. [8] explored how many edges need to be added to H to make G Hamiltonian with high probability. Recall that a graph is *Hamiltonian* if there exists a cycle where every vertex is visited only once.

It is known from Dirac's Theorem that if $d \geq 1/2$, then H is already Hamiltonian, so we can assume that $d < 1/2$.

Theorem 4.1 ([8]). *For a constant $0 < d < 1/2$ and for $H \in \mathcal{G}(n, d)$, let $G = G_{H,m} = (V, H \cup R)$, where R has size m and is chosen uniformly at random from $\binom{n}{2} \backslash E(H)$. Set $\theta = \ln d^{-1}$. Then:*

- *If $m \geq (30\theta + 13)n$, then $G_{H,m}$ is Hamiltonian with high probability; and*
- *For $d \leq 1/10$, there exist graphs $H \in \mathcal{G}(n, d)$ such that if $m < \theta n/3$, then $G_{H,m}$ is not Hamiltonian with high probability.*

The following result is also from [8]. Let $\alpha = \alpha(H)$ be the size of a largest independent set in H (recall that a set of vertices is *independent* if it spans no edge).

Theorem 4.2 ([8]). *Suppose that $H \in \mathcal{G}(n, d)$ and $1 \leq \alpha < d^2 n/2$ and so $d > n^{-1/2}$ (d need not be a constant). Let $G = (V(H), E(H) \cup R)$ with R comprised of m elements chosen uniformly at random from $\binom{n}{2} \backslash E(H)$. If*

$$\frac{md^3}{\theta} \to \infty,$$

then $G_{H,m}$ is Hamiltonian with high probability.

For a proof of Theorem 4.2, see [8].

4.3 More Results

The results in this section were proven by Bohman et al. [7] in 2016.

The first concerns whether G contains a complete graph K_r of order r.

Theorem 4.3 ([7]). *Let $r > r_0 \geq 2$ be integers. Let*

$$d \in \left(\frac{r_0 - 2}{r_0 - 1}, \frac{r_0 - 1}{r_0} \right]$$

be a fixed constant. Then:

- *If $H \in \mathcal{G}(n,d)$ and $m = \omega(n^{2-2/\lceil r/r_0 \rceil - 1})$, then $G_{H,m}$ contains a K_r with high probability; and*
- *There exists a graph $H_0 \in \mathcal{G}(n,d)$ such that if $m = o(n^{2-2/\lceil r/r_0 \rceil - 1})$, then $G_{H,m}$ does not contain a K_r with high probability.*

The next result considers the diameter of a randomly perturbed G.

Theorem 4.4 ([7]).

- *If $H \in \mathcal{G}(n,d)$ and $m = \omega(1)$, then diam$(G_{H,m}) \leq 5$ with high probability.*
- *Let $H \in \mathcal{G}(n,d)$. If $m = \frac{1-d}{d^2} \log n + \omega(1)$, then diam$(G_{H,m}) \leq 3$ with high probability.*
- *If $d < 1/2$, then there exists a graph $H_0 \in \mathcal{G}(n,d)$ such that if $m = \frac{\log n}{-2\log(1-2d)} - \omega(1)$, then diam$(G_{H_0,m}) \geq 5$ with high probability.*
- *Let $H \in \mathcal{G}(n,d)$. If $m = \frac{1-d}{d} n \log n + \omega(n)$, then diam$(G_{H,m}) \leq 2$ with high probability.*

- *There exists a graph* $H_0 \in \mathcal{G}(n, 1/2)$ *such that if* $m = \frac{1}{2}n\log n - \omega(n)$*, then* $diam(G_{H_0,m}) \geq 3$ *with high probability.*

From the above results, we see that (at least for $d < 1/2$), keeping $diam(G_{H,m}) \geq 5$ requires adding very few edges. As soon as we've randomly added $\log n$ edges (or any $\omega(1)$ function), we'll have $diam(G_{H,m}) \leq 5$. And in order to push $diam(G_{H,m}) \leq 2$, we should randomly add about $n \log n$ edges.

We now present a result on being k-connected. (Recall that a graph is *k-connected* if there doesn't exist a vertex cut of order at most $k - 1$.)

Theorem 4.5 ([7]).

- *Let* $H \in \mathcal{G}(n, d)$*. If* $k = O(1)$ *and* $m = \omega(1)$*, then* $G_{H,m}$ *is k-connected with high probability.*
- *If* $\omega(1) \leq k \leq d^2 n/32$ *and* $m = 640k/d^2$*, then* $G_{H,m}$ *is k-connected with high probability.*
- *If* $d < 1/2$*, then there exists a graph* $H_0 \in \mathcal{G}(n, d)$ *such that with high probability* $G_{H_0,m}$ *fails to be k-connected for all k such that*

$$m \leq \frac{k}{2}\left\lfloor \frac{n}{dn+1} \right\rfloor.$$

Results relating to spanning trees ([11], [12], [26], [27], [28]), Ramsey properties ([15] [16]), tilings ([6]), fixed subgraphs ([29]), and weighted graphs ([24]) have also been explored.

4.4 Results Related to Rainbow Connectivity

In 2018, Anastos and Frieze [4] explored the rainbow connected property for randomly perturbed graphs. We visit their results in this section. We first have to alter the $G_{H,m}$

model slightly so that its edges are colored. We color every edge of $G_{H,m}$ independently and uniformly at random with colors from $[r] = \{1, 2, \ldots, r\}$. The resulting graph we denote by

$$G^r_{H,m} = (V, E \cup R, c).$$

Here, $c\colon E \cup R \to [r]$ is a function that assigns to each edge its color.

Theorem 4.6 ([4])**.** *For rainbow connectivity, the following assertions hold:*

1. *If $r = 3$ and $m \geq 60d^{-2} \log n$, then $G^r_{H,m}$ is rainbow connected with high probability;*
2. *For $d \leq 1/10$, there exists a graph $H_0 \in \mathcal{G}(n, d)$ such that if $m \leq \log n/2$, then $G^4_{H_0,m}$ is not rainbow connected with high probability; and*
3. *If $r \geq 7$ and $m = \omega(1)$, then $G^r_{H,m}$ is rainbow connected with high probability.*

We present the proof given in [4].

Proof. We begin by proving assertion 1.

We want to show that for R large enough, the following holds true with high probability: for any pair of vertices $u, v \in V$, there exist many edges in R between $N(u)$ and $N(v)$. If we can show this, then it will follow that there is a rainbow path between u and v with high probability.

Let $R = \{r_1, \ldots, r_m\}$ and C be the event that $G^r_{H,m}$ is rainbow connected. Let $C_3(u, v)$ be the event that there exists a rainbow path u, u_0, v_0, v where the edges uu_0 and vv_0 are in $E(H)$ and the edge u_0v_0 is in R. Let $B(u, v)$ be the event that there exist fewer than $10 \log n$ such paths in $G^r_{H,m}$.

Given $r_1, \ldots, r_{i-1}$, either there exist $10 \log n$ such paths or r_i creates such a path with probability at least

$$\frac{dn(dn - 10\log n)}{\binom{n}{2}} \geq d^2.$$

It follows from the Chernoff Inequality (Corollary 1.4) that

$$\begin{aligned} Pr(B(u,v)) &\leq Pr(Binomial(60d^{-2}\log n, d^2) < 10\log n) \\ &\leq \exp\left\{-\frac{1}{2}\cdot\frac{1}{9}\cdot 60\log n\right\} \\ &\leq n^{-3}. \end{aligned}$$

We note that a path in $G^r_{H,m}$ of length 3 is rainbow with probability

$$1 \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}.$$

(This follows from the fact that we can choose any color for the first edge. We then have our choice of the remaining two colors for the second edge, and we are left with one color for the third edge.) It follows that

$$\begin{aligned} Pr(\text{not } C) &\leq \sum_{u,v\in V, u\neq v} Pr(\text{not } C_3(u,v)) \\ &\leq \sum_{u,v\in V, u\neq v} Pr(B(u,v)) + Pr(\text{not } C_3(u,v) \,|\, \text{not } B(u,v)) \\ &\leq \sum_{u,v\in V, u\neq v} \left(n^{-3} + \left(1 - \frac{2}{9}\right)^{10\log n}\right) \\ &\leq n^2(n^{-3} + n^{-20/9}) \\ &= o(1). \end{aligned}$$

Therefore, with high probability, $G^r_{H,m}$ is rainbow connected.

We proceed to assertion 2.

We will create a counterexample consisting of two disjoint copies of $G(\frac{n}{2}, p)$ with $p = 0.11$. We attempt to show that if the size of R is not sufficiently large, it will not cover every vertex in the neighborhoods of some vertices in either copy.

Let $d \leq 1/10$. We partition V into two sets, V_1, V_2, both of size $n/2$. Now H is generated by including in $E(H)$ every edge of $V_1 \times V_1$ or of $V_2 \times V_2$ independently with probability 0.22. As $(0.22)(n/2) = 0.11n$, the Chernoff Inequality (Corollary 1.4) implies that for all $v \in V$, the degree $d(v)$ of v satisfies $0.1n < d(v) < 0.12n$. This implies that $H \in \mathcal{G}(n, 0.1)$ with high probability.

In the case that there exist both $v \in V_1$ and $u \in V_2$ such that R contains no edge with an endpoint in each of

$$(\{v\} \cup N(v)) \times (\{u\} \cup N(v)),$$

then u and v are at distance at least 5 in $G^r_{H,m}$. Thus, if we only have 4 colors, it is impossible to rainbow color this path of length exceeding 4.

Note that with high probability, R covers sets $R_1 \subset V_1$ and $R_2 \subset V_2$ both of size at most $\log n$. Thus, a vertex in $V_1 \backslash R_1$ has at least one neighbor in R_1 independently with probability

$$1 - (1 - 0.22)^{|R_1|} \leq 1 - n^{-1/2}.$$

Therefore,

$$Pr(\exists v \in V_1 : (\{v\} \cup N(v)) \cap R = \emptyset) \geq 1 - (1 - n^{-1/2})^{0.5n} = 1 - o(1),$$

and similarly

$$Pr(\exists v \in V_2 : (\{v\} \cup N(v)) \cap R = \emptyset) = 1 - o(1).$$

Finally, with high probability, $G^4_{H,m}$ is not rainbow connected.

This leaves assertion 3, where we interrupt the Anastos-Frieze proof (that we've been following from [4]) because the main result (Theorem 5.2) of this dissertation is an improvement on that part. □

CHAPTER 5: Main Theorem

Recall the following result of Anastos and Frieze [4] (Theorem 4.6), which we restate here for convenience.

Theorem 5.1 ([4]). *For rainbow connectivity, the following assertions hold:*

1. *If $r = 3$ and $m \geq 60d^{-2} \log n$, then $G^r_{H,m}$ is rainbow connected w.h.p.;*
2. *For $d \leq 1/10$, there exists a graph $H_0 \in \mathcal{G}(n, d)$ such that if $m \leq \log n/2$, then $G^4_{H_0,m}$ is not rainbow connected w.h.p.; and*
3. *If $r \geq 7$ and $m = \omega(1)$, then $G^r_{H,m}$ is rainbow connected w.h.p.*

With regard to part 3, even though Anastos and Frieze proved it for $r = 7$, they conjectured that the true value was either 5 or 6. This conjecture was proved in [5], the authors of which include the author of this dissertation. We present that result. No attempt is made to optimize constants. Additionally, all non-integers are taken as ceilings.

Theorem 5.2. *If $r \geq 5$ and $m \geq 20d^{-2}$, then, with high probability, $G^r_{H,m}$ is rainbow connected.*

Proof. We begin by fixing $r = 5$ and $0 < d < 1$ and letting $m = m(d)$ be a large enough constant depending on d. Let H be a graph chosen randomly from $\mathcal{G}(n, d)$, that is, the set of graphs with vertex set $[n]$ and minimum degree $\delta \geq dn$. We randomly add to H a set R

of m edges chosen uniformly at random from $\binom{[n]}{2}$, where duplicate edges are ignored. We now randomly color the edges with 5 colors independently; the resulting graph is $G^5_{H,m}$.

Note that by ignoring duplicate edges we have a slightly weaker assumption than choosing R from $\binom{[n]}{2}\backslash E(H)$, but it will simplify our arguments.

We set $t = 100 \log n$ and select t sets $S_1, S_2, \ldots, S_t$, each made up of $k = k(d)$ vertices chosen uniformly at random from $V(H)$. Set $\bigcup S_i = S$.

Let C be the event that for every pair of vertices a, b in a particular S_i, there exists an edge in $G^5_{H,m}$ between $N(a)\backslash S$ and $N(b)\backslash S$. Note that $N(a)\backslash S$ and $N(b)\backslash S$ are not necessarily disjoint, so such an edge may be in their intersection.

Claim 1. $Pr(C) > 0.99$ *for* n *large enough.*

Proof. As $|S| = |\bigcup S_i| \leq kt = 100k \log n$, we can choose n large enough that $N(a)\backslash S$ and $N(b)\backslash S$ both have size at least $dn/2$. Thus, the probability that there exists no edge between $N(a)\backslash S$ and $N(b)\backslash S$ is at most

$$\begin{aligned} Pr(\text{not } C) &\leq \left(\frac{\binom{n}{2} - \binom{dn/2}{2}}{\binom{n}{2}} \right)^m \\ &\leq \left(1 - \frac{d^2}{4}\right)^m \\ &< \exp\left(-\frac{d^2 m}{4}\right) \\ &< 0.01 \end{aligned}$$

when $m \geq 20d^{-2}$. This proves the claim. $\square$

Claim 2. *For every pair of vertices* $u, v \in V(G^5_{H,m})$*, there is an index set* $I_{u,v} \subseteq [t]$ *such that* $|I_{u,v}| \geq 0.6t$*; also, for all* $i \in I_{u,v}$ *we have* $u, v \in N(S_i)$*.*

Proof. For distinct vertices u, v, observe the following:

$$\begin{aligned} Pr(u \notin N(S_i) \text{ or } v \notin N(S_i) &\leq Pr(u \notin N(S_i)) + Pr(v \notin N(S_i)) \\ &\leq 2Pr(u \notin N(S_i)) \\ &= 2Pr(S_i \cap N(u) = \emptyset) \\ &\leq \frac{\binom{(1-d)n}{k}}{\binom{n}{k}} \\ &\leq 2(1-d)^k \\ &< 2\exp(-dk) \\ &< 0.01 \end{aligned}$$

for $k > 6d^{-1}$. Thus, $Pr(u, v \in N(S_i)) \geq 0.99$.

For $i \in [t]$, let X_i be the indicator random variable for the event that $u, v \in N(S_i)$, and let $X = X_1 + X_2 + \cdots + X_t$ be the number of sets S_i such that $u, v \in N(S_i)$. Thus,

$$E[X] \geq 0.99t.$$

Since each S_i is chosen (uniformly at random) with replacement, the random variables X_i are independent. Therefore, by the Chernoff Inequality (Theorem 1.3), we obtain

$$Pr(X \leq (1-\alpha)E[X]) \leq \exp\left(-\frac{1}{2}\alpha^2 E[X]\right),$$

for $0 < \alpha < 1$. When $\alpha = 0.39$ this gives

$$\begin{aligned} Pr(X \leq 0.6t) &\leq Pr(X \leq 0.6039t) \\ &= Pr(X \leq (1-\alpha)(0.99t)) \\ &\leq \exp\left(-\frac{\alpha^2 E[X]}{2}\right) \\ &\leq \exp\left(-\frac{\alpha^2(0.99)(100\log n)}{2}\right) \\ &< n^{-7} \\ &< n^{-2}. \end{aligned}$$

Therefore, by the union bound (Theorem 1.1), u and v are contained in at least $0.6t$ sets of S_i with positive probability. This completes the proof of the claim. □

We continue by fixing u and v as arbitrary vertices in $G^5_{H,m}$. We want to estimate the probability that there is a rainbow uv-path with length at most 5. If uv is an edge, then we have such a path with probability 1, so we shall assume that uv is not an edge.

Let $I_{u,v}$ be the index set guaranteed by Claim 2. For $i \in I_{u,v}$, we proceed to estimate the probability that there is a rainbow uv-path of length at most 5 using vertices in S_i. We let $a \in S_i$ be a neighbor of u and let $b \in S_i$ be a neighbor of v. In the case that $a = b$, we have a uv-path of length 2, which is rainbow with probability

$$\frac{5 \cdot 4}{5^2} = \frac{4}{5}.$$

Therefore, we assume that $a \neq b$; the situation at this point is depicted in Figure 1.

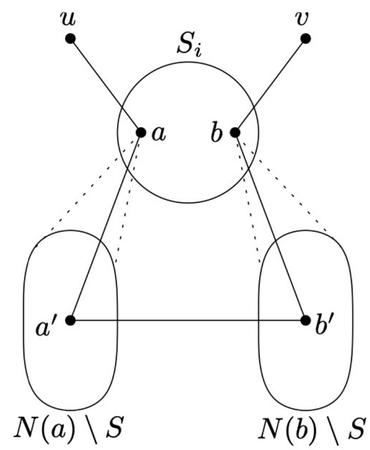

Figure 1: A u–v path of length 5.

It follows from Claim 1 that with probability at least 0.99, there exists an edge $a'b'$ between $N(a)\backslash S$ and $N(b)\backslash S$. If we condition on the existence of $a'b'$, then the probability that $uaa'b'bv$ is a rainbow path is $4!/5^4$ (having already used one color for $a'b'$ there are $4!$ ways to color the remaining edges of the path without repeating a color). It follows that the probability that there exists a rainbow uv-path using vertices u, v, a, b is at least

$$0.99 \cdot \frac{4!}{5^4}.$$

Therefore, we have

$$\begin{aligned}
Pr(\nexists \text{ rainbow } uv\text{-path of length } \leq 5) &\leq \left(1 - 0.99 \cdot \frac{4!}{5^4}\right)^{|I_{u,v}|} \\
&\leq \left(1 - 0.99 \cdot \frac{4!}{5^4}\right)^{0.6t} \\
&\leq \exp\left(-0.99 \cdot \frac{4!}{5^4} \cdot 0.6t\right) \\
&= \exp\left(-0.99 \cdot \frac{4!}{5^4} \cdot 0.6 \cdot 100 \log n\right) \\
&< n^{-2.25} \\
&= o(n^{-2}).
\end{aligned}$$

Applying the union bound (Theorem 1.1), we see that the probability that there is a pair u, v not connected by a rainbow path of length at most five is $o(1)$; i.e., $G^5_{H,m}$ is w.h.p. rainbow connected. □

BIBLIOGRAPHY

[1] Martin Aigner and Günter M Ziegler. *Proofs from the Book*. Springer, Berlin, 1999.

[2] Miklós Ajtai, Vašek Chvátal, Monroe M Newborn, and Endre Szemerédi. Crossing-free subgraphs. In *Theory and practice of combinatorics*, volume 60 of *North-Holland Math. Stud.*, pages 9–12. North-Holland, Amsterdam, 1982.

[3] Noga Alon and Joel H Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2008.

[4] Michael Anastos and Alan Frieze. How many randomly colored edges make a randomly colored dense graph rainbow Hamiltonian or rainbow connected? *Journal of Graph Theory*, 92(4):405–414, 2019.

[5] József Balogh, John Finlay, and Cory Palmer. Rainbow connectivity of randomly perturbed graphs. *arXiv preprint* `arXiv:2112.13277`, 5pp., 2021.

[6] József Balogh, Andrew Treglown, and Adam Zsolt Wagner. Tilings in randomly perturbed dense graphs. *Combinatorics, Probability and Computing*, 28(2):159–176, 2019.

[7] Tom Bohman, Alan Frieze, Michael Krivelevich, and Ryan Martin. Adding random edges to dense graphs. *Random Structures & Algorithms*, 24(2):105–117, 2004.

[8] Tom Bohman, Alan Frieze, and Ryan Martin. How many random edges make a dense graph hamiltonian? *Random Structures & Algorithms*, 22(1):33–42, 2003.

[9] Béla Bollobás. Random graphs. In *Modern graph theory*, chapter 7, pages 215–252. Springer, New York, 1998.

[10] Béla Bollobás and Andrew Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[11] Julia Böttcher, Jie Han, Yoshiharu Kohayakawa, Richard Montgomery, Olaf Parczyk, and Yury Person. Universality for bounded degree spanning trees in randomly perturbed graphs. *Random Structures & Algorithms*, 55(4):854–864, 2019.

[12] Julia Böttcher, Richard Montgomery, Olaf Parczyk, and Yury Person. Embedding spanning bounded degree graphs in randomly perturbed graphs. *Mathematika*, 66(2):422–447, 2020.

[13] Yair Caro, Arie Lev, Yehuda Roditty, Zsolt Tuza, and Raphael Yuster. On rainbow connection. *Electron. J. Combin.*, 15(1):Research paper 57, 13pp., 2008.

[14] Gary Chartrand, Garry L Johns, Kathleen A McKeon, and Ping Zhang. Rainbow connection in graphs. *Mathematica Bohemica*, 133(1):85–98, 2008.

[15] Shagnik Das, Patrick Morris, and Andrew Treglown. Vertex Ramsey properties of randomly perturbed graphs. *Random Structures & Algorithms*, 57(4):983–1006, 2020.

[16] Shagnik Das and Andrew Treglown. Ramsey properties of randomly perturbed graphs: cliques and cycles. *Combinatorics, Probability and Computing*, 29(6):830–867, 2020.

[17] Morris H DeGroot and Mark J Schervish. *Probability and statistics*. Pearson Education, Essex UK, fourth edition, 2012.

[18] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 2005.

[19] Paul Erdős. Graph theory and probability. *Canadian Journal of Mathematics*, 11:34–38, 1959.

[20] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[21] A B Ericksen. A matter of security. *Graduating Engineer & Computer Careers*, 24:28, 2007.

[22] Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016.

[23] Alan Frieze and Charalampos E Tsourakakis. Rainbow connectivity of sparse random graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 541–552. Springer, Boston, 2012.

[24] Alan M Frieze. The effect of adding randomly weighted edges. *SIAM Journal on Discrete Mathematics*, 35(2):1182–1200, 2021.

[25] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[26] Felix Joos and Jaehoon Kim. Spanning trees in randomly perturbed graphs. *Random Structures & Algorithms*, 56(1):169–219, 2020.

[27] Michael Krivelevich, Matthew Kwan, and Benny Sudakov. Bounded-degree spanning trees in randomly perturbed graphs. *SIAM Journal on Discrete Mathematics*, 31(1):155–171, 2017.

[28] Michael Krivelevich, Daniel Reichman, and Wojciech Samotij. Smoothed analysis on connected graphs. *SIAM Journal on Discrete Mathematics*, 29(3):1654–1669, 2015.

[29] Michael Krivelevich, Benny Sudakov, and Prasad Tetali. On smoothed analysis in dense graphs and formulas. *Random Structures & Algorithms*, 29(2):180–193, 2006.

[30] Xueliang Li, Yongtang Shi, and Yuefang Sun. Rainbow connections of graphs: a survey. *Graphs and combinatorics*, 29(1):1–38, 2013.

[31] Michael Mitzenmacher and Eli Upfal. *Probability and computing*. Cambridge University Press, Cambridge, second edition, 2017.

[32] Frank P Ramsey. On a problem of formal logic. In *Classic Papers in Combinatorics*, pages 1–24. Springer, Boston, 2009.

[33] Sheldon M Ross. *A first course in probability*. Prentice Hall, Upper Saddle River, NJ, sixth edition, 2002.