2016

# On Crown-free Set Families, Diffusion State Difference, and Non-uniform Hypergraphs

Edward Lawrence Boehnlein
*University of South Carolina*

# On Crown-free Set Families, Diffusion State Difference, and Non-uniform Hypergraphs

by

Edward Lawrence Boehnlein

Bachelor of Science
Arizona State University 2009
Master of Science
Miami University 2011

---

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Mathematics

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

Linyuan Lu, Major Professor

Éva Czabarka, Committee Member

Anton Schep, Committee Member

László Székely, Committee Member

Stephen Fenner, External Examiner

Cheryl L. Addy, Vice Provost and Dean of The Graduate School

# Acknowledgments

First, I want to thank my advisor, Dr. Lincoln Lu, to whom I am greatly indebted for all his help, ideas, suggestions, and patience with me over the last several years. I appreciate all the time you've taken to lead me through this program.

Brandon and Chris, I can't say enough to express my gratitude for your support over the past several years. Though life has grown more challenging since our glory days at ASU, our friendship has only gotten stronger, and you both continue to inspire me. Thank you, Julie, as well. I'm grateful to all of you for my home away from home.

Danielle, thank you for steering me clear of so many traps, both calculus and otherwise. To Dan and Emily, it's safe to say that I never would have gotten this far without having you in my corner.

There are a few other past and present USC peers that I must acknowledge, namely Ann, Bethany, Bill, Blake (or should I say Seymour?), Garner, Laura, Maria, Melanie (and on that note, Vash too for some critical input), Morgan, and Tippie.

Additionally, thanks to Jennifer, Connor, and Adam. Also, thank you to the Speakeasy crew, in particular to Jeremy, Travis, Dan, and Charlie.

And finally, but most importantly, I want to thank my family. To my mom and dad, I am eternally grateful for your love and support. I am only here because of the example you set, and the work ethic you infused in me. Jackie, my gratitude to you, well, it's as big as an ox! (or is it a house?) Also, I would be remiss not to thank Daisy and Sunny for their roles.

# Abstract

We present results in three different arenas of discrete mathematics.

Let $\mathrm{La}(n, H)$ denote the cardinality of the largest family on the Boolean lattice that does not contain $H$ as a subposet. Denote $\pi(H) := \lim_{n\to\infty} \frac{\mathrm{La}(n,H)}{\binom{n}{\lfloor n/2 \rfloor}}$. A crown $\mathcal{O}_{2k}$ for $k \geq 2$ is a poset on 2 levels whose Hasse diagram is a cycle. Griggs and Lu (2009) showed $\pi(\mathcal{O}_{4k}) = 1$ for $k \geq 2$. Lu (2014) proved $\pi(\mathcal{O}_{2k}) = 1$ for odd $k \geq 7$. We prove that the maximum size of a $\mathcal{O}_6$-free family, when restricted to the middle two levels of $\mathcal{B}_n$, is no greater than $1.56\binom{n}{\lfloor n/2 \rfloor}$. This section is joint work with Linyuan Lu.

The diffusion state distance (DSD) was introduced by Cao-Zhang-Park-Daniels-Crovella-Cowen-Hescott (2013) to capture functional similarity in protein-protein interaction networks. They proved the convergence of DSD for non-bipartite graphs. We extend the DSD to bipartite graphs using lazy-random walks and consider the general $L_q$-version of DSD. We discovered the connection between the DSD $L_q$-distance and Green's function, which was studied by Chung and Yau (2000). Based on that, we computed the DSD $L_q$-distance for Paths, Cycles, Hypercubes, as well as random graphs $G(n, p)$ and $G(w_1, \ldots, w_n)$. We also examined the DSD distances of two biological networks. This section is joint work with Peter Chin, Linyuan Lu, and Amit Sinha.

Motivated by the recent work on the Turán problems on non-uniform hypergraphs, we study when a fixed non-uniform hypergraph $H$ occurs in random hypergraphs with high probability. To be more precise, for a given set of positive integers $R := \{k_1, k_2, \ldots, k_r\}$ and probabilities $\mathbf{p} = (p_1, p_2, \ldots, p_r) \in [0, 1]^r$, let $G^R(n, \mathbf{p})$ be

the random hypergraph $G$ on $n$ vertices so that for $1 \leq i \leq r$ each $k_i$-subset of vertices appears as an edge of $G$ with probability $p_i$ independently. We ask for what probability vector $\mathbf{p}$, $G^R(n, \mathbf{p})$ almost surely contains a given subhypergraph $H$. Note that the Erdős-Rényi model $G(n, p)$ is the special case of $G^R(n, \mathbf{p})$ with $R = \{2\}$. The question of the threshold of the occurence of a fixed graph $H$ in $G(n, p)$ is well-studied in the literature. We generalize these results to non-uniform hypergraphs. Surprisingly, those $\mathbf{p}$ for which $G^R(n, \mathbf{p})$ almost surely contains $H$, form a convex region (depending on $H$) in a log-scale drawing. We also consider the associated extension problems. This section is joint work with Linyuan Lu.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

A graph $G = (V, E)$ is a set $V$ of vertices together with a set $E$ of pairs of distinct vertices called edges. A hypergraph $H = (V, E)$ is a set $V$ of vertices together with an edge set $E$ containing nonempty subsets of $V$. We say $H$ is $r$-uniform if every edge $e \in E(H)$ contains exactly $r$ distinct vertices. If $r = 2$, we just refer to it as a graph, or simple graph. In contrast, a non-uniform hypergraph has edges of varying vertex cardinalities.

## 1.1 Posets

A partially ordered set $G = (S, \leq)$, or *poset* for short, is a set $S$ with a partial ordering $\leq$. $G$ contains another poset $H = (S', \leq')$ as a *subposet* if there exists an injective map $f : S' \to S$ such that for all $u, v \in S'$, if $u \leq' v$ then $f(u) \leq f(v)$. Posets can be represented with a *Hasse diagram*, a graph whose vertices are the sets, and edges connect pairs of comparable sets, and we suppress any edges implied by transitivity.

Let $[n] := \{0, 1, \ldots, n-1\}$. The poset of concern here is the *Boolean lattice*, is defined $\mathcal{B}_n := (2^{[n]}, \subseteq)$ for $n \in \mathbb{N}$. Any family $\mathcal{F} \subseteq 2^{[n]}$ will be considered a subposet of $\mathcal{B}_n$. For any poset $H$, we say $\mathcal{F}$ is *$H$-free* if $H$ is not a subposet of $\mathcal{F}$. For any $n \in \mathbb{N}$, $\mathrm{La}(n, H)$ is defined to be the cardinality of the largest family $\mathcal{F} \subseteq \mathcal{B}_n$ that is $H$-free. We may consider $\mathrm{La}(n, H)$ as the poset analog of $\mathrm{ex}(n, H)$ in Turán theory, which stands for the maximum number of edges possible on a graph on $n$ vertices that does not contain $H$ as a subgraph. Figure 1.1 shows the Hasse diagram of the Boolean lattice for $n = 2$.

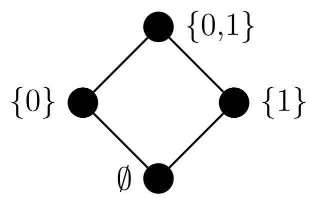


Figure 1.1 The Hasse diagram for $\mathcal{B}_2 = (2^{[2]}, \subseteq)$, also known as the diamond.

The history of determining $\mathrm{La}(n, H)$ dates back to Sperner [68] in 1928, when he proved that $\mathrm{La}(n, H) = \binom{n}{2}$ if $H$ is a pair of comparable elements ($P_2$). That is, the maximum size of an antichain, which means no two elements of $H$ are comparable, is $\binom{n}{2}$. In particular, the bound is attained by taking the middle row of $\mathcal{B}_n$, which is the row with the most elements. If $n$ is odd, then either of $\binom{n}{\lceil n/2 \rceil}$ and $\binom{n}{\lfloor n/2 \rfloor}$ will do. For convenience, the asymptotic value of $\mathrm{La}(n, H)$ is abbreviated as

$$\pi(H) := \lim_{n \to \infty} \frac{\mathrm{La}(n, H)}{\binom{n}{\lfloor n/2 \rfloor}}.$$

Griggs and Lu [44] conjectured that $\pi(H)$ exists and is an integer for all finite posets $H$. In Table 1.1, we summarize some of the posets for which $\pi(H)$ has already been determined.

Table 1.1 Known values of $\pi(H)$ in literature.

| $H$ | Name | $\pi(H)$ | Reference |
|---|---|---|---|
| $A_1 \subset \cdots \subset A_k$ | chain or $P_k$ | $k-1$ | [34] |
| $A \subset B_i$ for $i \in [r]$ | $r$-fork or $V_r$ | 1 | [28] |
| $A \subset C, D$ and $B \subset C$ | “N” | 1 | [42] |
| $A \cup B \subset C \cap D$ | butterfly | 2 | [29] |
| $A_1, \ldots, A_s \subset B_1, \ldots, B_t$ | $K_{s,t}$ | 2 | [50] |
| $A_1 \subset \cdots \subset A_k \subset B_1, \ldots B_s$ | $P_k(s)$ | $k$ | [28] |

A *chain* of *length* $k$ is a sequence of sets $A_1 \subset A_2 \subset \cdots \subset A_k$. The *height* of a poset is the length of its longest chain. Bukh [11] proved that for any poset $H$ whose Hasse diagram is a tree with length $k$, that $\pi(H) = k - 1$.

While $\mathrm{La}(n, H)$ has been determined for some simple posets, there is still much work to be done. Improving bounds on $\mathrm{La}(n, H)$ is still an active endeavor. The diamond (a copy of $\mathcal{B}_2$, see Figure 1.1) for instance, has been especially scrutinized in recent years. Table 1.2 summarizes the improvements on its bound in recent years.

Table 1.2 The upper bound for $\pi(\mathcal{B}_2)$, provided the limit exists.

| Authors | Bound on $\pi(\mathcal{B}_2)$ | Reference |
|---|---|---|
| Axenovich, Manske, and Martin | 2.283 | [3] |
| Griggs, Li, and Lu | 2.273 | [43] |
| Kramer, Martin, and Young | 2.25 | [57] |
| Grósz, Methuku, and Tompkins | 2.207 | [45] |

We can also consider problems involving induced posets. We say $G$ contains $H$ as an *induced subposet* if for any $u, v \in S'$, $u \leq' v$ if and only if $f(u) \leq f(v)$. Then, $\mathrm{La}^*(n, H)$ represents the cardinality of the largest family that does not contain $H$ as an induced subposet. Also, we can define

$$\pi^*(H) := \lim_{n\to\infty} \frac{\mathrm{La}^*(n, H)}{\binom{n}{\lfloor n/2 \rfloor}}.$$

Once can see that $\mathrm{La}(n, H) \leq \mathrm{La}^*(n, H)$ and $\pi(H) \leq \pi^*(H)$ for any given $H$. In general, determining $\mathrm{La}^*(n, H)$ is a much more difficult task than $\mathrm{La}(n, H)$ for a given poset $H$, and consequently there are not many known values of $\pi^*(H)$ found in literature. Carroll and Katona [13] showed $\pi^*(V_2) = 1$; Boehnlein and Jiang [9] proved that for $H$ whose Hasse diagram is a tree of height $k$, that $\pi^*(H) = k - 1$.

Of particular interest to us, however, are crowns. A *crown*, notated $\mathcal{O}_{2k}$ for $k \geq 2$, is a poset with height 2 whose Hasse diagram is a cycle. Often $\mathcal{O}_4$ is known as the butterfly. Figure 1.1 shows an example of a crown, namely the Hasse diagram of the 6-crown, which happens to be the middle two levels of $\mathcal{B}_3$.

Griggs and Lu [44] showed that for $k \geq 2$, $\pi(\mathcal{O}_{4k}) = 1$, and bounded $\pi(\mathcal{O}_{4k-2}) \leq 1.707$. Later, Lu [59] extended this, proving that for odd $k \geq 7$, $\pi(\mathcal{O}_{2k}) = 1$. This leaves only $\mathcal{O}_6$ and $\mathcal{O}_{10}$ as those crowns $\mathcal{O}_{2k}$ for which $\pi(\mathcal{O}_{2k})$ remains unknown. In chapter 2, we will present an upper bound for $\pi(\mathcal{O}_6)$, that carries an extra stipulation.

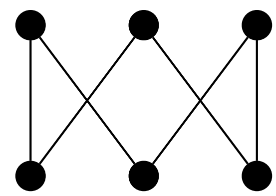

Figure 1.2 The 6-crown, $\mathcal{O}_6$, or the middle two rows of $\mathcal{B}_3$.

An important tool used in proving many of the above results is the *Lubell function*, which is defined

$$h_n(\mathcal{F}) := \sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}}.$$

In practice, the Lubell function is the expected number of elements from $\mathcal{F}$ that fall on a random full chain (one that contains an element of each cardinality, so having height $n+1$) in $\mathcal{B}_n$. The use of this function dates back to Lubell's [61] proof of Sperner's Theorem. Aside from arguments using the Lubell function, an increasingly popular tool is with flag algebras, the method we will discuss in detail in chapter 2, and use to prove our main result.

We exploit some linear algebra in the course of the flag algebra method. A real, symmetric $n$ by $n$ matrix $A$ is *positive semidefinite* if for any $\mathbf{x} \in \mathbb{R}^n$,

$$x^T A x \geq 0.$$

An equivalent condition is that the eigenvalues of $A$ are all non-negative. If $a, b \geq 0$, and symmetric $n$ by $n$ matrix $B$ is also positive semidefinite, then $aA + bB$ is positive semidefinite as well.

## 1.2 Random Walks on Graphs

A *walk* of length $k$ is an alternating sequence of vertices and edges $v_0, e_1, v_1, \ldots, e_k, v_k$. We permit vertices and edges to be repeated. Let $d_v$ denote the degree, or number of neighbors, of a vertex $v \in V(G)$. A *random walk* of length $k$ begins at a specified vertex $v_0$, and for $i \in [k]$ is traversed by randomly choosing a neighbor of $v_{i-1}$ (each having probability $1/d_{v_{i-1}}$ of being chosen, to be the next vertex $v_i$ in the sequence. Fixing $\alpha \in (0,1)$, an *$\alpha$-lazy random walk* is walk in which we remain at our present vertex $v$ with probability $\alpha$, or decide to step to one of its neighboring vertices with probability $(1-\alpha)/d_v$.

A *path* is a walk where no vertex is repeated. While there is a plethora of applications for walks and paths in graph theory, here we will only concern ourselves with their use in comparing the structures of graphs. A naïve way of comparing networks is by looking at the distribution of the lengths of the shortest paths between all vertex pairs. However, Cao et al [12] noted that in protein-protein interaction (PPI) networks, most networks have small diameters; that is, most vertices are relatively close to all other vertices. Their solution was to employ a new metric, called the diffusion state difference, which is based on random walks. A drawback is that their result excludes bipartite graphs. In chapter 3, we extend the diffusion state difference to cover all graphs using $\alpha$-lazy random walks instead.

In order extend the result of Cao [12], our methods centered around Green's function. For a graph $G$, let $A = (a_{ij})$ denote its adjacency matrix, where $a_{ij} := 1$ if $v_iv_j \in E(G)$, and $a_{ij} := 0$ otherwise. Define $D = (d_{ij})$ to be the diagonal matrix such that $d_{ii} = d_{v_i}$ and $d_{ij} := 0$ for $i \neq j$. Also, let $I$ denote the $n$ by $n$ identity matrix where $|V(G)| = n$. Then, the *discrete Laplacian* is defined

$$L := I - D^{-1}A.$$

But $L$ is not symmetric. So to achieve symmetry, the *normalized Laplacian* is defined

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

Note that $L$ and $\mathcal{L}$ have the same eigenvalues. For a full overview of spectral graph theory refer to [15]. Then, *Green's function* is defined to be the left inverse of $\mathcal{L}$.

Green's function first appeared in 1828, in a paper by George Green [41] using partial differential equations for applications to electricity and magnetism. William

Thomson (Lord Kelvin) [70, 71] revisited Green's functions years later, bringing them more attention. Later, Chung and Yau [22] thoroughly explored the application of Green's function to graphs.

## 1.3 Hypergraphs

The questions of the average behavior and the extremal behavior are frequently asked for many discrete objects. They are often the motivations for the growth of the discrete areas.

For a hypergraph $H = (V, E)$, define the *edge type* of $H$, $R(H) := \{|F| : F \in E(H)\}$. For a fixed set $R$ of positive integers, we say a hypergraph $H$ is an $R$-graph if $R(H) \subseteq R$. We often denote by $H_n^R$, an $R$-graph on $n$ vertices.

The extremal problems of non-uniform hypergraphs are considered by Johnston and Lu [49]. They generalized several important properties of the Turán density to non-uniform hypergraphs: supersaturation, blow-up, suspension, etc. For a given $R$-graph $H$, the Turán density $\pi(H)$ is the smallest number $\alpha$ such that for any $\epsilon > 0$ and any $R$-graph $G$ on $n$ vertices with Lubell value $h_n(G) := \sum_{F\in E(G)} \frac{1}{\binom{n}{|F|}}$ of at least $\alpha + \epsilon$ must contain a copy of $H$ for sufficiently large $n$. This definition generalizes the classical definition of Turán density of $k$-uniform hypergraphs. For $R = \{2\}$, the graph case, Erdős-Stone-Simonovits proved $\pi(G) = 1 - \frac{1}{\chi(G)-1}$ for any graph $G$ with chromatic number $\chi(G) \geq 3$. Johnston and Lu generalized Erdős-Stone-Simonovits' theorem to $\{1, 2\}$-graphs and determined the value of $\pi(H)$ for all $\{1, 2\}$-graph $H$. There are a few uniform hypergraphs whose Turán density has been determined: the Fano plane [40, 52], expanded triangles [53], 3-books, 4-books [39], $F_5$ [37], extended complete graphs [62], etc. In particular, Baber and Talbot [4] recently found the Turán density of many 3-uniform hypergraphs using flag algebra methods. However, no single value of $\pi(K_k^r)$ is known for any complete $r$-graph on $k$-vertices with $k > r \geq 3$. Turán conjectured [72] that $\pi(K_4^3) = 5/9$. Erdős [33] offered \$500 for determining any $\pi(K_k^r)$ with $k > r \geq 3$ and \$1000 for answering it for all $k$ and $r$. The upper bounds for $\pi(K_4^3)$ have been sequentially improved: 0.6213 (de Caen [30]), 0.5936 (Chung-Lu [21]), 0.56167 (Razborov [63], using the flag algebra method.) For a more complete survey of methods and results on uniform hypergraphs see Peter Keevash's survey paper [51].

The question of average behavior asks when a fixed hypegraph $H$ will occur in a random hypergraph. For any fixed set $R$ of positive integers, and any probability vector $\mathbf{p} \in [0,1]^R$, we define the random hypergraph $G^R(n, \mathbf{p}) = (V, E)$ with $V := [n]$, the set of first $n$ positive integers; and for $r \in R$, an $r$-set $F \in \binom{V}{r}$ belongs to $E$ independently with probabilty $p_r$. Additionally, we write the probability that $G^R(n, \mathbf{p})$ satisfies a certain property $A$ as $\Pr[G^R(n, \mathbf{p}) \models A]$. For $R = \{2\}$, this definition is precisely the Erdős-Rényi model $G^R(n, p)$ of the classical random graphs, originally described in [35]. In recent years, the same concept has been generalized to apply to uniform hypergraphs with $R = \{r\}$, such as in [54], [24], and [31]. A graph $H$ on $n$ vertices with $e$ edges is called *balanced* if for every subgraph $H' \subset H$, then $\rho(H') \leq \rho(H)$, where $\rho(H) = \frac{e}{n}$. Even stronger, $H$ is called *strictly balanced* if $\rho(H') < \rho(H)$ for all proper subgraphs $H' \subsetneq H$. Given a fixed graph $H$, the threshold of the occurence of a strictly balanced graph $H$ in $G^{\{2\}}(n, p)$ is given by $p = cn^{-v/e}$ by Alon and Spencer [1]. In this case, for any $c > 0$,

$$\lim_{n \to \infty} \Pr[G^{\{2\}}(n, p) \models A] = \exp\left(-c^e / |\mathrm{Aut}(H)|\right).$$

# Chapter 2

# 6-Crowns

## 2.1 Flag Preliminaries

Recall that $h_n(\mathcal{F})$ refers to the Lubell function. Let $\mathcal{H}$ denote the family of all 6-crown-free posets consisting only of sets from the middle two rows of $\mathcal{B}_5$. Define $p_H$ to be the probability that a random subset of the middle two rows of $\mathcal{B}_5$ is isomporphic to $H$. One can observe that:

$$\begin{aligned} h_n(\mathcal{F}) &= \sum_{H \in \mathcal{H}} p_H h_n(H) && (2.1) \\ &\leq \max_{H \in \mathcal{H}} h_n(H). && (2.2) \end{aligned}$$

However, under specific circumstances, a careful selection of constants $c_H$ for each set $H \in \mathcal{H}$, we can potentially improve this bound as follows:

**Proposition 2.1.**

$$\begin{aligned} h_n(\mathcal{F}) &\leq \sum_{H \in \mathcal{H}} (p_H h_n(H) + c_H p_H) && (2.3) \\ &= \sum_{H \in \mathcal{H}} (h_n(H) + c_H) p_H && (2.4) \\ &\leq \max_{H \in \mathcal{H}} (h_n(H) + c_H), && (2.5) \end{aligned}$$

where the $c_H$ come from entries in a positive semidefinite matrix.

The selection of the values of $c_H$ are far from arbitrary. We will rigorously justify Proposition 2.1 with a stronger claim in Proposition 2.3 in the next section. For now, we note that the bound in Proposition 2.1 leads to our main result.

**Theorem 2.2.** *For a 6-crown-free set family $\mathcal{F} \subset \mathcal{B}_n$, whose sets are restricted to the middle two rows of $\mathcal{B}_n$,*

$$|\mathcal{F}| \leq 1.56\binom{n}{\lfloor n/2 \rfloor}.$$

Properly choosing the values of $c_H$ to use is the most difficult task here. To do so, we employ the method of flag algebras, introduced in the next section.

## 2.2 Flag Algebras

The chief technique we use in the proof of Theorem 2.2 is that of flag algebras. Flag algebras are a strategy that is presently in vogue in discrete mathematics, particularly graph theory, that boils down to shrewd application of the Cauchy-Schwarz inequality. The notoriety of flag algebras can be attributed to Razborov's [64] seminal paper on the method, in which he demonstrated their use with an abundance of applications, in particular improving numerous results in Turán theory. Since then, flag algebras have become increasingly popular appearing in papers from Baber and Talbot [5] and Keevash [51] on hypergraph Turán theory, Balogh et al [7] on hypercubes with forbidden subgraphs, and Kramer, Martin, and Young [57] on diamond-free posets, just to name a few.

We present the flag algebra strategy in the context of a more general, graph theoretic version of Proposition 2.1. First we introduce the appropriate, analogous notation. We follow the setup of Baber and Talbot [5], who used flag algebras to tackle hypergraphs. For a graph $G$ on $n$ vertices, the *density* of a graph $G$ is given by

$$d(G) := \frac{|E(G)|}{\binom{n}{2}}.$$

This can be adapted to a $k$-regular hypergraph by amending the denominator to $\binom{n}{k}$, but we stick to the classic graph for simplicity. In the poset case we use the Lubell function instead. A graph $G$ is said to be *$F$-free* if there is no subgraph of

$G$ that is isomporphic of $F$. Let $\mathcal{H}$ be the family of graphs on $\ell$ vertices that are $F$-free. Then for any graph $H \in \mathcal{H}$, and an $F$-free graph $G$, define $p(H;G)$ to be the probability that the subgraph induced by a randomly chosen subset of $\ell$ vertices from $G$ is isomorphic to $H$. (Note that we used simply $p_H$ earlier instead of $p(H;G)$ because the size of the host graph was fixed.)

**Proposition 2.3** (Baber and Talbot, [5]).

$$d(G) \leq \sum_{H \in \mathcal{H}} (d(H) + c_H) p(H;G) + o(1).$$

In particular,

$$d(G) \leq \max_{H \in \mathcal{H}} (d(H) + c_H).$$

*Proof.* We reproduce their proof for completeness, and in particular, because it highlights the flag algebra setup. From the definition of density, observe that

$$d(G) = \sum_{H \in \mathcal{H}} d(H) p(H;G). \tag{2.6}$$

And so,

$$d(G) \leq \max_{H \in \mathcal{H}} d(H). \tag{2.7}$$

Note that (2.6) and (2.7) are the analogs of (2.1) and (2.2), respectively. However these bounds can be improved if we exploit the intersection of two subgraphs $H, H' \in \mathcal{H}$, whereas the latter observation only captures information on disjoint subgraphs.

Let $\theta : [\ell] \to V(H)$ be a bijection. Then we define a *flag* to be an ordered pair $F_1 = (H, \theta)$. Further, if $\sigma$ is a flag, we say $F_1$ is a $\sigma$-flag if $F_1$ is isomorphic to $\sigma$. Next we denote $\mathcal{F}_m^\sigma$ to be the set of all $\sigma$-flags, up to isomorphism, of order $m \leq (\ell + |\sigma|)/2$. The upper bound on $m$ is necessary for attaining subgraphs that intersect nontrivially, namely in $|\sigma|$ vertices. Next, denote $\Theta$ as the set of all injective maps $\theta : [|\sigma|] \to V(G)$. Then if $F \in \mathcal{F}_m^\sigma$ and $\theta \in \Theta$, assign $p(F, \theta; G)$ to the probability that a random subset $V'$ with $|V'| = m$ and $\mathrm{im}(\theta) \subseteq V' \subseteq V(G)$ induces a $\sigma$-flag isomorphic to $F$.

Now, if $F_a, F_b \in \mathcal{F}_m^\sigma$ and $\theta \in \Theta$, we set $p(F_a, F_b, \theta; G)$ to be the probability that given two randomly chosen $m$-subsets of $V(G)$, namely $V'_a$ and $V'_b$, such that $\mathrm{im}(\theta) \subseteq V'_a$ and $V'_a \cap V'_b = \mathrm{im}(\theta)$, then the flags $(G[V'_a], \theta)$ and $G([V'_b], \theta)$ are isomorphic to $F_a$ and $F_b$, respectively. While it would be rather convenient if

$$p(F_a, \theta; G)p(F_b, \theta; G) = p(F_a, F_b, \theta; G), \tag{2.8}$$

this is not always true because the vertices are chosen with replacement on the left side but without replacement on the right side of (2.8). However, this is not an issue if we assume $V(G)$ to be sufficiently large, as Baber and Talbot proved in a a special case of a lemma from Razborov [64]:

**Lemma 2.4** (Baber and Talbot, [5])**.** *For any $F_a, F_b \in \mathcal{F}_m^\sigma$ and $\theta \in \Theta$,*

$$p(F_a, \theta; G)p(F_b, \theta; G) = p(F_a, F_b, \theta; G) + o(1).$$

*In particular, the $o(1)$ term goes to zero as $|V(G)| \to \infty$.*

Consequently, if $\theta \in \Theta$ is chosen uniformly at random,

$$\mathbb{E}_{\theta\in\Theta}[p(F_a, \theta; G)p(F_b, \theta; G)] = \mathbb{E}_{\theta\in\Theta}[p(F_a, F_b, \theta; G)] + o(1).$$

However, that if we let $\Theta_H$ denote the set of all injective functions $\theta : [|\sigma|] \to V(H)$, then

$$\mathbb{E}_{\theta\in\Theta}[p(F_a, F_b, \theta; G)] = \sum_{H\in\mathcal{H}} \mathbb{E}_{\theta\in\Theta_H}[p(F_a, F_b, \theta; G)]p(H; G).$$

Next suppose $Q = (Q_{ab})$ is a positive semi-definite matrix of dimension $|\mathcal{F}_m^\sigma|$. Define the vector $\mathbf{p}_\theta := (p(F, \theta; G) : F \in \mathcal{F}_m^\sigma)$. And so,

$$\mathbb{E}_{\theta\in\Theta}[\mathbf{p}_\theta^T Q \mathbf{p}_\theta] = \sum_{H\in\mathcal{H}} c_H(\sigma, m, Q)p(H; G) + o(1),$$

where

$$c_H(\sigma, m, Q) = \sum_{F_a, F_b\in\mathcal{F}_m^\sigma} q_{ab}\mathbb{E}_{\theta\in\Theta_H}[p(F_a, F_b, \theta; H)].$$

Next, suppose $\sigma_i$ is a type, $m_i \leq \frac{1}{2}(\ell + |\sigma_i|)$, and $Q_i$ is a positive semidefinite matrix of dimension $|\mathcal{F}^{\sigma_i}_{m_i}|$. Fix $t \in \mathbb{Z}$ to be the number of choices for $(\sigma_i, m_i, Q_i)$. Then, for any $H \in \mathcal{H}$ define:

$$c_H = \sum_{i=1}^{t} c_H(\sigma_i, m_i, Q_i).$$

But exploiting that each matrix $Q_i$ is positive semidefinite, we obtain:

$$\sum_{H \in \mathcal{H}} p(H; G) + o(1) \geq 0.$$

And so,

$$d(G) \leq \sum_{H \in \mathcal{H}} (d(H) + c_H) p(H; G) + o(1).$$

But since $\sum_{H \in \mathcal{H}} p(H; G) = 1$, then it follows

$$d(G) \leq \max_{H \in \mathcal{H}} (d(H) + c_H).$$

□

Returning to the poset setting, by changing density to Lubell function, Proposition 2.3 implies Proposition 2.1.

## 2.3 Proof of Theorem 2.2

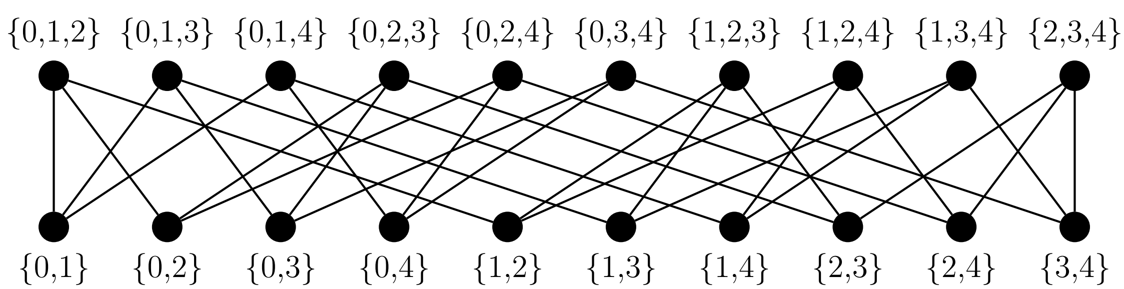


Figure 2.1 The middle two levels of $\mathcal{B}_5$.

We identify $\mathcal{H}$ as the set of all posets that do not contain a 6-crown, and whose members are restricted to the middle two rows of $\mathcal{B}_5$ (shown in Figure 2.1). For convenience, we may refer to any $H \in \mathcal{H}$ by the graph induced by its Hasse diagram.

To determine the flags, using the notation established with Proposition 2.3, we set $G$ as the middle two levels of $\mathcal{B}_5$, and $\ell = 10$. There are 8400 such graphs, up to isomorphism, found with computer assistance. We denote these graphs as $H_0, H_1, \ldots, H_{8399}$, sorted in increasing order of their Lubell function. A few of these graphs appear in Figures 2.2 and 2.3.

**Remark 2.5.** Among these 8400 graphs, the highest Lubell value is 1.6, which is attained only by $H_{8399}$. Hence, for a 6-crown free family $\mathcal{F}$ on the middle two rows, $|\mathcal{F}| \leq 1.6\binom{n}{\lfloor n/2 \rfloor}$.

Now the families $\{\{0,1\}\}$ and $\{\{1,2\}\}$ are isomorphic, but neither of these are isomorphic to $\{\{0,1,2\}\}$, as we do care to differentiate between the upper and lower rows. The difference will also be exploited in the use of the duals of the graphs. The *dual* of $H \in \mathcal{H}$ is an inverted copy of $H$. That is, rather than considering $H \in (2^{[n]}, \subseteq)$, we find the dual of $H$ by instead considering $H$ a member of $(2^{[n]}, \supseteq)$. In Figure 2.2, graphs $H_{7202}$ is the dual of $H_{7968}$, and vice versa. Also, $H_{8399}$ is self dual.

Now we can simply use $p_H$ rather than $p(H; G)$ since $G$ is fixed. Set also $m = 4$, so the flags are composed of copies of $\mathcal{B}_2$. Finally, $|\sigma| = 1$, as the flags only overlap in one element, namely $\{0\}$. The result is a total of 6 flags, shown in Figure 2.4. Vertices that are included are shaded black, hollow vertices are excluded.

Using computer assistance, a multiplication table for all possible pairs of flags was produced. As an example, consider the product of $p_1$ and $p_2$. The multiplication process is shown in Figure 2.5. The element 0 is fixed in both flags, but we consider $p_2$ on the elements $0, 3, 4$ rather than $0, 1, 2$. Their product is the linear combination of all subgraphs of the middle two levels of $\mathcal{B}_5$, that contain the sets $\{0,1\}$ (from $p_1$) and $\{0,3,4\}$ (from $p_2$) but not $\{0,2\}$, $\{0,1,2\}$, $\{0,3\}$, nor $\{0,4\}$. Note that the element $\{0\}$ can be disregarded at this point, as it does not lie in the middle 2 rows of $\mathcal{B}_5$. The Hasse diagram for this product is given in Figure 2.5, where the 14 elements

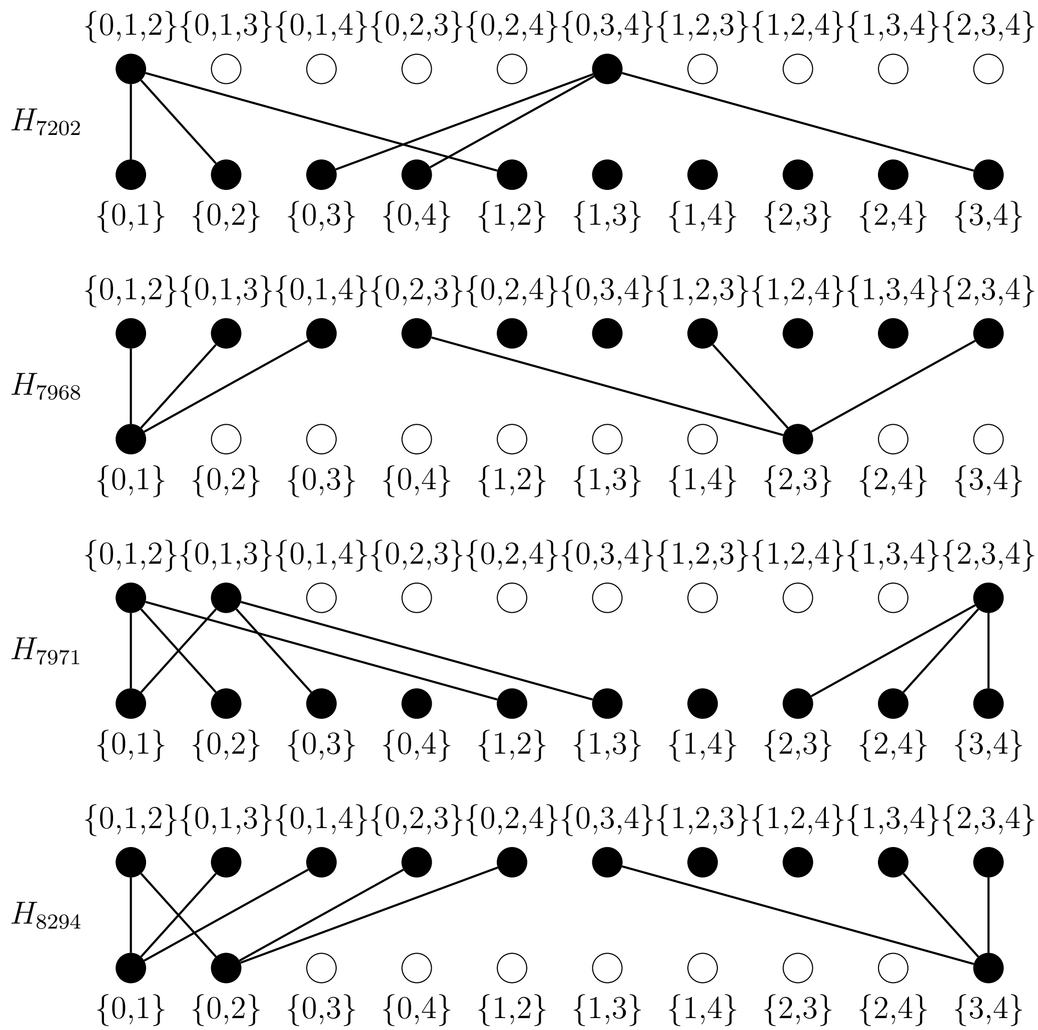


Figure 2.2 Some examples of the graphs $H_i$

that may or may not be included are shaded gray. Not counting isomorphisms, there are $2^{14}$ such graphs with this prescribed property. However, even before considering isomorphisms, note that not all of these $2^{14}$ are realizable, as many will contain a 6-crown. However, the coefficients in this linear combination are the number of copies of each graph from $\mathcal{H}$, counting isomorphisms.

The resulting multiplication table has its columns indexed by the graphs, and its rows organized by pairwise products of flags. An abbreviated version of the table is provided in Table 2.1, transposed for formatting constraints. After determining the products of all pairs of flags (which are commutative, as noted earlier), we are able

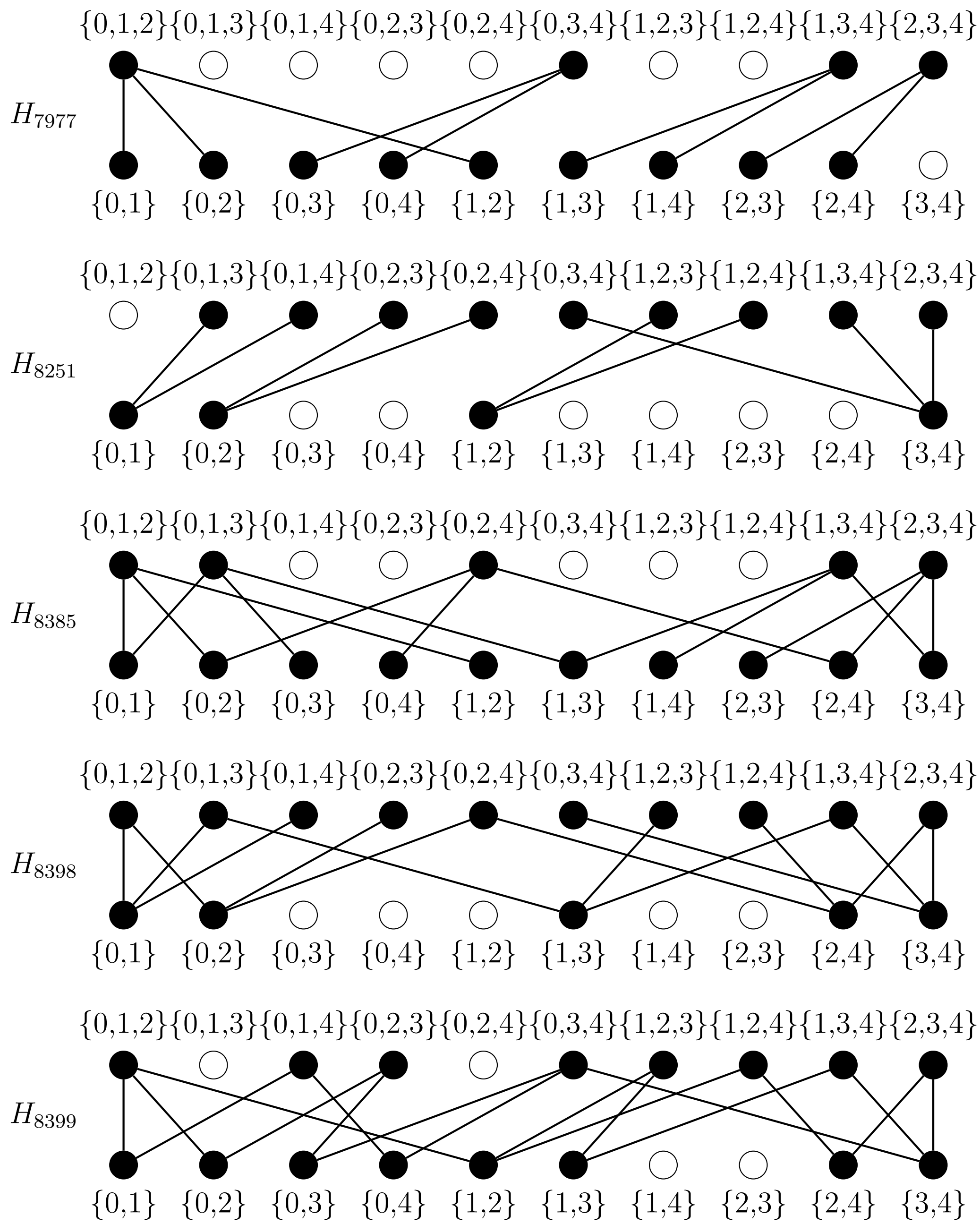


Figure 2.3  More examples of the graphs $H_i$

to set up a semidefinite program problem. The program has the form:

$$\text{minimize } v$$

$$\text{subject to } v \geq h_n(H_i) + c_{H_i} \text{ for all } i \in [8400]$$

$$v \in \mathbb{R},\ Q \text{ is semidefinite.}$$

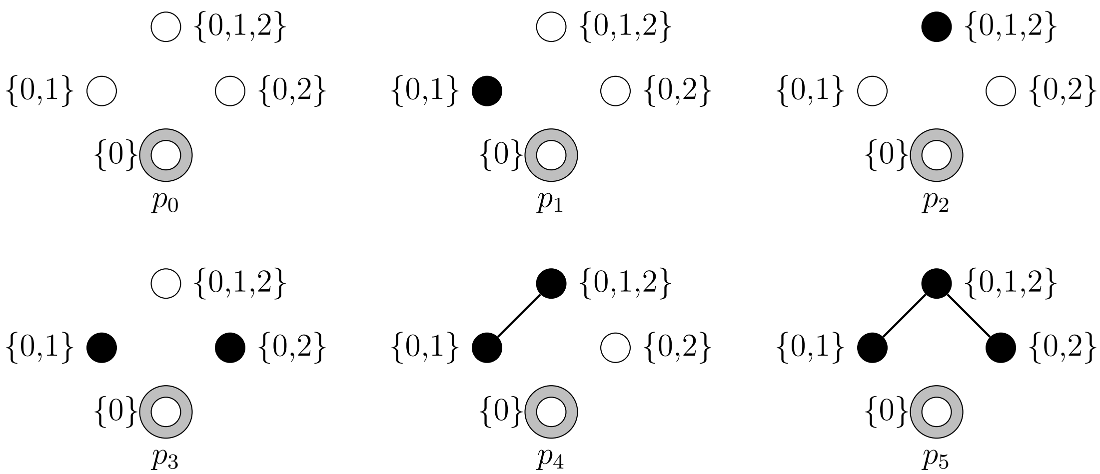


Figure 2.4 The flags.

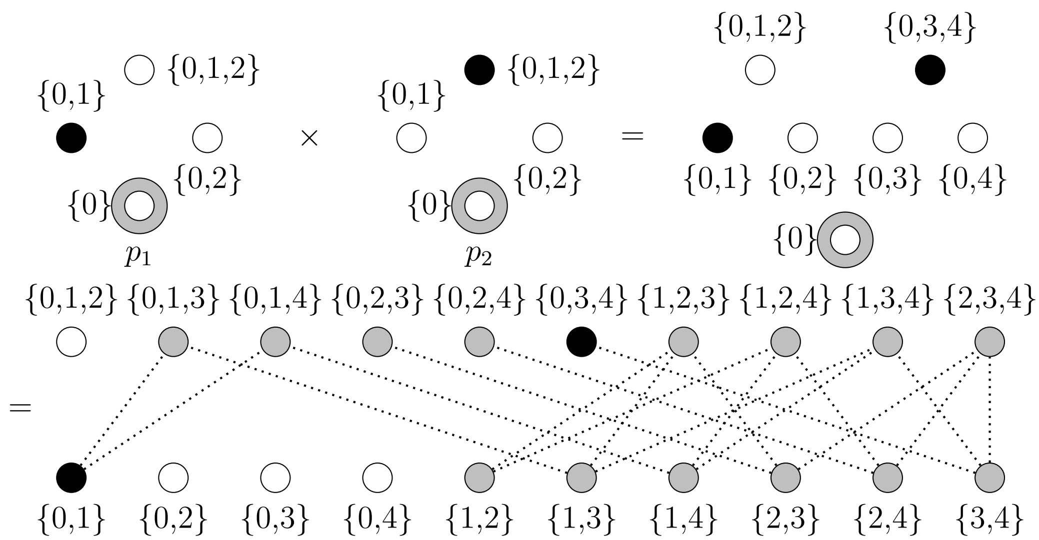


Figure 2.5 The product of flags $p_1$ and $p_2$.

There are 6 flags, so a 6 by 6 semidefinite matrix will be computed, having the

Table 2.1 Number of copies of $H_k$ obtained from combining flags $p_i$ and $p_j$

| | $p_2$, $p_2$ | $p_2$, $p_3$ | $p_2$, $p_4$ | $p_2$, $p_5$ | $p_3$, $p_3$ | $p_3$, $p_4$ | $p_3$, $p_5$ | $p_4$, $p_4$ | $p_4$, $p_5$ | $p_5$, $p_5$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $H_{4325}$ | 120 | 0 | 0 | 0 | 120 | 0 | 0 | 0 | 0 | 0 |
| $H_{5924}$ | 120 | 0 | 0 | 0 | 120 | 0 | 0 | 0 | 0 | 0 |
| $H_{7202}$ | 24 | 0 | 24 | 0 | 80 | 0 | 16 | 0 | 0 | 8 |
| $H_{7968}$ | 24 | 0 | 24 | 0 | 80 | 0 | 16 | 0 | 0 | 8 |
| $H_{7971}$ | 0 | 0 | 24 | 4 | 64 | 0 | 20 | 4 | 0 | 16 |
| $H_{7977}$ | 0 | 12 | 12 | 0 | 48 | 12 | 0 | 12 | 0 | 24 |
| $H_{8251}$ | 0 | 12 | 12 | 0 | 48 | 12 | 0 | 12 | 0 | 24 |
| $H_{8294}$ | 0 | 0 | 24 | 4 | 64 | 0 | 20 | 4 | 0 | 16 |
| $H_{8385}$ | 0 | 0 | 0 | 20 | 40 | 0 | 20 | 20 | 0 | 40 |
| $H_{8398}$ | 0 | 0 | 0 | 20 | 40 | 0 | 20 | 20 | 0 | 40 |
| $H_{8399}$ | 0 | 0 | 0 | 0 | 16 | 16 | 0 | 0 | 32 | 32 |

form

$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} & q_{15} & q_{16} \\ q_{21} & q_{22} & q_{23} & q_{24} & q_{25} & q_{26} \\ q_{31} & q_{32} & q_{33} & q_{34} & q_{15} & q_{36} \\ q_{41} & q_{42} & q_{43} & q_{44} & q_{45} & q_{46} \\ q_{51} & q_{52} & q_{53} & q_{54} & q_{55} & q_{56} \\ q_{61} & q_{62} & q_{63} & q_{64} & q_{65} & q_{66} \end{pmatrix},$$

which is symmetric ($q_{ij} = q_{ji}$) and positive semidefinite. Now $c_{H_i}$ is computed by multiplying the vector

$$(q_{11}, 2q_{12}, 2q_{13}, \ldots, q_{22}, 2q_{23}, \ldots, 2q_{56}, q_{66})$$

with the corresponding column (transposed to row here) $H_i$ in the multiplication Table 2.1. So the result has the form

$$c_{H_i} = \sum_{1 \le i,j \le 6} \alpha_{ij} q_{ij},$$

for some $\alpha_{ij} \in \mathbb{R}$.

The program was then solved using the CSDP solver [10].

Now from Proposition 2.1,

$$h_n(\mathcal{F}) \leq \max_{H \in \mathcal{H}} (h_n(H) + c_H). \tag{2.9}$$

In order to drive the bound lower, the entire process can be repeated using the duals of the graphs instead. Figure 2.6 shows how the flags for the dual case are simply those in Figure 2.4 turned upside down, rooted in the $\binom{5}{4}$ level rather than the $\binom{5}{1}$ level.

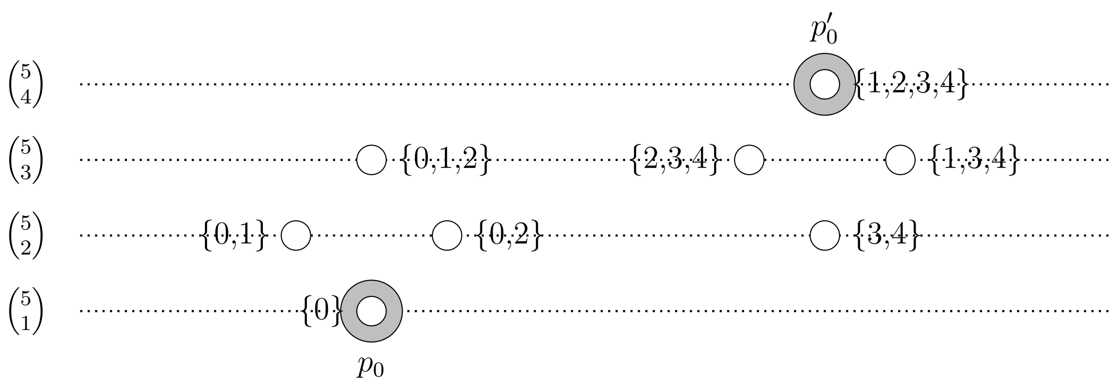


Figure 2.6 A comparison of the flags for the original (left) and dual (right) cases

That is, we can consider the flags being turned upside down, being rooted in the $\binom{5}{4}$ level of $\mathcal{B}_5$ rather than the $\binom{5}{1}$ level. As a result, we obtain a new inequality,

$$h_n(\mathcal{F}) \leq \max_{H \in \mathcal{H}} (h_n(H) + c'_H), \tag{2.10}$$

where $c'_H$. Next, we can average the constants in (2.9) and (2.10) to obtain another bound,

$$h_n(\mathcal{F}) \leq \max_{H \in \mathcal{H}} \left( h_n(H) + \frac{c_H + c'_H}{2} \right). \tag{2.11}$$

The resulting matrix, after rounding the entries to a reasonably close rational number, was

$$Q \approx Q_0 = \frac{1}{2400}\begin{pmatrix} 12 & 12 & 4 & 0 & 6 & -5 \\ 12 & 40 & 5 & 0 & 9 & -7 \\ 4 & 5 & 9 & 0 & 3 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 9 & 3 & 0 & 5 & -4 \\ -5 & -7 & -6 & 0 & -4 & 5 \end{pmatrix}.$$

With computer assistance, $Q_0$ is confirmed to be positive semidefinite. A summary of the results is given in Table 2.2, including only the graphs with a Lubell value of 1.5 or 1.6.

From (2.11), the bound of $|\mathcal{F}| \leq 1.56\binom{n}{\lfloor n/2 \rfloor}$ is attained.

Table 2.2 Relevant results for graphs $H$ with large Lubell value.

| $H$ | $h_n(H)$ | $\frac{c_H + c'_H}{2}$ | $h_n(H) + \frac{c_H + c'_H}{2}$ |
|---|---|---|---|
| $H_{8385}$ | 1.5 | 0.025 | 1.525 |
| $H_{8386}$ | 1.5 | 0.01167 | 1.51167 |
| $H_{8387}$ | 1.5 | -0.02333 | 1.47667 |
| $H_{8388}$ | 1.5 | 0.01 | 1.51 |
| $H_{8389}$ | 1.5 | -0.08 | 1.42 |
| $H_{8390}$ | 1.5 | 0 | 1.5 |
| $H_{8391}$ | 1.5 | -0.05667 | 1.44333 |
| $H_{8392}$ | 1.5 | 0 | 1.5 |
| $H_{8393}$ | 1.5 | -0.05667 | 1.44333 |
| $H_{8394}$ | 1.5 | -0.08 | 1.42 |
| $H_{8395}$ | 1.5 | 0.01 | 1.51 |
| $H_{8396}$ | 1.5 | -0.02333 | 1.47667 |
| $H_{8397}$ | 1.5 | 0.01167 | 1.51167 |
| $H_{8398}$ | 1.5 | 0.025 | 1.525 |
| $H_{8399}$ | 1.6 | -0.04 | **1.56** |

**Remark 2.6.** After finishing this proof, we discovered that a better bound has already been proven. Restricting $\mathcal{F}$ to 2 rows, Kramer [56] showed $|\mathcal{F}| \leq (2\sqrt{3} - 2)\binom{n}{\lfloor n/2 \rfloor}$ without the use of flag algebras.

## 2.4 Future Work

Maintaining the restriction to the middle 2 levels of $\mathcal{B}_n$, it seems plausible that the bound can be pushed further down. This could potentially be accomplished by using larger flags (namely $\mathcal{B}_3$), though this would be more computationally intensive. Short of gettting a bound for $\pi(\mathcal{O}_6)$ outright without any restrictions, it may be worthwhile to attempt a bound assuming a restriction to the middle 3 levels. Flag algebras seem to be the most promising method to use.

It would also be interesting to see whether flag algebras could give us any nontrivial bounds on $\mathcal{O}_{10}$, even if it requires another restriction to the middle 2 levels. However, if we employed a similar setup to that presented here, the flags would consist of the middle 2 levels of $\mathcal{B}_3$, of which there are 20, up to isomorphism. Then the product of any 2 flags yields $2^{29}$ possible graphs, before accounting for isomorphisms or checking for copies of $\mathcal{O}_6$. Thus, this appears to be more computationally daunting.

# Chapter 3

# Computing Diffusion State Distance using Green's Function and Heat Kernel on Graphs[1]

## 3.1 Introduction

Recently, the diffusion state distance (DSD, for short) was introduced in [12] to capture functional similarity in protein-protein interaction (PPI) networks. The diffusion state distance is much more effective than the classical shortest-path distance for the problem of transferring functional labels across nodes in PPI networks, based on evidence presented in [12]. The definition of DSD is purely graph theoretic and is based on random walks.

Let $G = (V, E)$ be a simple undirected graph on the vertex set $\{v_1, v_2, \ldots, v_n\}$. For any two vertices $u$ and $v$, let $He^{\{k\}}(u, v)$ be the expected number of times that a random walk starting at node $u$ and proceeding for $k$ steps, will visit node $v$. Let $He^{\{k\}}(u)$ be the vector $(He^{\{k\}}(u, v_1), \ldots, He^{\{k\}}(u, v_n))$. The diffusion state distance (or DSD, for short) between two vertices $u$ and $v$ is defined as

$$DSD(u, v) = \lim_{k\to\infty} \left\| He^{\{k\}}(u) - He^{\{k\}}(v) \right\|_1$$

provided the limit exists (see [12]). Here the $L_1$-norm is not essential. Generally, for $q \geq 1$, one can define the DSD $L_q$-distance as

$$DSD_q(u, v) = \lim_{k\to\infty} \left\| He^{\{k\}}(u) - He^{\{k\}}(v) \right\|_q$$

[1]E. Boehnlein, P. Chin, A. Sinha, and L. Lu, *Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings*, Springer International Publishing, **8882** (2014), 79–95. Reprinted here with permission of Springer.

provided the limit exists. (We use $L_q$ rather than $L_p$ to avoid confusion, as $p$ will be used as a probability throughout the paper.)

In [12], Cowen et al. showed that the above limit always exists whenever the random walk on $G$ is ergodic (i.e., $G$ is connected non-bipartite graph). They also prove that this distance can be computed by the following formula:

$$DSD(u,v) = \|(1_u - 1_v)(I - D^{-1}A + W)^{-1}\|_1$$

where $D$ is the diagonal degree matrix, $A$ is the adjacency matrix, and $W$ is the constant matrix in which each row is a copy of $\pi$ , $\pi = \frac{1}{\sum_{i=1}^{n} d_i}(d_1, \ldots, d_n)$ is the unique steady state distribution.

A natural question is how to define the diffusion state distance for a bipartite graph. We suggest to use the lazy random walk. For a given $\alpha \in (0,1)$, one can choose to stay at the current node $u$ with probability $\alpha$, and choose to move to one of its neighbors with probability $(1-\alpha)/d_u$. In other words, the transitive matrix of the $\alpha$-lazy random walk is

$$T_\alpha = \alpha I + (1-\alpha)D^{-1}A.$$

Similarly, let $He_\alpha^{\{k\}}(u,v)$ be the expected number of times that the $\alpha$-lazy random walk starting at node $u$ and proceeding for $k$ steps, will visit node $v$. Let $He_\alpha^{\{k\}}(u)$ be the vector $(He_\alpha^{\{k\}}(u,v_1), \ldots, He_\alpha^{\{k\}}(u,v_n))$. The $\alpha$-diffusion state distance $L_q$-distance between two vertices $u$ and $v$ is

$$DSD_q^\alpha(u,v) = \lim_{k\to\infty} \left\| He_\alpha^{\{k\}}(u) - He_\alpha^{\{k\}}(v) \right\|_q.$$

**Theorem 3.1.** *For any connected graph $G$ and $\alpha \in (0,1)$, the $DSD_q^\alpha(u,v)$ is always well-defined and satisfies*

$$DSD_q^\alpha(u,v) = (1-\alpha)^{-1}\|(\mathbf{1}_u - \mathbf{1}_v)\mathbb{G}\|_q. \tag{3.1}$$

*Here $\mathbb{G}$ is the matrix of Green's function of $G$.*

Observe that $(1-\alpha)DSD_q^{\alpha}(u,v)$ is independent of the choice of $\alpha$. Naturally, we define the DSD $L_q$-distance of any graph $G$ as:

$$DSD_q(u,v) := \lim_{\alpha\to 0}(1-\alpha)DSD_q^{\alpha}(u,v) = \|(\mathbf{1}_u - \mathbf{1}_v)\mathbb{G}\|_q.$$

This definition extends the original definition for non-bipartite graphs.

With properly chosen $\alpha$, $\|He_{\alpha}^{\{k\}}(u) - He_{\alpha}^{\{k\}}(v)\|_q$ converges faster than $\|He^{\{k\}}(u) - He^{\{k\}}(v)\|_q$. This fact leads to a faster algorithm to estimate a single distance $DSD_q(u,v)$ using random walks. We will discuss it in Remark 3.2.

Green's function was introduced in 1828 by George Green [41] to solve some partial differential equations, and it has found many applications (e.g. [6], [22],[16], [32], [47], [69]).

The Green's function on graphs was first investigated by Chung and Yau [22] in 2000. Given a graph $G = (V,E)$ and a given function $g\colon V \to \mathbb{R}$, consider the problem to find $f$ satisfying the discrete Laplace equation

$$Lf = \sum_{y\in V}(f(x) - f(y))p_{xy} = g(x).$$

Here $p_{xy}$ is the transition probability of the random walk from $x$ to $y$. Roughly speaking, Green's function is the left inverse operator of $L$ (for the graphs with boundary). It is closely related to the Heat kernel of the graphs (see also [27]) and the normalized Laplacian.

In this paper, we will use Green's function to compute the DSD $L_q$-distance for various graphs. The maximum DSD $L_q$-distance varies from graphs to graphs. The maximum DSD $L_q$-distance for paths and cycles are at the order of $\Theta(n^{1+1/q})$ while the $L_q$-distance for some random graphs $G(n,p)$ and $G(w_1,\ldots,w_n)$ are constant for some ranges of $p$. The hypercubes are somehow between the two classes. The DSD $L_1$-distance is $\Omega(n)$ while the $L_q$-distance is $\Theta(1)$ for $q>1$. Our method for random graphs is based on the strong concentration of the Laplacian eigenvalues.

The paper is organized as follows. In Section 2, we will briefly review the terminology on the Laplacian eigenvalues, Green's Function, and heat kernel. The proof of Theorem 3.1 will be proved in Section 3. In Section 4, we apply Green's function to calculate the DSD distance for various symmetric graphs like paths, cycles, and hypercubes. We will calculate the DSD $L_2$-distance for random graphs $G(n, p)$ and $G(w_1, w_2, \ldots, w_n)$ in Section 5. In the last section, we examined two brain networks: a cat and a Rhesus monkey. The distributions of the DSD distances are calculated.

### 3.2 Notation

In this paper, we only consider undirected simple graph $G = (V, E)$ with the vertex set $V$ and the edge set $E$. For each vertex $x \in V$, the *neighborhood* of $x$, denoted by $N(x)$, is the set of vertices adjacent to $x$. The *degree* of $x$, denoted by $d_x$, is the cardinality of $N(x)$. We also denote the maximum degree by $\Delta$ and the minimum degree by $\delta$.

Without loss of generalization, we assume that the set of vertices is ordered and assume $V = [n] = \{1, 2, \ldots, n\}$. Let $A$ be the adjacency matrix and $D = \mathrm{diag}(d_1, \ldots, d_n)$ be the diagonal matrix of degrees. For a given subset $S$, let the volume of $S$ to be $\mathrm{vol}(S) := \sum_{i\in S} d_i$. In particular, we write $\mathrm{vol}(G) = \mathrm{vol}(V) = \sum_{i=1}^{n} d_i$.

Let $V^*$ be the linear space of all real functions on $V$. The *discrete Laplace operator* $L\colon V^* \to V^*$ is defined as

$$L(f)(x) = \sum_{y\in N(x)} \frac{1}{d_x}(f(x) - f(y)).$$

The Laplace operator can also written as a $(n \times n)$-matrix:

$$L = I - D^{-1}A.$$

Here $D^{-1}A$ is the transition probability matrix of the (uniform) random walk on $G$. Note that $L$ is not symmetric. We consider a symmetric version

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

which is so called the *normalized Laplacian*. Both $L$ and $\mathcal{L}$ have the same set of eigenvalues. The eigenvalues of $\mathcal{L}$ can be listed as

$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq 2.$$

The eigenvalue $\lambda_1 > 0$ if and only if $G$ is connected while $\lambda_{n-1} = 2$ if and only if $G$ is a bipartite graph. Let $\phi_0, \phi_1, \ldots, \phi_{n-1}$ be a set of orthogonal unit eigenvectors. Here $\phi_0 = \frac{1}{\sqrt{\mathrm{vol}(G)}}(\sqrt{d_1}, \ldots, \sqrt{d_n})$ is the positive unit eigenvector for $\lambda_0 = 0$ and $\phi_i$ is the eigenvector for $\lambda_i$ $(1 \leq i \leq n-1)$.

Let $O = (\phi_0, \ldots, \phi_{n-1})$ and $\Lambda = \mathrm{diag}(0, \lambda_1, \ldots, \lambda_{n-1})$. Then $O$ is an orthogonal matrix and $\mathcal{L}$ be diagonalized as

$$\mathcal{L} = O\Lambda O'. \tag{3.2}$$

Equivalently, we have

$$L = D^{-1/2} O\Lambda O' D^{1/2}. \tag{3.3}$$

The *Green's function* $\mathbb{G}$ is the matrix with its entries, indexed by vertices $x$ and $y$, defined by a set of two equations:

$$\mathbb{G}L(x,y) = I(x,y) - \frac{d_y}{\mathrm{vol}(G)}, \tag{3.4}$$

$$\mathbb{G}\mathbf{1} = 0. \tag{3.5}$$

(This is the so-called Green's function for graphs without boundary in [22].)

The *normalized Green's function* $\mathcal{G}$ is defined similarly:

$$\mathcal{G}L(x,y) = I(x,y) - \frac{\sqrt{d_x d_y}}{\mathrm{vol}(G)}.$$

The matrices $\mathbb{G}$ and $\mathcal{G}$ are related by

$$\mathcal{G} = D^{1/2}\mathbb{G}D^{-1/2}.$$

Alternatively, $\mathcal{G}$ can be defined using the eigenvalues and eigenvectors of $\mathcal{L}$ as follows:

$$\mathcal{G} = O\Lambda^{\{-1\}}O',$$

where $\Lambda^{\{-1\}} = \text{diag}(0, \lambda_1^{-1}, \ldots, \lambda_{n-1}^{-1})$. Thus, we have

$$\mathbb{G}(x, y) = \sum_{l=1}^{n-1} \frac{1}{\lambda_l} \sqrt{\frac{d_y}{d_x}} \phi_l(x)\phi_l(y). \tag{3.6}$$

For any real $t \geq 0$, the heat kernel $\mathcal{H}_t$ is defined as

$$\mathcal{H}_t = e^{-t\mathcal{L}}.$$

Thus,

$$\mathcal{H}_t(x, y) = \sum_{l=0}^{n-1} e^{-\lambda_i t} \phi_l(x)\phi_l(y).$$

The heat kernel $\mathcal{H}_t$ satisfies the heat equation

$$\frac{d}{dt}\mathcal{H}_t f = -\mathcal{L}\mathcal{H}_t f.$$

The relation of the heat kernel and Green's function is given by

$$\mathcal{G} = \int_0^\infty \mathcal{H}_t dt - \phi_0' \phi_0.$$

The heat kernel can be used to compute Green's function for the Cartesian product of two graphs. We will omit the details here. Readers are directed to [22] and [15] for the further information.

## 3.3 Proof of Theorem 3.1

*Proof.* Rewrite the transition probability matrix $T_\alpha$ as

$$\begin{aligned} T_\alpha &= \alpha I + (1 - \alpha) D^{-1} A. \\ &= D^{-1/2}(\alpha I + (1 - \alpha) D^{-1/2} A D^{-1/2}) D^{1/2} \\ &= D^{-1/2}(\alpha I + (1 - \alpha)(I - \mathcal{L})) D^{1/2} \\ &= D^{-1/2}(I - (1 - \alpha)\mathcal{L}) D^{1/2}. \end{aligned}$$

For $k = 0, 1, \ldots, n-1$, let $\lambda_k^* = 1 - (1 - \alpha)\lambda_k$ and $\Lambda^* = diag(\lambda_0^*, \ldots, \lambda_{n-1}^*) = I - (1 - \alpha)\Lambda$. Applying Equation (3.3), we get

$$T_\alpha = D^{-1/2} O \Lambda^* O' D^{1/2} = (O' D^{1/2})^{-1} \Lambda^* O' D^{1/2}.$$

Then for any $t \geq 1$, the $t$-step transition matrix is $T_\alpha^t = (OD^{1/2})^{-1}\Lambda^{*t}OD^{1/2} = D^{-1/2}O\Lambda^{*t}O'D^{1/2}$. Denote $p_\alpha^{\{t\}}(u,j)$ as the $(u,j)^{th}$ entry in $T_\alpha^t$.

$$\begin{aligned} p_\alpha^{\{t\}}(u,j) &= \sum_{l=0}^{n-1}(\lambda_l^*)^t\sqrt{\frac{d_j}{d_u}}\phi_l(u)\phi_l(j) \\ &= \frac{d_j}{\mathrm{vol}(G)} + \sum_{l=1}^{n-1}(\lambda_l^*)^t\sqrt{\frac{d_j}{d_u}}\phi_l(u)\phi_l(j). \end{aligned}$$

Thus,

$$He_\alpha^{\{k\}}(u,j) - He_\alpha^{\{k\}}(v,j) = \sum_{t=0}^{k}\sum_{l=1}^{n-1}(\lambda_l^*)^t d_j^{1/2}\phi_l(j)(d_u^{-1/2}\phi_l(u) - d_v^{-1/2}\phi_l(v)).$$

The limit $\lim_{k\to\infty} He_\alpha^{\{k\}}(u,j) - He_\alpha^{\{k\}}(v,j)$ forms the sum of $n$ geometric series:

$$\sum_{t=0}^{\infty}\sum_{l=1}^{n-1}(\lambda_l^*)^t d_j^{1/2}\phi_l(j)(d_u^{-1/2}\phi_l(u) - d_v^{-1/2}\phi_l(v)).$$

Note each geometric series converges since the common ratio $\lambda_l^* \in (-1,1)$. Thus,

$$\begin{aligned} \lim_{k\to\infty}\left(He_\alpha^{\{k\}}(u,j) - He_\alpha^{\{k\}}(v,j)\right) &= \sum_{t=0}^{\infty}\sum_{l=1}^{n-1}(\lambda_l^*)^t d_j^{1/2}\phi_l(j)(d_u^{-1/2}\phi_l(u) - d_v^{-1/2}\phi_l(v)) \\ &= \sum_{l=1}^{n-1} d_j^{1/2}\phi_l(j)(d_u^{-1/2}\phi_l(u) - d_v^{-1/2}\phi_l(v))\sum_{t=0}^{\infty}(\lambda_l^*)^t \\ &= \sum_{l=1}^{n-1}\frac{1}{1-\lambda_l^*}d_j^{1/2}\phi_l(j)(d_u^{-1/2}\phi_l(u) - d_v^{-1/2}\phi_l(v)) \\ &= \frac{1}{1-\alpha}\sum_{l=1}^{n-1}\frac{1}{\lambda_l}d_j^{1/2}\phi_l(j)(d_u^{-1/2}\phi_l(u) - d_v^{-1/2}\phi_l(v)) \\ &= \frac{1}{1-\alpha}(\mathbb{G}(u,j) - \mathbb{G}(v,j)). \end{aligned}$$

We have

$$\lim_{k\to\infty} He_\alpha^{\{k\}}(u) - He_\alpha^{\{k\}}(v) = \frac{1}{1-\alpha}(\mathbf{1}_u - \mathbf{1}_v)\mathbb{G}.$$

$\square$

**Remark 3.2.** Observe that the convergence rate of $He_\alpha^{\{k\}}(u) - He_\alpha^{\{k\}}(v)$ is determined by $\bar{\lambda}^* := \max\{1-(1-\alpha)\lambda_1, (1-\alpha)\lambda_{n-1} - 1)$. It is critical that we assume $\alpha \neq 0$.

When $\alpha = 0$ then $\bar{\lambda}^* < 1$ holds only if $\lambda_{n-1} < 2$, i.e. $G$ is a non-bipartite graph (see [12]).

When $\lambda_1 + \lambda_{n-1} > 2$, $\bar{\lambda}^*$ (as a function of $\alpha$) achieves the minimum value $\frac{\lambda_{n-1} - \lambda_1}{\lambda_{n-1} + \lambda_1}$ at $\alpha = 1 - \frac{2}{\lambda_1 + \lambda_{n-1}}$. This is the best mixing rate that the $\alpha$-lazy random walk on $G$ can achieve. Using the $\alpha$-lazy random walks (with $\alpha = 1 - \frac{2}{\lambda_1 + \lambda_{n-1}}$) to approximate the DSD $L_q$-distance will be faster than using regular random walks.

Equation (3.6) implies $\|\mathbb{G}\|_2 \le \frac{1}{\lambda_1}\sqrt{\frac{\Delta}{\delta}}$. Combining with Theorem 3.1, we have

**Corollary 3.3.** *For any connected simple graph $G$, and any two vertices $u$ and $v$, we have $DSD_2(u, v) \le \frac{\sqrt{2}}{\lambda_1}\sqrt{\frac{\Delta}{\delta}}$.*

Note that for any connected graph $G$ with diameter $m$ (Lemma 1.9, [15])

$$\lambda_1 > \frac{1}{m \operatorname{vol}(G)}.$$

This implies a uniform bound for the DSD $L_2$ distances on any connected graph $G$ on $n$ vertices.

$$DSD_2(u, v) \le \sqrt{\frac{2\Delta}{\delta}} m \operatorname{vol}(G) < \sqrt{2} n^{3.5}.$$

This is a very coarse upper bound. But it does raise an interesting question "How large can the DSD $L_q$-distance be?"

### 3.4 Some Examples of the DSD Distance

In this section, we use Green's function to compute the DSD $L_q$-distance (between two vertices of the distance reaching the diameter) for paths, cycles, and hypercubes.

## The path $P_n$

We label the vertices of $P_n$ as $1, 2, \ldots, n$, in sequential order. Chung and Yau computed the Green's function $\mathcal{G}$ of the weighed path with no boundary (Theorem 9,

[22]). It implies that Green's function of the path $P_n$ is given by: for any $u \le v$,

$$\begin{aligned}
\mathcal{G}(u,v) &= \frac{\sqrt{d_u d_v}}{4(n-1)^2}\Bigg(\sum_{z<u}(d_1+\ldots+d_z)^2+\sum_{v\le z}(d_{z+1}+\cdots+d_n)^2 \\
&\quad -\sum_{u\le z<v}(d_1+\cdots+d_z)(d_{z+1}+\cdots+d_n)\Bigg) \\
&= \frac{\sqrt{d_u d_v}}{4(n-1)^2}\left(\sum_{z=1}^{u-1}(2z-1)^2+\sum_{z=v}^{n-1}(2n-2z-1)^2-\sum_{z=u}^{v-1}(2z-1)(2n-2z-1)\right) \\
&= \frac{\sqrt{d_u d_v}}{4(n-1)^2}\left(\sum_{z=1}^{n-1}(2z-1)^2+\sum_{z=v}^{n-1}(2n-2)(2n-4z)-\sum_{z=u}^{v-1}(2z-1)(2n-2)\right) \\
&= \frac{\sqrt{d_u d_v}(2n-1)(2n-3)}{12(n-1)}+\frac{\sqrt{d_u d_v}}{2(n-1)}\left(\sum_{z=v}^{n-1}(2n-4z)-\sum_{z=u}^{v-1}(2z-1)\right) \\
&= \frac{\sqrt{d_u d_v}}{2(n-1)}\left((u-1)^2+(n-v)^2-\frac{2n^2-4n+3}{6}\right).
\end{aligned}$$

When $u > v$, we have

$$\mathcal{G}(u,v)=\mathcal{G}(v,u)=\frac{\sqrt{d_u d_v}}{2(n-1)}\left((v-1)^2+(n-u)^2-\frac{2n^2-4n+3}{6}\right).$$

Applying $\mathbb{G}(u,v)=\frac{\sqrt{d_v}}{\sqrt{d_u}}\mathcal{G}(u,v)$, we get

$$\mathbb{G}(u,v)=\begin{cases}\frac{d_v}{2(n-1)}\left((u-1)^2+(n-v)^2-\frac{2n^2-4n+3}{6}\right) & \text{if } u\le v;\\ \frac{d_v}{2(n-1)}\left((v-1)^2+(n-u)^2-\frac{2n^2-4n+3}{6}\right) & \text{if } u>v.\end{cases}$$

We have

$$\begin{aligned}
\mathbb{G}(1,1) &= \frac{4n^2-8n+3}{12(n-1)};\\
\mathbb{G}(1,j) &= \frac{1}{n-1}\left((n-j)^2-\frac{2n^2-4n+3}{6}\right) \qquad \text{for } 2\le j\le n-1;\\
\mathbb{G}(1,n) &= -\frac{2n^2-4n+3}{12(n-1)};\\
\mathbb{G}(n,1) &= -\frac{2n^2-4n+3}{12(n-1)};\\
\mathbb{G}(n,j) &= \frac{1}{n-1}\left((j-1)^2-\frac{2n^2-4n+3}{6}\right) \qquad \text{for } 2\le j\le n-1;\\
\mathbb{G}(n,n) &= \frac{4n^2-8n+3}{12(n-1)}.
\end{aligned}$$

Thus,

$$\mathbb{G}(1,j)-\mathbb{G}(n,j)=\begin{cases}\frac{n-1}{2} & \text{if } j=1;\\ n+1-2j & \text{if } 2\le j\le n-1;\\ -\frac{n-1}{2} & \text{if } j=n.\end{cases} \tag{3.7}$$

**Theorem 3.4.** *For any $q \geq 1$, the DSD $L_q$-distance of the Path $P_n$ between $1$ and $n$ satisfies*

$$DSD_q(1,n)=(1+q)^{-1/q}n^{1+1/q}+O(n^{1/q}).$$

*Proof.*

$$\begin{aligned}DSD_q(1,n)&=\left(2\left(\frac{n-1}{2}\right)^q+\sum_{j=2}^{n-1}|n+1-2j|^q\right)^{1/q}\\&=\left(\frac{1}{1+q}n^{1+q}+O(n^q)\right)^{1/q}\\&=(1+q)^{-1/q}n^{1+1/q}+O(n^{1/q}).\end{aligned}$$

□

For $q=1$, we have the following exact result:

$$\begin{aligned}DSD_1(1,n)&=\sum_{j=1}^{n}|\mathbb{G}(1,j)-\mathbb{G}(n,j)|\\&=\begin{cases}2k^2-2k+1 & \text{if } n=2k\\ 2k^2 & \text{if } n=2k+1.\end{cases}\end{aligned}$$

## The cycle $C_n$

Now we consider Green's function of cycle $C_n$. For $x, y \in \{1, 2, \ldots, n\}$, let $|x-y|_c$ be the graph distance of $x, y$ in $C_n$. We have the following Lemma.

**Lemma 3.5.** *For even $n = 2k$, Green's function $\mathbb{G}$ of $C_n$ is given by*

$$\mathbb{G}(x,y)=\frac{1}{2k}(k-|x-y|_c)^2-\frac{k}{6}-\frac{1}{12k}.$$

*For odd $n = 2k+1$, Green's function $\mathbb{G}$ of $C_n$ is given by*

$$\mathbb{G}(x,y) = \frac{2}{2k+1}\binom{k+1-|x-y|_c}{2} - \frac{k^2+k}{3(2k+1)}.$$

*Proof.* We only prove the even case here. The odd case is similar and will be left to the readers.

For $n = 2k$, it suffices to verify that $\mathbb{G}$ satisfies Equations (3.4) and (3.5). To verify Equation (3.4), we need show

$$\mathbb{G}(x,y) - \frac{1}{2}\mathbb{G}(x,y-1) - \frac{1}{2}\mathbb{G}(x,y+1) = \begin{cases} -\frac{1}{n} \text{ if } x \neq y; \\ 1 - \frac{1}{n} \text{ if } x = y. \end{cases}$$

Let $z = \frac{k}{6} + \frac{1}{12k}$ and $i = |x-y|_c$. For $x \neq y$, we have

$$\begin{aligned} &\mathbb{G}(x,y) - \frac{1}{2}\mathbb{G}(x,y-1) - \frac{1}{2}\mathbb{G}(x,y+1) \\ &\qquad = (\frac{1}{2k}(k-i)^2 - z) - \frac{1}{2}(\frac{1}{2k}(k-i-1)^2 - z) - \frac{1}{2}(\frac{1}{2k}(k-i+1)^2 - z) \\ &\qquad = -\frac{1}{2k} \\ &\qquad = -\frac{1}{n}. \end{aligned}$$

When $x = y$, we have

$$\begin{aligned} &\mathbb{G}(x,y) - \frac{1}{2}\mathbb{G}(x,y-1) - \frac{1}{2}\mathbb{G}(x,y+1) \\ &\qquad = \frac{1}{2k}k^2 - z - \frac{1}{2}\left(\frac{1}{2k}(k-1)^2 - z\right) - \frac{1}{2}\left(\frac{1}{2k}(k-1)^2 - z\right) \\ &\qquad = \frac{2k-1}{2k} \\ &\qquad = 1 - \frac{1}{n}. \end{aligned}$$

To verify Equation (3.5), it is enough to verify

$$1^2 + 2^2 + \cdots + (k-1)^2 + k^2 + (k-1)^2 + \cdots + 1^2 = \frac{2k^3+k}{3} = n^2 z.$$

This can be done by induction on $k$. $\square$

**Theorem 3.6.** *For any $q \geq 1$, the DSD $L_q$-distance of the Cycle $C_n$ between 1 and $\lfloor \frac{n}{2} \rfloor + 1$ satisfies*

$$DSD_q(1, \lfloor \frac{n}{2} \rfloor + 1) = \left( \frac{4}{1+q} \right)^{1/q} \left( \frac{n}{4} \right)^{1+1/q} + O(n^{1/q}).$$

*Proof.* We only verify the case of even cycle here. The odd cycle is similar and will be omitted.

For $n = 2k$, the difference of $\mathbb{G}(1, j)$ and $\mathbb{G}(1 + k, j)$ have a simple form:

$$\mathbb{G}(1, j) - \mathbb{G}(1 + k, j) = \frac{1}{2k}((k - i)^2 - i^2) = \frac{k}{2} - i,$$

where $i = |j - 1|_c$. Thus,

$$\begin{aligned}
DSD_q(1, 1 + k) &= \left( 2 \sum_{i=0}^{k-1} \left| \frac{k}{2} - i \right|^q \right)^{1/q} \\
&= \left( \frac{4}{1+q} \left( \frac{k}{2} \right)^{1+q} + O(k^q) \right)^{1/q} \\
&= \left( \frac{4}{1+q} \right)^{1/q} \left( \frac{n}{4} \right)^{1+1/q} + O(n^{1/q}).
\end{aligned}$$

□

## The hypercube $Q_n$

Now we consider the hypercube $Q_n$, whose vertices are the binary strings of length $n$ and whose edges are pairs of vertices differing only at one coordinate. Chung and Yau [22] computed the Green's function of $Q_n$: for any two vertices $x$ and $y$ with distance $k$ in $Q_n$,

$$\begin{aligned}
\mathbb{G}(x, y) &= 2^{-2n} \left( -\sum_{j<k} \frac{(\binom{n}{0} + \cdots + \binom{n}{j})(\binom{n}{j+1} + \cdots + \binom{n}{n})}{\binom{n-1}{j}} + \sum_{k \leq j} \frac{(\binom{n}{j+1} + \cdots + \binom{n}{n})^2}{\binom{n-1}{j}} \right) \\
&= 2^{-2n} \sum_{j=0}^{n} \frac{(\binom{n}{j+1} + \cdots + \binom{n}{n})^2}{\binom{n-1}{j}} - 2^{-n} \sum_{j<k} \frac{\binom{n}{j+1} + \cdots + \binom{n}{n}}{\binom{n-1}{j}}.
\end{aligned}$$

We are interested in the DSD distance between a pair of antipodal vertices. Let $\mathbf{0}$ denote the all-0-string and $\mathbf{1}$ denote the all-1-string. For any vertex $x$, if the distance between $\mathbf{0}$ and $x$ is $i$ then the distance between $\mathbf{1}$ and $x$ is $n-i$. We have

$$\begin{aligned}\mathbb{G}(\mathbf{0},x)-\mathbb{G}(\mathbf{1},x) &= -2^{-n}\sum_{j<k}\frac{\binom{n}{j+1}+\cdots+\binom{n}{n}}{\binom{n-1}{j}} + 2^{-n}\sum_{j<n-k}\frac{\binom{n}{j+1}+\cdots+\binom{n}{n}}{\binom{n-1}{j}} \\ &= 2^{-n}\sum_{j=k}^{n-k-1}\frac{\binom{n}{j+1}+\cdots+\binom{n}{n}}{\binom{n-1}{j}}. \end{aligned} \tag{3.8}$$

Here we use the convention that $\sum_{j=b}^{a} c_j = -\sum_{j=a}^{b} c_j$ for $b > a$.

**Theorem 3.7.** *For any $q \geq 1$, the DSD $L_q$-distance of the hypercube $Q_n$ between $\mathbf{0}$ and $\mathbf{1}$ satisfies*

$$DSD_q(\mathbf{0},\mathbf{1}) = \left(\sum_{k=0}^{n}\binom{n}{k}\left|2^{-n}\sum_{j=k}^{n-k-1}\frac{\binom{n}{j+1}+\cdots+\binom{n}{n}}{\binom{n-1}{j}}\right|^q\right)^{1/q}. \tag{3.9}$$

*In particular, $DSD_q(\mathbf{0},\mathbf{1}) = \Theta(1)$ when $q > 1$ while $DSD_1(\mathbf{0},\mathbf{1}) = \Omega(n)$.*

*Proof.* Equation (3.9) follows from the definition of DSD $L_q$-distance and Equation (3.8). Let

$$a_k = \binom{n}{k}\left|2^{-n}\sum_{j=k}^{n-k-1}\frac{\binom{n}{j+1}+\cdots+\binom{n}{n}}{\binom{n-1}{j}}\right|^q.$$

Observe that $a_k = a_{n-k}$, we only need to estimate $a_k$ for $0 \leq k \leq n/2$. Also we can throw away the terms in the second summation for $j > n/2$ since that part is at most half of $a_k$. For $k \leq j \leq n/2$,

$$\frac{1}{2} \leq 2^{-n}\left(\binom{n}{j+1}+\cdots+\binom{n}{n}\right) \leq 1.$$

Thus $a_k$ has the same magnitude as $b_k := \binom{n}{k}\left(\sum_{j=k}^{n/2}\frac{1}{\binom{n-1}{j}}\right)^q$.

For $q > 1$, we first bound $b_k$ by $b_k \leq \binom{n}{k}\left(\frac{n/2}{\binom{n-1}{k}}\right)^q = O(n^{(1-q)k+q})$. When $k > \frac{q+2}{q-1}$, we have $b_k = O(n^{-2})$. The total contribution of those $b_k$'s is $O(n^{-1})$, which is negligible. Now consider the term $b_k$ for $k = 0, 1, \ldots, \lfloor\frac{q+2}{q-1}\rfloor$. We bound $b_k$ by

$$b_k \leq \binom{n}{k}\left(\frac{1}{\binom{n-1}{k}} + \frac{n/2}{\binom{n-1}{k+1}}\right)^q = O(1).$$

This implies $DSD_q(\mathbf{0},\mathbf{1}) = O(1)$. The lower bound $DSD_q(\mathbf{0},\mathbf{1}) \geq 1$ is obtained by taking the term at $k = 0$. Putting together, we have $DSD_q(\mathbf{0},\mathbf{1}) = \Theta(1)$ for $q > 1$.

For $q = 1$, note that

$$b_k = \sum_{j=k}^{n/2} \frac{\binom{n}{k}}{\binom{n-1}{j}} > \frac{\binom{n}{k}}{\binom{n-1}{k}} = \frac{n}{n-k} > 1.$$

Thus, $DSD_1(\mathbf{0},\mathbf{1}) = \Omega(n)$. $\square$

## 3.5 Random Graphs

In this section, we will calculate the DSD $L_q$-distance in two random graphs models. For random graphs, the non-zero Laplacian eigenvalues of a graph $G$ are often concentrated around 1. The following Lemma is useful to the DSD $L_q$-distance.

**Lemma 3.8.** *Let $\lambda_1, \ldots, \lambda_{n-1}$ be all non-zero Laplacian eigenvalues of a graph $G$. Suppose there is a small number $\epsilon \in (0, 1/2)$, so that for $1 \leq i \leq n-1$, $|1 - \lambda_i| \leq \epsilon$. Then for any pairs of vertices $u, v$, the DSD $L_q$-distance satisfies*

$$|DSD_q(u,v) - 2^{1/q}| \leq \frac{\epsilon}{1-\epsilon}\sqrt{\frac{\Delta}{d_u} + \frac{\Delta}{d_v}} \quad \text{if } q \geq 2, \tag{3.10}$$

$$|DSD_q(u,v) - 2^{1/q}| \leq n^{\frac{1}{q}-\frac{1}{2}}\frac{\epsilon}{1-\epsilon}\sqrt{\frac{\Delta}{d_u} + \frac{\Delta}{d_v}} \quad \text{for } 1 \leq q < 2. \tag{3.11}$$

*Proof.* Rewrite the normalized Green's function $\mathcal{G}$ as

$$\mathcal{G} = I - \phi_0'\phi_0 + \Upsilon.$$

Note that the eigenvalues of $\Upsilon := \mathcal{G} - I + \phi_0\phi_0'$ are $0, \frac{1}{\lambda_1} - 1, \ldots, \frac{1}{\lambda_{n-1}} - 1$. Observe that for each $i = 1, 2, \ldots, n-1$, $|\frac{1}{\lambda_i} - 1| \leq \frac{\epsilon}{1-\epsilon}$. We have

$$\|\Upsilon\| \leq \frac{\epsilon}{1-\epsilon}.$$

Thus,

$$\begin{aligned}
\mathrm{DSD}_q(u,v) &= \|(\mathbf{1}_u - \mathbf{1}_v)D^{-1/2}\mathcal{G}D^{1/2}\|_q \\
&= \|(\mathbf{1}_u - \mathbf{1}_v)D^{-1/2}(I - \phi_0'\phi + \Upsilon)D^{1/2}\|_q \\
&\leq \|(\mathbf{1}_u - \mathbf{1}_v)D^{-1/2}(I - \phi_0'\phi)D^{1/2}\|_q + \|(\mathbf{1}_u - \mathbf{1}_v)D^{-1/2}\Upsilon D^{1/2}\|_q.
\end{aligned}$$

Viewing $\Upsilon$ as the error term, we first calculate the main term.

$$\begin{aligned}
&\|(\mathbf{1}_u - \mathbf{1}_v)D^{-1/2}(I - \phi_0'\phi)D^{1/2}\|_q \\
&\quad = \|(\mathbf{1}_u - \mathbf{1}_v)(I - W)\|_q \\
&\quad = \|(\mathbf{1}_u - \mathbf{1}_v)\|_q \\
&\quad = 2^{1/q}.
\end{aligned}$$

The $L_2$-norm of the error term can be bounded by

$$\begin{aligned}
&\|(\mathbf{1}_u - \mathbf{1}_v)D^{-1/2}\Upsilon D^{1/2}\|_2 \\
&\quad \leq \|(\mathbf{1}_u - \mathbf{1}_v)D^{-1/2}\|_2\|\Upsilon\|\|D^{1/2}\| \\
&\quad \leq \sqrt{\frac{1}{d_u} + \frac{1}{d_v}}\frac{\epsilon}{1-\epsilon}\sqrt{\Delta} \\
&\quad = \frac{\epsilon}{1-\epsilon}\sqrt{\frac{\Delta}{d_u} + \frac{\Delta}{d_v}}.
\end{aligned}$$

To get the bound of $L_q$-norm from $L_2$-norm, we apply the following relation of $L_q$-norm and $L_2$-norm to the error term. For any vector $x \in \mathbb{R}^n$,

$$\|x\|_q \leq \|x\|_2 \quad \text{for } q \geq 2.$$

and

$$\|x\|_q \leq n^{\frac{1}{q}-\frac{1}{2}}\|x\|_2 \quad \text{for } 1 \leq q < 2.$$

The inequalities (3.10) and (3.11) follow from the triangular inequality of the $L_q$-norm and the upper bound of the error term. $\square$

Now we consider the classical Erdős-Renyi random graphs $G(n,p)$. For a given $n$ and $p \in (0,1)$, $G(n,p)$ is a random graph on the vertex set $\{1, 2, \ldots, n\}$ obtained by adding each pair $(i,j)$ to the edges of $G(n,p)$ with probability $p$ independently.

There are plenty of references on the concentration of the eigenvalues of $G(n,p)$ (for example, [23], [26],[58], and [60]). Here we list some facts on $G(n,p)$.

1. For $p > \frac{(1+\epsilon)\log n}{n}$, almost surely $G(n,p)$ is connected.

2. For $p \gg \frac{\log n}{n}$, $G(n,p)$ is "almost regular"; namely for all vertex $v$, $d_v = (1 + o_n(1))np$.

3. For $np(1-p) \gg \log^4 n$, all non-zero Laplacian eigenvalues $\lambda_i$'s satisfy (see [60])

$$|\lambda_i - 1| \leq \frac{(3 + o_n(1))}{\sqrt{np}}. \tag{3.12}$$

Apply Lemma 3.8 with $\epsilon = \frac{(3+o_n(1))}{\sqrt{np}}$, and note that $G(n,p)$ is almost-regular. We get the following theorem.

**Theorem 3.9.** *For $p(1-p) \gg \frac{\log^4 n}{n}$, almost surely for all pairs of vertices $(u,v)$, the DSD $L_q$-distance of $G(n,p)$ satisfies*

$$DSD_q(u,v) = 2^{1/q} \pm O\left(\frac{1}{\sqrt{np}}\right) \qquad \text{if } q \geq 2,$$

$$DSD_q(u,v) = 2^{1/q} \pm O\left(\frac{n^{\frac{1}{q}-\frac{1}{2}}}{\sqrt{np}}\right) \qquad \text{if } 1 \leq q < 2.$$

Now we consider the random graphs with given expected degree sequence $G(w_1,\ldots,w_n)$ (see [8], [18], [17], [16], [48]). It is defined as follows:

1. Each vertex $i$ (for $1 \leq i \leq n$) is associated with a given positive weight $w_i$.

2. Let $\rho = \frac{1}{\sum_{i=1}^n w_i}$. For each pair of vertices $(i,j)$, $ij$ is added as an edge with probability $w_i w_j \rho$ independently. ($i$ and $j$ may be equal so loops are allowed. Assume $w_i w_j \rho \leq 1$ for $i,j$.)

Let $w_{min}$ be the minimum weight. There are many references on the concentration of the eigenvalues of $G(w_1,\ldots,w_n)$ (see [19], [20], [23], [26], [60]). The version used here is in [60].

1. For each vertex $i$, the expected degree of $i$ is $w_i$.

2. Almost surely for all $i$ with $w_i \gg \log n$, then the degree $d_i = (1 + o(1))w_i$.

3. If $w_{min} \gg \log^4 n$, all non-zero Laplacian eigenvalues $\lambda_i$ (for $1 \leq i \leq n-1$),

$$|1 - \lambda_i| \leq \frac{3 + o_n(1)}{\sqrt{w_{min}}}. \tag{3.13}$$

**Theorem 3.10.** *Suppose $w_{min} \gg \log^4 n$, almost surely for all pairs of vertices $(u, v)$, the DSD $L_q$-distance of $G(w_1, \ldots, w_n)$ satisfies*

$$DSD_q(u, v) = 2^{1/q} \pm O\left(\frac{1}{\sqrt{w_{min}}}\sqrt{\frac{w_{max}}{w_u} + \frac{w_{max}}{w_v}}\right) \qquad \text{if } q \geq 2,$$

$$DSD_q(u, v) = 2^{1/q} \pm O\left(\frac{n^{\frac{1}{q}-\frac{1}{2}}}{\sqrt{w_{min}}}\sqrt{\frac{w_{max}}{w_u} + \frac{w_{max}}{w_v}}\right) \qquad \text{if } 1 \leq q < 2.$$

### 3.6 Examples of Biological Networks

In this section, we will examine the distribution of the DSD distances for some biological networks. The set of graphs analyzed in this section include three graphs of brain data from the Open Connectome Project [73] and two more graphs built from the *S. cerevisiae* PPI network and *S. pombe* PPI network used in [12]. Figure 1 and 2 serves as a visual representation of one of the two brain data graphs: the graph of a cat and the graph of a Rhesus monkey. The network of the cat brain has 65 nodes and 1139 edges while the network of rhesus monkey brain has 242 nodes and 4090 edges.

Each node in the Rhesus graph represents a region in the cerebral cortex originally analyzed in [46]. Each edge represents axonal connectivity between regions and there is no distinction between strong and weak connections in this graph [46]. The Cat data-set follows a similar pattern where each node represents a region of the brain and each edge represents connections between them. The Cat data-set represents 18 visual regions, 10 auditory regions, 18 somatomotor regions, and 19 frontolimbic regions[65].

For each network above, we calculated all-pair DSD $L_1$-distances. Divide the possible values into many small intervals and compute the number of pairs falling

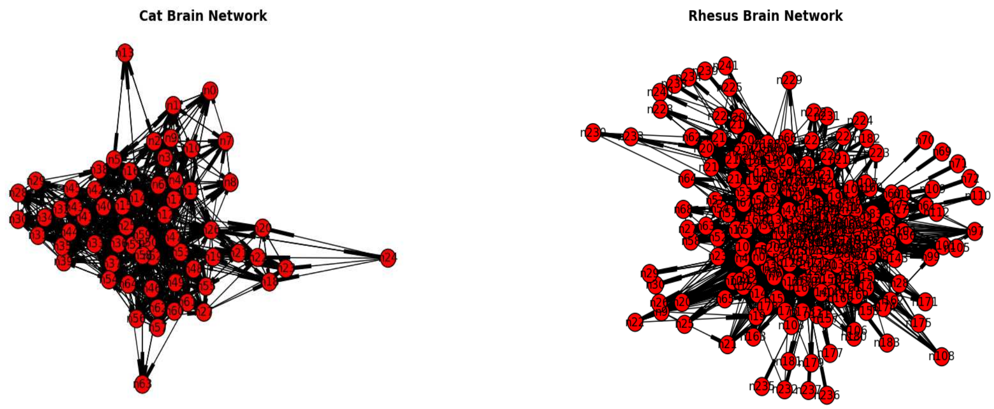


Figure 3.1 The brain networks: (a), a Cat; (b): a Rhesus Monkey

into each interval. The results are shown in Figure 3.1. The patterns are quite surprising to us.

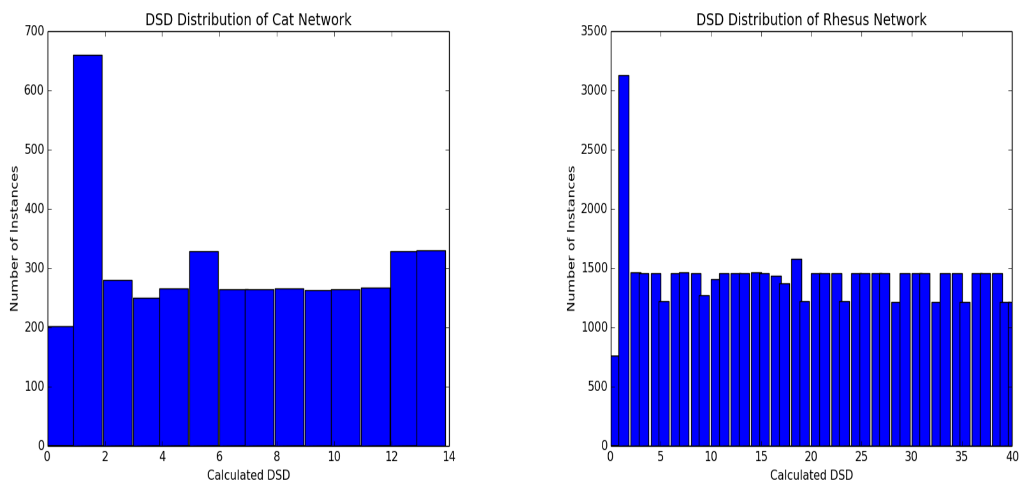


Figure 3.2 The distribution of the DSD $L_1$-distances of brain networks: (a), a Cat; (b): a Rhesus Monkey

Both graphs have a small interval consisting of many pairs while other values are more or less uniformly distributed. We think, that phenomenon might be caused by the clustering of a dense core. The two graphs have many branches sticking out. Since we are using $L_1$-distance, it doesn't matter the directions of these branches sticking out when they are embedded into $\mathbb{R}^n$ using Green's function.

When we change $L_1$-distance to $L_2$-distance, the pattern should be broken. This is confirmed in Figure 3.3. The actual distributions are mysterious to us.

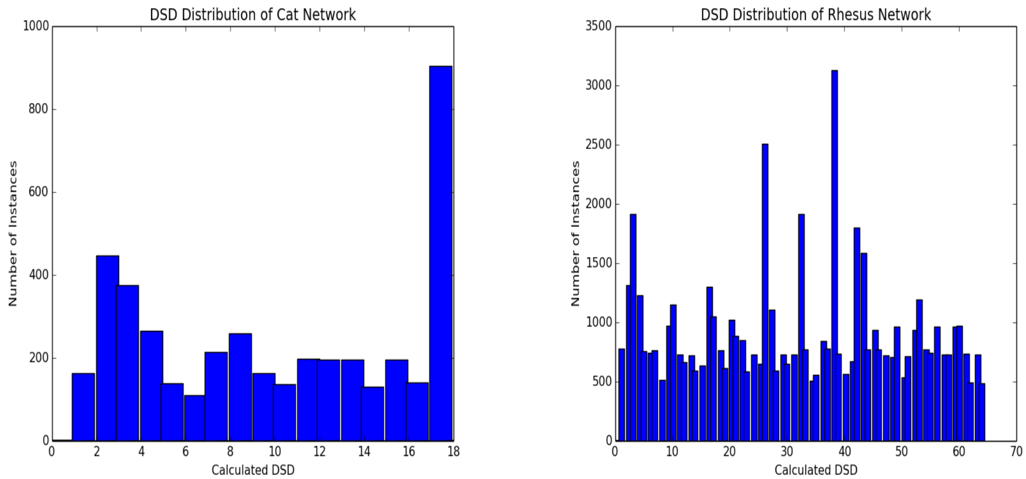


Figure 3.3 The distribution of the DSD $L_2$-distances of brain networks: (a), a Cat; (b): a Rhesus Monkey

# Chapter 4

# Non-uniform Hypergraphs

For each $i \in [r]$, where $R = \{k_1, \ldots, k_r\}$, define $f_i(H)$ to be the number of $k_i$-edges belonging to $E(H)$. Also, for each $p_i \in [0,1]$, set $\alpha_i := -\log p_i$. Let $H := H_v^R$ be an $R$-graph on $v$ vertices. Define

$$\phi(H, \mathbf{p}) = |V(H)| - \sum_{i=1}^{r} \alpha_i f_i(H),$$

and then,

$$q(H, \mathbf{p}) := \min_{H' \subset H} \phi(H', \mathbf{p}).$$

We can now state the main result.

**Theorem 4.1.** *Let $G$ be the random hypergraph $G^R(n, \mathbf{p})$. For any $\epsilon > 0$, for $n$ sufficiently large, we have:*

1. *If $q(H, \mathbf{p}) < -\epsilon$ then $G^R(n, \mathbf{p})$ is almost surely $H$-free.*
2. *If $q(H, \mathbf{p}) \geq \epsilon$, then almost surely $G^R(n, \mathbf{p})$ contains $(1 + o(1))\frac{1}{|Aut(H)|} n^{\phi(H,\mathbf{p})}$ copies of $H$.*

Using the same technique as in Theorem 4.1, we can make a claim regarding extensions, as well. First define $f_i^S(H)$ denote the number of edges from $H$ of type $i$ contained in $H|_S$. Next, define for each subhypergraph $H' \subset H$, $\phi_S(H', \mathbf{p}) := v - |S| - \sum_{i=1}^{r} (\alpha_i f_i(H') - \alpha_i f_i^S(H'))$, and $q_s(H, \mathbf{p}) := \min_{H' \subset H} \phi_S(H', \mathbf{p})$.

**Theorem 4.2.** *Suppose $G = G^R(n, \mathbf{p})$, and let a subset $S \subset V(G)$ be given. Assume $H|_S \subset G|_S$ almost surely. If $q_S(H, \mathbf{p}) \geq \epsilon$, then $H|_S$ can almost certainly be extended to a copy of $H$ in $G$.*

## 4.1 Notation, Methods, and Examples

It is not immediately clear how to extend the notions of 'balanced' or 'strictly balanced' to the non-uniform case. Consider the related extremal poset problem. One can ask what conditions are required in order to force the presence of a poset $P$ to appear in some set family $\mathcal{F}$. A necessary condition for $P$ to appear is that every subposet $P' \subset P$ appears in the family $\mathcal{F}$. With this analogy in mind, for $|R| > 1$, we propose a new definition for the hypergraph. Let $V_{H'}$ denote the number of vertices of a subhypergraph $H' \subset H$. A set of hypergraphs $H_1, \ldots, H_k$ with $H = \cup_{j=1}^{k} H_j$ is *balanced* if for all subgraphs $H' \subset H$ and all $1 \leq j \leq k$, $\sum_{i=1}^{r} \alpha_i f_i(H_j) \leq V_{H_j}$ implies $\sum_{i=1}^{r} \alpha_i f_i(H') \leq V_{H'}$ , where each $\alpha_i \geq 0$.

As an example, we consider the hypergraph $H$ pictured in Figure 4.1. The 1-edges are represented with solid circles, and the 2-edges are represented in the customary way. The set $\{H_1, H_2\}$ constitutes the balanced subgraphs of $H$, which was determined by plotting the half-planes $\alpha_1 f_1(H_i) + \alpha_2 f_2(H_i) = V_{H_i}$ for each subgraph $H_i \subset H$, as seen in Figure 4.2. Consider the intersection of these half-planes in the first quadrant, which is shaded. The lines that compose the boundary of this intersection correspond to the graphs in the balanced set. Lines that did not lie on the boundary of the image were omitted here. Now for any ordered pair $(\alpha_1, \alpha_2)$ lying in the shaded region of the plot, consider the probability vector $\mathbf{p} = (n^{-\alpha_1}, n^{-\alpha_2})$. By Theorem 4.1, the random graph $G \sim G(n, \mathbf{p})$ almost surely contains a copy of $H$.

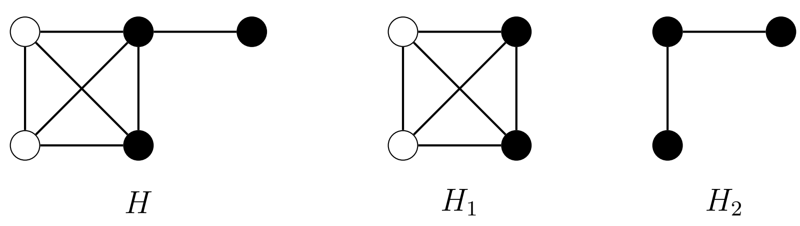


Figure 4.1 A hypergraph $H$ with $R = \{1, 2\}$, and its balanced subgraphs

Now fix any hypergraph $H = H_v^R$ on $v$ vertices with the edge type $R = \{k_1, \ldots, k_r\}$.

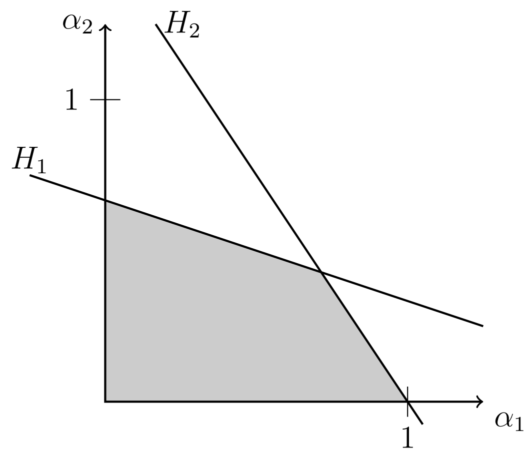


Figure 4.2  The probabilities for which $G(n, \mathbf{p})$ almost surely contains $H$

The goal is to count the number of copies of $H$ contained in the random hypergraph $G = G^R(n, \mathbf{p})$. For any $S \in \binom{n}{v}$, define an indicator variable

$$X_S := \begin{cases} 1 & \text{if } G|_S \supseteq H \\ 0 & \text{otherwise.} \end{cases}$$

contained inside $G|_S$. So set

$$X = \sum_{S \in \binom{n}{v}} X_S.$$

Then,

$$\begin{aligned} \mathbb{E}(X) &= \sum_{S \in \binom{n}{v}} \mathbb{E}(X_S) \\ &= \binom{n}{v} \frac{v!}{|\mathrm{Aut}(H)|} \prod_{i=1}^{r} p_i^{f_i(H)} \\ &\approx \frac{1}{|\mathrm{Aut}(H)|} n^{v - \sum_{i=1}^{r} \alpha_i f_i(H)}, \end{aligned}$$

where $p_i = 1/n^{\alpha_i}$.

The proof depends on an inequality from Kim and Vu. First, set $\mu = \mathbb{E}(X)$, $E' = \max_{H'} \dfrac{\partial X}{\partial H'}$, and $E = \max\{\mu, E'\}$.

**Theorem 4.3** ([55]). *For any $\lambda > 1$, if $X$ is polynomial, then*

$$Pr[|X - \mu| > a_k \sqrt{EE'} \lambda^k] < d_k e^{-\lambda} n^{k-1}$$

*where $a_k = 8^k \sqrt{k!}$ and $d_k = 2e^2$.*

To illustrate our methods more clearly, we must invoke derivatives. Take any injection $\varphi : V(H) \to [n]$. For each $F \in E(H)$, define an indicator variable $t_{\varphi(F)}$ to be 1 if $\varphi(F)$ is an edge in $G$, and 0 otherwise. Hence, we can write $X$, the number of copies of $H$ in $G$, as the polynomial:

$$X = \sum_{\varphi} \prod_{F \in E(H)} t_{\varphi(F)}$$

Next, we may define the partial derivatives of $X$ as follows. Suppose $H' \subset H$ is subgraph spanning $V(H)$. Let $S = S(H')$ denote the set of vertices incident to an edge in $E(H')$. We then restrict our attention to considering only those subgraphs $H'$ such that $H[S] \subseteq H' \subseteq H$. Let $\varphi^*$ denote those maps $\varphi$ that fix $S(H')$. Then define:

$$\frac{\partial X}{\partial H'} = \frac{\partial X}{\prod_{E \in E(H')} \partial t_E} = \sum_{\varphi^*} \prod_{F \in E(H) \setminus E(H')} t_{\varphi^*(F)}.$$

Now we examine an example hypergraph $H$ with $V(H) = \{1, 2, 3, 4\}$ and $E(H) = \{1, 4, 12, 23, 234\}$, as shown in Figure 4.3. The shaded triangle represents the 3-edge.

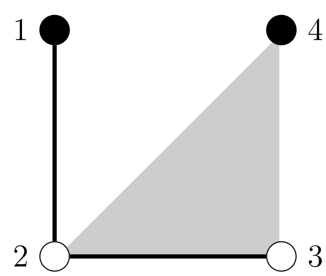


Figure 4.3 A hypergraph with $R = \{1, 2, 3\}$

Say $\varphi$ maps $(1, 2, 3, 4) \mapsto (i, j, k, \ell)$. So $X = \sum_{\varphi} t_i t_j t_{ij} t_{jk} t_{jk\ell}$. Now consider the (vertex spanning) subgraph $H'$ given by $E(H') = \{1, 12, 23\}$. Then $S(H') = \{1, 2, 3\}$.

Then taking the derivative of $H$ with respect to $H'$, we obtain:

$$\begin{aligned}\frac{\partial X}{\partial H'} &= \frac{\partial X}{\partial t_1 \partial t_{12} \partial t_{23}} = \sum_{\ell} t_\ell t_{jk\ell} \\ &= \sum_{\varphi^*} t_{\varphi^*(4)} t_{\varphi^*(2)\varphi^*(3)\varphi^*(4)}\end{aligned}$$

with expected value

$$\mathbb{E}\left(\frac{\partial X}{\partial H'}\right) = \mathbb{E}\left(\frac{\partial X}{\partial t_1 \partial t_{12} \partial t_{23}}\right) \leq np_1p_3 = n^{1-\alpha_1-\alpha_3},$$

as we have $n$ choices for $\varphi^*(4)$, the only vertex whose destination is not already determined from our choice of $H'$. For another example, we consider the subgraph $H'$ given by $E(H') = \{23, 234, 4\}$. Here, we obtain:

$$\begin{aligned}\frac{\partial X}{\partial H'} &= \frac{\partial X}{\partial t_{23} \partial t_{234} \partial t_4} = \sum_i t_i t_{ij} = \sum_{\varphi^*} t_{\varphi^*(1)} t_{\varphi^*(1)\varphi^*(2)} \\ \mathbb{E}\left(\frac{\partial X}{\partial t_{23} \partial t_{234} \partial t_4}\right) &= \mathbb{E}\left(\frac{\partial X}{\partial H'}\right) \leq np_1p_2 = n^{1-\alpha_1-\alpha_2}.\end{aligned}$$

### 4.2 Proof of Theorem 4.1

Taking $p_i \sim 1/n^{\alpha_i}$, recall that

$$\begin{aligned}\mathbb{E}(X) &= \frac{v!}{|\mathrm{Aut}(H)|}\binom{n}{v}\prod_{i=1}^{r} p_i^{f_i(H)} \\ &\approx \frac{1}{|\mathrm{Aut}(H)|} n^{v-\sum_{i=1}^{r}\alpha_i f_i(H)} \\ &= \frac{1}{|\mathrm{Aut}(H)|} n^{\phi(H,\mathbf{p})}.\end{aligned}$$

Now if $q(H, \mathbf{p}) < -\epsilon < 0$, then there exists a subgraph $H' \subset H$ such that

$$\phi(H', \mathbf{p}) = V(H') - \sum_{i=1}^{r} \alpha_i f_i(H') < -\epsilon.$$

Letting $X(H')$ denote the number of copies of $H'$ in $G$. Then,

$$\begin{aligned}E(X(H')) &\approx \frac{1}{|\mathrm{Aut}(H')|} n^{\phi(H',\mathbf{p})} \\ &= o(1).\end{aligned}$$

That is, $G$ almost surely does not contain a copy of $H'$, and thus almost surely does not contain $H$. For the other case, when $\phi(H', \mathbf{p}) \geq \epsilon > 0$ for all subgraphs $H'$, we require some more machinery. Recalling the definitions of our derivatives, $E$, and $E'$, observe that:

$$E' = \frac{1}{v!} n^{\max_{H' \subset H} q(H) - \phi(H')} \leq E n^{-\epsilon}.$$

Now for any subgraph $H' \subset H$,

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial X}{\partial H'}\right) &= \mathbb{E}\left(\sum_{\varphi^*} \prod_{F \in E(H) \setminus E(H')} t_F\right) \\
&\leq \mathbb{E}\left(\sum_{\varphi^*} \frac{\prod_{F \in E(H)} t_F}{\prod_{F \in (H')} t_F}\right) \\
&= \sum_{\varphi^*} \frac{n^{-\sum_{i=1}^r \alpha_i f_i(H)}}{n^{-\sum_{i=1}^r \alpha_i f_i(H')}} \\
&\approx n^{|S(H')|} \frac{n^{-\sum_{i=1}^r \alpha_i f_i(H)}}{n^{-\sum_{i=1}^r \alpha_i f_i(H')}} \\
&\leq \frac{n^{|V(H)|}}{n^{|V(H')|}} \frac{n^{-\sum_{i=1}^r \alpha_i f_i(H)}}{n^{-\sum_{i=1}^r \alpha_i f_i(H')}} \\
&\leq n^{\phi(H, \mathbf{p}) - \phi(H', \mathbf{p})} \\
&\leq n^{\phi(H, \mathbf{p}) - q(H', \mathbf{p})} \\
&\leq \frac{n^{\phi(H, \mathbf{p})}}{n^{\epsilon}} \\
&= (1 + o(1))|\mathrm{Aut}(H)| \frac{\mu}{n^{\epsilon}}
\end{aligned}$$

Thus, $E' \leq (1 + o(1))|\mathrm{Aut}(H)| \frac{\mu}{n^{\epsilon}}$.

But since $\mu = n^{q(H, \mathbf{p})} |\mathrm{Aut}(H)|^{-1}$, then by Theorem 4.3,

$$\Pr\left[|X - \mu| > 8^v \sqrt{v! \mu |\mathrm{Aut}(H)| \mu n^{-\epsilon}} \lambda^v\right] \; < \; 2e^2 e^{-\lambda} n^{v-1}.$$

Now taking $\lambda = v \ln n$, we obtain:

$$\Pr\left[|X - \mu| > 8^v (v! |\mathrm{Aut}(H)|)^{1/2} \mu n^{-\epsilon/2} v^v (\ln n)^v\right] \; < \; 2e^2 n^{-1}$$

However, for $\epsilon > 0$, note that $8^v (v! |\mathrm{Aut}(G)|)^{1/2} \mu n^{-\epsilon/2} v^v (\ln n)^v \to 0$ as $n \to \infty$. Therefore in this case we expect to find $(1 + o(1))\mu$ copies of $H$ inside $G$.

## 4.3 Proof of Theorem 4.2

Fix any subset $S \subset V(G)$. We wish to determine a sufficient condition for us to extend $H|_S$ to a copy of $H$ contained in $G$ for some given hypergraph $H$. To be more precise, we adopt a rigorous definition similar to that for the classic graph case in [66] and [67]. Assume $H|_S \subset G|_S$ as subgraphs. We say $H|_S$ can be extended to a copy of $H$ in $G$ if there exists a set $Z \subset V(G)$ such that for each $F \subset S \cup Z$, if $F \in E(H)$ then $F \in E(G)$.

Let $\psi : V(G) \to V(G)$ be an injection such that $\psi|_S$ is the identity map. Let $Y$ denote the number of such extensions. As before, let $|V(G)| = n$ and $|V(H)| = v$. Then,

$$Y = \sum_{\psi} \prod_{F \in E(H) \setminus E(H|_S)} t_{\psi(F)}$$

Now recall $f_i^S(H)$ denote the number of edges from $H$ of type $i$ contained in $H|_S$, $\phi_S(H', \mathbf{p}) = v - |S| - \sum_{i=1}^{r} (\alpha_i f_i(H') - \alpha_i f_i^S(H'))$ , and $q_S(H, \mathbf{p}) = \min_{H' \subset H} \phi_S(H', \mathbf{p})$. Thus,

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}\left( \sum_{\psi} \prod_{F \in E(H) \setminus E(H|_S)} t_{\psi(F)} \right) \\ &= \binom{n - |S|}{v - |S|} \frac{(v - |S|)!}{|\mathrm{Aut}(H|_{S^c})|} \prod_{i \in R} p_i^{f_i(H) - f_i^S(H)} \\ &\approx \frac{1}{|\mathrm{Aut}(H|_{S^c})|} n^{\phi_S(H, \mathbf{p})} \end{aligned}$$

So set $\mu_Y = \mathbb{E}(Y)$. Next we turn our attention to the partial derivatives of $Y$. In light of our earlier definitions of derivatives, the event that we can extend $H|_S$ to a copy of $H$ given a subgraph $H'$ already in place, is given by $\dfrac{\partial Y}{\partial \left(H|_S \cup H'\right)}$. So since $\mathbb{E}(X) \approx \mathbb{E}(Y)$, we can approximate

$$\max_{H' \subset H} \frac{\partial Y}{\partial \left(H|_S \cup H'\right)} = \max_{H' \subset H} \frac{\partial X}{\partial H'} = E'$$

So set $E'_Y = E'$ and $E_Y = \max\{\mu_Y, E'_Y\}$. Then by Theorem 4.3,

$$\Pr\left[ |Y - \mu_Y| > 8^{v-|S|} \sqrt{(v - |S|)! \mu_Y |\mathrm{Aut}(H|_{S^c})| \mu_y n^{-\epsilon}} \lambda^{v-|S|} \right] < 2e^2 e^{-\lambda} n^{v-|S|-1}.$$

Now taking $\lambda = (v - |S|) \ln n$,

$$\Pr\left[|Y - \mu_Y| > 8^{v-|S|}(v - |S|)!^{1/2}|\mathrm{Aut}(H|_{S^c})|^{1/2}\mu_Y n^{-\epsilon/2}(v - |S|)^{v-|S|}\right] < 2e^2 n^{-1}.$$

However, for $\epsilon > 0$, note that $8^{v-|S|}(v-|S|)!^{1/2}|\mathrm{Aut}(H|_{S^c})|^{1/2}\mu_Y n^{-\epsilon/2} v - |S|^{v-|S|} \to 0$ as $n \to \infty$. Thus, we can almost certainly extend $H|_S$ into a copy of $H$ inside $G$. In particular, we expect to find $(1 + o(1))\mu_Y$ many extensions of $H|_S$.

### 4.4 Improving Theorem 4.1

In Theorem 4.1, $q(H, \mathbf{p}) \geq \epsilon > 0$ was a sufficient condition for obtaining copies of $H$ in $G^R(n, \mathbf{p})$. It turns out that we can invoke Chebyshev's inequality to lower this threshold to $o(1)$. Let $\omega(n)$ denote a function that tends slowly to $\infty$ as $n \to \infty$.

**Theorem 4.4.** *If $q(H, \mathbf{p}) \geq \frac{\log \omega(n)}{\log n}$, then almost surely $G^R(n, \mathbf{p})$ contains $(1 + o(1))\frac{1}{|Aut(H)|}n^{\phi(H,\mathbf{p})}$ copies of $H$.*

Theorem 4.4 follows from the following lemma.

**Lemma 4.5.** *If $\mathbb{E}(X) = \Omega(\omega(n))$, then*

$$Pr(|X - \mathbb{E}(X)| \geq \lambda) = o(1),$$

*where $\lambda = \max\left\{\frac{1}{n}\mathbb{E}(X), \sqrt{\mathbb{E}(X)\omega(n)}\right\}$.*

*Proof.* Observe that

$$\begin{aligned}
\sum_S \mathrm{Var}\,(X_S) &= \sum_S (\mathbb{E}(X_S^2) - (\mathbb{E}(X_S))^2) \\
&\leq \sum_S \mathbb{E}(X_S^2) \\
&= \sum_S \mathbb{E}(X_S) \\
&= \mathbb{E}(X).
\end{aligned}$$

If $|S \cap S'| \leq 1$, then $X_S$ and $X_{S'}$ are independent, in which case $\mathrm{Covar}(X_S, X_{S'}) = 0$. Hence,

$$\begin{aligned}
\sum_{S \neq S'} \mathrm{Covar}(X_S, X_{S'}) &\leq \sum_{\substack{S \neq S' \\ |S \cap S'| \geq 2}} |\mathrm{Covar}(X_S, X_{S'})| \\
&= \sum_{\substack{S \neq S' \\ |S \cap S'| \geq 2}} |\mathbb{E}(X_S X_{S'}) - \mathbb{E}(X_S)\mathbb{E}(X_{S'})| \\
&\leq \sum_{S} \mathbb{E}(X_S) \sum_{\substack{S' \neq S \\ |S \cap S'| \geq 2}} |\mathbb{E}(X_{S'}|X_S) - \mathbb{E}(X_{S'})| \\
&\leq \sum_{S} \mathbb{E}(X_S) \sum_{S' \neq S} \mathbb{E}(X_{S'}|X_S) + \sum_{S} \mathbb{E}(X_S) \sum_{\substack{S' \neq S \\ |S \cap S'| \geq 2}} \mathbb{E}(X_{S'}) \\
&\leq \sum_{S} \mathbb{E}(X_S) \sum_{H'} \left( \left| n^{-\sum_{i=1}^{r} \alpha_i f_i(H')} \right| + \left| n^{-\sum_{i=1}^{r} \alpha_i f_i(H)} \right| \right) \\
&\leq \sum_{S} \mathbb{E}(X_S) \left( n^{q(H,\mathbf{p})} + \frac{1}{n^2}\mathbb{E}(X) \right) \\
&\leq \mathbb{E}(X) n^{q(H,\mathbf{p})} + \frac{1}{n^2}(\mathbb{E}(X))^2.
\end{aligned}$$

Putting these together,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}\left( \sum_{S} X_S \right) && (4.1) \\
&= \sum_{S} \mathrm{Var}\,(X_S) + \sum_{S \neq S'} \mathrm{Covar}(X_S, X_{S'}) && (4.2) \\
&\leq \mathbb{E}(X)(1 + n^{q(H,\mathbf{p})}) + \frac{1}{n^2}(\mathbb{E}(X))^2. && (4.3)
\end{aligned}$$

Now Chebyshev's inequality states that for $\lambda > 0$,

$$\Pr\left(|X - \mathbb{E}(X)| \geq \lambda\right) \leq \frac{\mathrm{Var}(X)}{\lambda^2}. \tag{4.4}$$

Choose $\lambda = \max\left\{ \frac{1}{n}\mathbb{E}(X), \sqrt{\mathbb{E}(X)\omega(n)} \right\}$. Thus $\lambda = o(\mathbb{E}(X))$. Hence, applying (4.3) to (4.4) implies

$$\begin{aligned}
\Pr\left(|X - \mathbb{E}(X)| \geq \lambda\right) &\leq \max\left\{ \frac{1}{n^2}, \frac{1}{\omega(n)} \right\} \\
&= o(1).
\end{aligned}$$

$\square$

# Bibliography


[1] N. Alon and J. Spencer, *The Probabilistic Method*, 2nd ed., Wiley, New York, 2000.

[2] N. Alon and R. Yuster, Threshold functions for $H$-factors, *Combin. Probab. Comput.* **2** (1993), 137-144.

[3] M. Axenovich, J. Manske and R.R. Martin, $\mathcal{Q}_2$-free families in the Boolean lattice, *Order* **29** (1) (2012), 177-191.

[4] R. Baber and J. Talbot, New Turán densities for 3-graphs, *Elect. J. Combin.*, (2012), P22, 21p.

[5] R. Baber and J. Talbot, Hypergraphs do jump, *Combinatorics, Probability and Computing*, **20** (2011), 161-171.

[6] S. Baroni, P. Giannozzi, and A. Testa, Green's-function approach to linear response in solids. *Physical Review Letters*, **58**(18) (1987), 1861-1864.

[7] J. Balogh, P. Hu, B. Lidický, and H. Liu, Upper bounds on the size of 4- and 6-cycle-free subgraphs of the hypercube, *European J. Combin.*, **35** (2014), 75-85.

[8] S. Bhamidi, R.W. van der Hofstad, and J.S.H. van Leeuwaarden, Scaling limits for critical inhomogeneous random graphs with finite third moments, *Electronic Journal of Probability*, **15**(54) (2010), 1682-1702.

[9] E. Boehnlein and T. Jiang, Set families with a forbidden induced subposet, *Combinatorics, Probability, and Computing*, **21** (2012), 496–511.

[10] B. Borchers, CSDP, a C library for semidefinite programming,*Optimization Methods and Software,* **11** (1) (1999), 613–623.

[11] B. Bukh, Set families with a forbidden subposet, *Electronic Journal of Combinatorics*, **16** (2009) #R142.

[12] M. Cao, H. Zhang, J. Park, N.M. Daniels, M.E. Crovella, L.J. Cowen, and B. Hescott, Going the distance for protein function prediction: a new distance metric for protein interaction networks, PLoS ONE (2013), **8**(10):e76339.

[13] T. Carroll and G.O.H. Katona, Bounds on maximal families of sets not containing three sets with $A \cup B \subset C$, $A \not\subset B$, *Order*, **25** (2008), 229–236.

[14] S.P. Chin, E. Reilly, and L. Lu. Finding structures in large-scale graphs. SPIE Defense, Security, and Sensing. International Society for Optics and Photonics, 2012.

[15] F. Chung, *Spectral graph theory*, AMS publications, 1997.

[16] F. Chung and L. Lu, *Complex graphs and networks*, CBMS Regional Conference Series in Mathematics; number 107, (2006), 264+vii pages. ISBN-10: 0-8218-3657-9, ISBN-13: 978-0-8218-3657-6.

[17] F. Chung and L. Lu. The average distances in random graphs with given expected degrees, *Proc. Natl. Acad. Sci.* **99** (2002), 15879–15882.

[18] F. Chung and L. Lu. Connected components in a random graph with given degree sequences, *Annals of Combinatorics*, **6** (2002), 125–145.

[19] F. Chung, L. Lu, and V. H. Vu, Eigenvalues of random power law graphs, *Ann. Comb.*, **7** (2003), 21–33.

[20] F. Chung, L. Lu, and V. H. Vu, Spectra of random graphs with given expected degrees, *Proc. Natl. Acad. Sci. USA*, **100**(11) (2003), 6313–6318.

[21] F. Chung, L. Lu, An upper bound for the Turán number $t_3(n, 4)$. *J. Combin. Theory Ser. A* **87** (1999), no. 2, 381–389.

[22] F. Chung and S.-T. Yau, Discrete Green's functions, *J. Combinatorial Theory (A)* **91** (2000), 191-214.

[23] F. Chung and M. Radcliffe, On the spectra of general random graphs, *Electron. J. Combin.*, **18**(1) (2011), P215.

[24] C. Cooper, A. Frieze, M. Molloy, and B. Reed. Perfect matchings in random $r$-regular, $s$-uniform hypergraphs. *Combin. Probab. Comput.* **5** (1996), 1-14.

[25] A. Coja-Oghlan, On the Laplacian eigenvalues of $G(n,p)$, *Combin. Probab. Comput.*, **16**(6) (2007), 923–946.

[26] A. Coja-Oghlan and A. Lanka, The spectral gap of random graphs with given expected degrees, *Electron. J. Combin.*, **16** (2009), R138.

[27] E.B. Davies, *Heat Kernels and Spectral Theory*, Vol. 92, Cambridge University Press, 1990.

[28] A. De Bonis and G.O.H. Katona, Largest families without an $r$-fork, *Order*, **24** (2007), 181–191.

[29] A. De Bonis, G.O.H. Katona, and K. Swanepoel, Largest family without $A \cup B \subseteq C \cap D$ *J. Combin. Theory Ser. A*, **111** (2005), 331–336.

[30] D. de Caen, The current status of Turán's problem on hypergraphs, *Extremal problems for finite sets* (Visegrád, 1991), 187–197, Bolyai Soc. Math. Stud., 3, János Bolyai Math. Soc., Budapest, 1994.

[31] A. Dudek and A.M. Frieze, Loose Hamilton cycles in random uniform hypergraphs. *Electronic Journal of Combinatorics* **18** (2011), P48.

[32] D.G. Duffy, *Green's Functions with Applications*, CRC Press, 2010.

[33] P. Erdős, On the combinatorial problems which I would like to see solved, *Combinatorica* **1** (1981), 25–42.

[34] P. Erdős, On a lemma of Littlewood and Offord, *Bulletin of the American Mathematical Society*, **51** (1945), 898–902.

[35] P. Erdős, and A. Rényi, On Random Graphs I, *Publ. Math Debrecen* **6** (1959), 290–297.

[36] P. Erdős, M. Simonovits, Supersaturated graphs and hypergraphs. *Combinatorica* **3** (1983), 181–192.

[37] P. Frankl, Z. Füredi, A new generalization of the Erdős-Ko-Rado theorem, *Combinatorica* **3** (1983), 341–349.

[38] P. Frankl, Z. Füredi, An exact result for 3-graphs, *Discrete Math.* **50** (1984), 323–328.

[39] Z. Füredi, D. Mubayi, and O. Pikhurko, Quadruple systems with independent neighborhoods, *J. Comb. Theory Ser. A* **115** (2008), 1552–1560.

[40] Z. Füredi and M. Simonovits, Triple systems not containing a Fano configuration, *Combin. Probab. Comput.* **14** (2005), 467–484.

[41] G. Green, *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism*, Nottingham, 1828.

[42] J.R. Griggs and G.O.H. Katona, No four subsets forming an N, *J. Combin. Theory Ser. A*, **115** (2008): 677–685.

[43] J.R. Griggs, W.-T. Li, and L. Lu, Diamond-free families, *J. Combin. Th. Ser. A*, **119**(2) (2012), 310–322.

[44] J.R. Griggs and L. Lu, On families of subsets with a forbidden subposet, *Combinatorics, Probability, and Computing*, **18** (2009), 731–748.

[45] D. Grósz, A. Methuku, and C. Tompkins, An upper bound on the size of diamond-free families of sets, *ArXiv*:1601.06332, 2016.

[46] L. Harriger, M. P. van den Heuvel, and O. Sporns, Rich club organization of macaque cerebral cortex and its role in network communication, PloS one 7.9 (2012): e46497.

[47] L. Hedin, New method for calculating the one-particle Green's function with application to the electron-gas problem, *Physical Review*, **139**(3A) (1965), 796-823.

[48] R.W. van der Hofstad, Critical behavior in inhomogeneous random graphs, *Random Structures and Algorithms*, **42**(4) (2013), 480-508.

[49] J. Johnston, L. Lu, Turán Problems on Non-uniform Hypergraphs, arxiv:1301.1870 [math.CO].

[50] G.O.H. Katona and T.G. Tarján, Extremal problems with excluded subgraphs in the $n$-cube,*Lecture Notes in Mathematics*, Springer, **1018** (1983), 84–93.

[51] P. Keevash, Hypergraph Turán problems, *Surveys in Combinatorics*, Cambridge University Press (2011), 80-140.

[52] P. Keevash and B. Sudakov, The Turán number of the Fano plane, *Combinatorica* **25** (2005), 561–574.

[53] P. Keevash and B. Sudakov, On a hypergraph Turán problem of Frankl, *Combinatorica* **25** (2005), 673–706.

[54] J.H. Kim, Perfect matchings in random uniform hypergraphs, *Random Structures & Algorithms* **23** (2003), 111–132.

[55] J.H. Kim, V. Vu, Concentration of multi-variate polynomials and its applications, *Combinatorica*, **20** (3) (2000), 417–434.

[56] L. Kramer, On diamond-free subposets of the Boolean lattice: An application of flag algebras, *Graduate Theses and Dissertations*, Paper 13894, 2014, http://lib.dr.iastate.edu/etd/13894.

[57] L. Kramer, R.R. Martin and M. Young, On diamond-free subposets of the Boolean lattice, *J. Combin. Theory Ser. A* **120** (3) (2013), 545–560.

[58] M. Krivelevich and B. Sudakov, The largest eigenvalue of sparse random graphs, *Combin. Probab. Comput.*, **12** (2003), 61–72.

[59] L. Lu, On crown-free families of subsets, *J. Combin. Theory Ser. A.*, **126** (2014), 216–231.

[60] L. Lu and X. Peng, Spectra of edge-independent random graphs, *Electronic Journal of Combinatorics*, **20** (4) (2013), P27.

[61] D. Lubell, A short proof of Sperner's lemma, *J. Combin. Theory* **1** (1966), 299.

[62] O. Pikhurko, Exact computation of the hypergraph Turán function for expanded complete 2-graphs, Accepted by *J. Combin. Theory Ser. B* **103** (2013), 220–225.

[63] A. Razborov, On 3-hypergraphs with forbidden 4-vertex configurations, *SIAM J. Disc. Math.* **24** (2010), 946–963.

[64] A. Razborov, Flag algebras, *Journal of Symbolic Logic*, **72**(4) (2007), 1239-1282.

[65] M.A. de Reus and M.P. van den Heuvel, Rich club organization and intermodule communication in the cat connectome, *The Journal of Neuroscience*, 33.32 (2013), 12929-12939.

[66] J. Spencer, Counting extensions, *J. Comb. Theory Series A*, **55** (1990), 247-255.

[67] J. Spencer, Threshold functions for extension statements, *J. Comb. Theory Series A* **53** (1990), 286–305.

[68] E. Sperner, (1928) Ein Satz uber Untermengen einer endlichen Menge, *Math. Z.* **27** (1928), 544–548.

[69] I. Stakgold and M.J. Holst, *Green's Functions and Boundary Value Problems*, John Wiley & Sons, 2011.

[70] W. Thompson, *Reprint of papers on electrostatics and magnetism*, Macmillian (1872).

[71] W. Thompson and P.G. Tait, *Treatise on natural philosophu, Part I, Part II*, Cambridge U.P. (1879, 1903).

[72] P. Turán, On an extremal problem in graph theory, *Mat. Fiz. Lapok* **48** (1941), 436–452.

[73] Open Connectome Project, Web. 25 June 2014, http://www.openconnectomeproject.org.

# Appendix A

# Permission to Reprint

Table A.1 Springer License Terms and Conditions

| | |
|---|---|
| License Number | 3920180571251 |
| License Date | Jul 27, 2016 |
| Licensed Content Publisher | Springer |
| Licensed Content Publication | Springer eBook |
| Licensed Content Title | Computing Diffusion State Distance Using Green's Function and Heat Kernel on Graphs |
| Licensed Content Author | Edward Boehnlein |
| Licensed Content Date | Jan 1, 2014 |
| Type of Use | I don't see my intended use |
| Requestor Type | University |
| Portion | Full text |
| Format | Print and Electronic |
| Will you be translating? | No |
| Print Run | 6 |
| Describe how you will be using this material | This article is a chapter in my dissertation, which my university (U. of South Carolina) will disseminate in UMI/ProQuest's database, and USC's institutional repository, Scholar Commons. |
| Author of this Springer Article | Yes and you are the sole author of the new work |
| Requestor Location | Edward Boehnlein<br>COLUMBIA, SC 29208<br>United States<br>Attn: Edward Boehnlein |
| Billing Type | Invoice |
| Billing Address | Edward Boehnlein<br>COLUMBIA, SC 29208<br>United States<br>Attn: Edward Boehnlein |
| Total | 0.00 USD |
| Terms and Conditions | |

Introduction

The publisher for this copyrighted material is Springer. By clicking “accept” in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. (”CCC”), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License

With reference to your request to reuse material on which Springer controls the copyright, permission is granted for the use indicated in your enquiry under the following conditions:

- Licenses are for one-time use only with a maximum distribution equal to the number stated in your request.
- Springer material represents original material which does not carry references to other sources. If the material in question appears with a credit to another source, this permission is not valid and authorization has to be obtained from the original copyright holder.
- This permission
  - is non-exclusive
  - is only valid if no personal rights, trademarks, or competitive products are infringed.
  - explicitly excludes the right for derivatives.
- Springer does not supply original artwork or content.

- According to the format which you have selected, the following conditions apply accordingly:

  - Print and Electronic: This License include use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.
  - Print: This License excludes use in electronic form.
  - Electronic: This License only pertains to use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.

  For any electronic use not mentioned, please contact Springer at permissions.springer@spi-global.com.

- Although Springer controls the copyright to the material and is entitled to negotiate on rights, this license is only valid subject to courtesy information to the author (address is given in the article/chapter).

- If you are an STM Signatory or your work will be published by an STM Signatory and you are requesting to reuse figures/tables/illustrations or single text extracts, permission is granted according to STM Permissions Guidelines: http://www.stm-assoc.org/permissions-guidelines/. For any electronic use not mentioned in the Guidelines, please contact Springer at permissions.springer@spi-global.com. If you request to reuse more content than stipulated in the STM Permissions Guidelines, you will be charged a permission fee for the excess content.

  Permission is valid upon payment of the fee as indicated in the licensing process. If permission is granted free of charge on this occasion, that does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

- If your request is for reuse in a Thesis, permission is granted free of charge under the following conditions:

  This license is valid for one-time use only for the purpose of defending your thesis and with a maximum of 100 extra copies in paper. If the thesis is going to be published, permission needs to be reobtained.

  - includes use in an electronic form, provided it is an author-created version of the thesis on his/her own website and his/her university's repository, including UMI (according to the definition on the Sherpa website: http://www.sherpa.ac.uk/romeo/);
  - is subject to courtesy information to the co-author or corresponding author.

Geographic Rights: Scope
Licenses may be exercised anywhere in the world.

Altering/Modifying Material: Not Permitted
Figures, tables, and illustrations may be altered minimally to serve your work. You may not alter or modify text in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s).

Reservation of Rights
Springer reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction and (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and

conditions.

License Contingent on Payment

While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Springer or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received by the date due, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Springer reserves the right to take any and all action to protect its copyright in the materials.

Copyright Notice: Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer book/journal title, chapter/article title, volume, year of publication, page, name(s) of author(s), (original copyright notice as given in the publication in which the material was originally published) "With permission of Springer" In case of use of a graph or illustration, the caption of the graph or illustration must be included, as it is indicated in the original publication.

Warranties: None

Springer makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity

You hereby indemnify and agree to hold harmless Springer and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you without Springer's written permission.

No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case of Springer, by CCC on Springer's behalf).

Objection to Contrary Terms

Springer hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other conditions:

V 12AUG2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.