PROPERTIES OF RANDOM THRESHOLD AND DIFFERENCE GRAPHS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By
Christopher Ross, MS
Graduate Program in Mathematics

The Ohio State University

2011

Dissertation Committee:
Dr. Boris Pittel, Advisor
Dr. G. Neil Robertson
Dr. Warren Sinnott
ABSTRACT

Generate a graph on $n$ vertices by randomly assigning to each vertex $v$ some weight $w(v) \in [0, 1]$, where each weight is taken independently and identically on the interval, and adding an edge between vertices $v_i$ and $v_j$ if and only if $|w(v_i) + w(v_j)| > 1$; the results of such processes are known as threshold graphs. Similarly, if we first partition the vertices into two sets $A$ and $B$, and only permit edges between vertices of different sets, then the result is a difference graph.

On these two probability spaces, we examine the behavior of the respective classes of graphs by finding the distribution of several graph invariants, such as the matching number, connectivity, and length of the longest cycle. We are aided in this task by an additional result which permits us to translate between the continuous sample spaces given above and the discrete probability space in which each such graph is chosen uniformly at random.
To my wife, Rachel
I owe my deepest debt of gratitude to my advisor, Boris Pittel. Not only did he instruct me in many of the techniques and approaches used herein, but he also introduced me to the threshold and difference graphs which have formed the core of my doctoral research. It is through his constant support, encouragement, and guidance that both my work and myself have developed to their current point.

I would also like to thank my fellow students, in particular Huseyin Acan, Nicholas Peterson, Daniel Poole, and Robert Woodruff, for their comments, questions, and criticism over many months of research. I am similarly appreciative of my committee members, Dr. Robertson and Dr. Sinnott, and their willingness to assist in the production of my dissertation.

Finally, I would like to thank the entire faculty and staff of Ohio State University, especially the Mathematics Department, who made this all possible.
VITA

1983 ................................. Born - Gaithersburg, MD

2005 ................................. B.S. in Mathematics, Rutgers University

2008 ................................. M.S. in Mathematics, Ohio State University

2005-Present ........................ Graduate Teaching Associate,

The Ohio State University

PUBLICATIONS

FIELDS OF STUDY

Major Field: Mathematics

Specialization: Probabilistic Combinatorics
# TABLE OF CONTENTS

Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ii
Dedication . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ii
Acknowledgments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iv
Vita . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . v
List of Figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . viii

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>2.1</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>9</td>
</tr>
<tr>
<td>2.3</td>
<td>12</td>
</tr>
<tr>
<td>2.4</td>
<td>13</td>
</tr>
<tr>
<td>2.5</td>
<td>18</td>
</tr>
<tr>
<td>2.6</td>
<td>26</td>
</tr>
<tr>
<td>2.7</td>
<td>29</td>
</tr>
<tr>
<td>2.8</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>39</td>
</tr>
<tr>
<td>3.4</td>
<td>41</td>
</tr>
<tr>
<td>3.5</td>
<td>43</td>
</tr>
<tr>
<td>3.6</td>
<td>45</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>6</td>
</tr>
<tr>
<td>1.3</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>19</td>
</tr>
<tr>
<td>2.3</td>
<td>20</td>
</tr>
<tr>
<td>2.4</td>
<td>21</td>
</tr>
<tr>
<td>4.1</td>
<td>67</td>
</tr>
<tr>
<td>4.2</td>
<td>68</td>
</tr>
<tr>
<td>4.3</td>
<td>68</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Preliminaries

Here we shall investigate two extremely similar structures, threshold graphs and difference graphs. We study them not for practical reasons, although these classes have applications in various fields, from computer science and operations research to education and mathematical psychology, but for their versatility. That is, both of these graphs represent transitions from relationships of several continuous variables to seemingly unrelated, purely discrete combinatorial structures. In fact, one of the major themes for these transformations is that of comparison, as exhibited by the following methods of generating threshold graphs:

Consider the first - and most straightforward - definition of a threshold graph. For each vertex $v$, we define some weight $w(v)$, then add an edge between distinct vertices $v_i$ and $v_j$ whenever the sum $w(v_i) + w(v_j)$ exceeds some threshold value $t$. While there are multiple ways to define $w$ and $t$, the entire graph is determined by comparing the weight sums to the threshold.

Alternatively, one can “grow” a threshold graph through an iterative process. Starting with an graph on a single vertex, introduce additional vertices one at a time; as each is added, we adjoin it to either all or none of the earlier vertices. After $n - 1$ stages, the result is a threshold graph on $n$ vertices, and as the graph resulted from a
series of \( n-1 \) binary decisions, we can “encode” the entire process as a binary sequence of \( n-1 \) digits. Here, the relationships between the vertex weights are determined by the process itself, for those vertices which are isolated from the previously-introduced brethren are of lesser weight, whereas those that are dominating must have greater weight. As such, comparisons between vertices are not so much made as defined.

A third method of constructing a threshold graph is by taking some existing system and finding the threshold graph that models the data. Specifically, this requires two sets of objects, denoted \( A \) and \( B \), where the elements of \( B \) are subsets of \( A \), and totally ordered by inclusion. (Example: \( A \) could be a set of students, and \( B \) could be the set

\[
B = \begin{cases}
    \text{Students that passed,} \\
    \text{Students that earned at least a 'B',} \\
    \text{Students that earned at least an 'A'}
\end{cases}
\]

which obviously satisfies the order restriction.) Then we define the vertex set to be the union \( A \cup B \), and add an edge between two vertices if both are elements of \( B \), or if one is an element of the other. (Returning to our example, the vertices representing sets of students would all be mutually adjacent, but a student-vertex is adjacent to a set-of-students-vertex if and only if said student is a member of the set.) Here, the comparisons are made in the construction of the set \( B \), which by its very nature is a process that winnows out elements of \( A \).

So of these three methods for generating a threshold graph, using vertex weights, iterative growth, or conversion of a model, all have their roots in the comparison of vertex elements. (Difference graphs, which behave similarly, have analoguous formation processes, and exhibit the same characteristics.) Later, we introduce methods of randomly generating these graphs, and examine the results of these processes. If we were to take a randomly-chosen threshold graph, for some value of “random”, }
would we expect it to be connected? Planar? What would be the size of the largest matching?

1.2 History

As threshold graphs can be obtained through several processes, some drastically dis-similar from others, it should come as no surprise that the idea has been independently “discovered” multiple times. However, the term itself, if not the common interpretation, stems from Chvátal and Hammer’s 1973 paper, “Set-packing problems and threshold graphs” [2]. Similarly, the notion of a difference graph was formally introduced in 1990 by Hammer, Peled, and Sun [8], although functionally equivalent structures had been under investigation for almost a decade prior.

In 1995, Mahadev and Peled compiled many of these disparate results into Threshold Graphs and Related Topics [12]. For example, they found no fewer than eight different characterizations of threshold graphs; fortunately, we need only concern ourselves with the three described above.

1.3 Applications and Examples

As we saw in the students-and-grades example, above, threshold graphs are extremely adept for modeling certain relationships. Formally speaking, let us consider the notion of a Guttman scale [7], a set of ordered survey items with binary responses such that a respondent’s response to the entire set can be encoded by a particular item. For example, consider this excerpt of the Bogardus Social Distance Scale [1], in use since 1925:

1. Would you be accepting of immigrants as relatives by marriage?

2. Would you be accepting of immigrants as close personal friends?
3. Would you be accepting of immigrants as neighbors on the same street?

4. Would you be accepting of immigrants as co-workers in the same occupation?

5. Would you be accepting of immigrants as citizens in your country?

Here, a response of “yes” to item three would necessarily imply “yes” to items four and five, while conveying no information about items one and two. So a subject’s answers to all five questions could be imparted by giving the lowest-numbered question to which he answered “yes”. (If someone answered “no” to all five, then his answers could be represented by the value six.)

Mathematically, we can say that a Guttman scale is just a set whose elements form a chain under inclusion. As before, we can create a threshold graph by adding one vertex for every question and one vertex for each respondent, adjoining each pair of questions, and adjoining each recipient to every question for which he responded in the affirmative.

To illustrate this, we shall eschew use of the somewhat political scale above in favor of the grade-based scale that we touched upon earlier:

1. Did you get at least an ‘A’?

2. Did you get at least a ‘B’?

3. Did you get at least a ‘C’?

4. Did you get at least a ‘D’?

Which, in a 2009 class of 21 students, has given rise to the graph shown in Figure 1.1.

In this manner, the same techniques we will use to generate and analyze random threshold graphs can be applied to the simulations of survey data. Admittedly,
this particular representation of the information does not benefit from any graph-theoretical concepts, serving mainly as an illustration of a concept, so let us now turn to more involved scenario.

In 1984, Doignon, Ducamp, and Falmagne [4] took the concept of a Guttman scale and doubled it. Whereas threshold graphs model the relationships between elements of a Guttman scale and some independent (in several senses) set, the researchers defined a biorder, which serves to simultaneously relate the elements of two chains. After a half-dozen years, these biorders would be identified with the newly-defined difference graphs.

For example, suppose that you are running a small mathematics department, and are trying to assign instructors to classes. There are four different courses, Math 101 through Math 104; as these courses cumulatively build on each other, anyone capable of teaching a class is also able to teach those classes of lower level. There are six possible instructors, labelled A through F in decreasing order of ability. Due to this
order, if a particular instructor can teach a given course, so too can all instructors that alphabetically precede him.

Consider the graphical representation of such a system, found in Figure 1.2, where an edge between a course and an instructor indicates a potentially valid assignment:

Now, with a bit of rearrangement, we can express this bipartite graph as one in which all edges lead up and to the right, as in Figure 1.3.

This second representation takes the linear orders inherent to the individual parts and combines them into a single ordering, using the existence of edges as a comparator on inter-part pairings. Using the representation in Figure 1.3, we read the vertices from left to right, replacing all of the instructors with zeroes and all of the classes as
ones. Thus, we can restate the entire graph as a single sequence, given in Figure 1.4.

\[0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1\]

Figure 1.3: Class Assignments, Representation 2

Figure 1.4: Class Assignments, Representation 3

From Figure 1.3, we see that the size of the maximum matching in the graph is three, so regardless of how you assign your faculty, there will be at least one class without appropriate instruction. (We will investigate difference graph matchings further in Chapter Four.) While this particular example is admittedly rather contrived, threshold and difference graphs have been applied to problems of parallel processing and resource allocation for several decades.
CHAPTER 2
THRESHOLD GRAPH PROPERTIES

2.1 Background

As we have already seen, there are several ways to randomly generate a threshold
graph on \( n \) vertices. In 2009, Reilly and Scheinerman showed that two of these
options, either by randomly taking \( n \) vertex weights uniformly on \([0, 1]\) or by choosing
a creation sequence uniformly from the set of all \((n - 1)\)-digit binary sequences, were
equivalent. That is, the probability of generating a particular threshold graph was
the same regardless of the method of production.

Using this, they were able to detect the presence or absence of various graph prop-
erties by analysis of the creation sequences. These properties included connectivity,
Hamiltonicity, perfect matchings, and more. In this chapter, we use their primary re-
sult about model equivalence to expand upon their discoveries, finding the likelihoods
and distributions of other invariants.

For example, if we want to find the probability of planarity, we can discard all
information about vertex weights. Instead, we look at what properties must be true
of the creation sequence in order for the graph to be planar, and count the number of
sequences that have said properties. Since we choose the creation sequences uniformly
at random, the probability of planarity would be the number of satisfactory sequences
divided by the total number of sequences.
2.2 Fundamentals

Before we can begin to discuss the existing research, let alone tread new ground, it is necessary to discuss some of the more basic and well-known properties of threshold graphs. For now we shall take as our primary definition the method suggested by the name of the class of graphs: we assign to each vertex \( v_i \) a weight \( w(v_i) \), and draw an edge between vertices \( v_i \) and \( v_j \) if and only if \( w(v_i) + w(v_j) > t \), where \( t \) is some threshold. Specifically, each weight is drawn independently and uniformly at random from the interval \([0, 1]\), and the threshold is \( t = 1 \).

This definition leads immediately to the first property:

**Proposition 2.2.1.** If \( G \) is a threshold graph, then every induced subgraph of \( G \) is a threshold graph.

*Proof.* Let \( A \subset V(G) \) be any subset of \( G \)'s vertices, and let \( H \) be the subgraph induced by \( A \).

As \( G \) is a threshold graph, there exists some weighting function \( w \) and threshold \( t \) such that for all \( v_i, v_j \in V(G) \), \( \{v_i, v_j\} \in E(G) \) if and only if \( w(v_i) + w(v_j) > t \). Then as \( H \) is an induced subgraph of \( G \), we see that for all \( v_i, v_j \in A \), \( \{v_i, v_j\} \in E(H) \) if and only if \( w(v_i) + w(v_j) > t \). And thus, \( H \) is a threshold graph as it inherits the same weights as \( G \). \( \square \)

**Proposition 2.2.2.** If \( G \) is a threshold graph, then \( G \) contains either a dominating vertex or an isolated vertex.

*Proof.* Let \( v_{\min} \) denote a vertex of \( G \) of minimum weight, and similarly, we define \( v_{\max} \) to be a vertex of maximum weight. Then there are two possible cases:

If \( w(v_{\max}) + w(v_{\min}) > t \), then for all vertices \( v_i \),

\[ w(v_{\max}) + w(v_i) \geq w(v_{\max}) + w(v_{\min}) > t, \]
and therefore \( v_{\text{max}} \) is adjacent to \( v_i \).

Alternatively, if \( w(v_{\text{max}}) + w(v_{\text{min}}) \leq t \), then for all vertices \( v_i \),

\[
 w(v_{\text{min}}) + w(v_i) \leq w(v_{\text{min}}) + w(v_{\text{max}}) \leq t,
\]

and therefore \( v_{\text{min}} \) is not adjacent to \( v_i \).

So in every threshold graph, there exists either a vertex that is isolated, or one that dominates the entire graph.

Note that in the case when \( G \) has but a single vertex, said vertex is both dominating and isolated, but for any larger threshold graph, exactly one of the conditions must be true. Thus, we can form any threshold graph on \( n \) vertices by sequentially adding \( n - 1 \) vertices, and deciding with each addition whether the new vertex is to be dominating or isolated.

During this “growth” process, we record our \( n - 1 \) decisions in the form of a binary sequence, marking a “0” whenever the new vertex is isolated and a “1” when it dominates. In this fashion, the entire threshold graph can be encoded in an \( n - 1 \)-digit string called the creation sequence.

To clarify, when given a threshold graph, one can recursively find the creation sequence by determining whether there is an isolated or dominating vertex, noting a “0” or “1”, respectively, and then appending that digit to the creation sequence for the subgraph induced on the remaining vertices. Similarly, when given a creation sequence, one can build the graph by starting with a single vertex and reading the sequence from left to right, adding either an isolated or dominating vertex at each step.

As an example of the encoding process, we turn to Figure 2.1. Note that the original graph has an isolated vertex, denoted \( B \), which tells us that the creation sequence ends with a 0. Removing \( B \), the induced subgraph has dominating vertex
A, so the penultimate digit is a 1. After removing both A and B, the resulting graph has two isolated vertices; removing them in either order, we have a pair of zeroes in the creation sequence.

![Figure 2.1: Threshold Example 1](image)

Thus, by recording these digits from right to left, we see that the corresponding creation sequence is 0110010. By the reverse process, when given such a sequence, reading it from left to right allows us to reconstruct the graph.

Due to this bijection between the set of \( n \)-vertex unlabelled threshold graphs and the set of \( n - 1 \)-digit binary sequences, we see that the number of threshold graphs on \( n \) vertices is exactly \( 2^{n-1} \).
For simplicity, we let the terms zero-vertices and one-vertices represent those vertices which correspond to the digits “0” and “1”, respectively, in the creation sequence. Note that, as all zeroes correspond to isolated vertices, no two zero-vertices can ever be adjacent. Similarly, as ones represent dominating vertices, any two one-vertices are automatically adjacent. And a zero-vertex is only adjacent to a one-vertex if the latter was added after the former; that is, when the “1” lies to the right of the “0”.

2.3 Definitions

For a random threshold graph $G_n$ with corresponding creation sequence $C_n$, we define random variables $U_n$ and $Z_n$ to be the numbers of ones and zeroes, respectively, in $C_n$. We can then enumerate the vertices according to their position in the creation sequence: for $i \in \{1, \ldots, Z_n\}$, we let $z_i$ denote the vertex corresponding to the $i$-th zero in $C_n$, and for $j \in \{1, \ldots, U_n\}$, $u_j$ denotes the vertex corresponding to the $j$-th one. Labeling the base vertex, which lacks a digit, as $v_0$, we have the following decomposition of the vertex set:

$$V(G_n) = \{v_0\} \cup \{u_1, \ldots, u_{U_n}\} \cup \{z_1, \ldots, z_{Z_n}\}$$

Please note that although we are dealing with unlabeled graphs, these labels are necessary for many of the proofs to come. The distinction is that the assignment of labels is a consequence of the graph, rather than an independent property. It is simply far more convenient to use “$z_i$” rather than “vertex of $G_n$ corresponding to the $i$-th zero from the left in $C_n$”.

To better describe $C_n$, we also introduce the notion of the index of a vertex $v$, which is the position of the corresponding digit in the creation sequence. For example, the index of $v_0$ would be zero, whereas the index of the one-vertex in the graph given
by 00010 would be four, and the index of the isolated vertex in Figure 2.1 would be seven.

It will also prove useful to examine the “tail” of a sequence \( S \), which consists of the subsequence of \( S \) that includes all digits after a certain index. In particular, if \( S \) is the \( n \)-long sequence of the form

\[
S = s_1 s_2 \ldots s_{n-1} s_n,
\]

then the \( i \)-th tail of \( S \) is the sequence \( s_i s_{i+1} \ldots s_n \). Similarly, let \( t^z_i(S) \) and \( t^u_i(S) \) denote the number of zeroes and ones, respectively, in the \( i \)-th tail of \( S \).

Thus equipped, we are now prepared to explore new properties of threshold graphs.

### 2.4 Matching Number

Given a graph \( G \), the matching number \( \nu(G) \) denotes the size (in edges) of the maximum matching in \( G \). Reilly and Scheinerman showed that the probability of a random threshold graph on \( n \) vertices having a perfect matching is approximately \((\pi n/2)^{-1/2}\). But what about the probability that the largest matching is of some other size?

Reilly and Scheinerman proved that for even values of \( n \), \( \nu(G_n) = \frac{n}{2} \) if and only if \( t^u_k(C_n) \geq t^z_k(C_n) \) for all \( k \). That is, there is a perfect matching if and only if every tail of the creation sequence contains at least as many ones as zeroes. In this vein, let us define a function \( h \) on the set of all finite binary sequences by, for any finite binary sequence \( S \),

\[
h(S) = \max_{0 \leq i \leq |S|+1} \{ t^z_i(S) - t^u_i(S) \}
\]

Note that \( h \) is always non-negative, as the tail corresponding to \( i = |S| + 1 \) is the empty sequence, without ones or zeroes. We will be making repeated use of the following lemma about \( h \):
Lemma 2.4.1. Let $S$ be a binary sequence with $h(S) > 0$, and let $S'$ be the binary sequence formed by removing the right-most $h(S)$ zeroes from $S$. Then $h(S') = 0$.

Proof. Suppose, for the sake of contradiction, that $h(S') > 0$. Then there exists some maximizing index $m$ such that

$$h(S') = t_m^z(S') - t_m^u(S') > 0$$

Then $t_m^z(S') \geq 1$, so there exists a zero in the $m$-th tail of $S'$. Since the right-most $h(S)$ zeroes were removed from $S$ to form $S'$, and zeroes within the $m$-th tail of $S'$ were not removed, this means that all of the removed zeroes must be in the $m$-th tail of $S$. And because the $m$-th tails of each sequence contain the same number of ones,

$$t_m^z(S) - t_m^u(S) = (t_m^z(S') + h(S)) - t_m^u(S') > h(S),$$

which contradicts the maximality of $h(S)$.

Observe that $h$ permits a generalization of the aforementioned requirement for a perfect matching. Specifically, when $h(C_n) = 0$, every tail has at least as many ones as zeroes, and there is a perfect matching, but as $h$ increases, so too does the number of “excess zeroes” across the tails of $C_n$, and we move further from the ideal case.

Proposition 2.4.2. For a threshold graph $G_n$ with creation sequence $C_n$,

$$\nu(G_n) = \left\lfloor \frac{n - h(C_n)}{2} \right\rfloor$$

Proof. Note that the case $h(C_n) = 0$ corresponds to the earlier result by Reilly and Scheinerman. Thus, we restrict ourselves to the case $h(C_n) \geq 1$.

As $2\nu(G_n)$ is the maximum number of vertices in a matching, $n - 2\nu(G_n)$ represents the number of vertices not present in any maximum matching.

Let $m$ denote a maximizing index such that $h(C_n) = t_m^z(C_n) - t_m^u(C_n)$. Then there are $h(C_n)$ more zeroes than ones in the $m$-th tail. Since those zeroes can only
be adjacent to ones in that tail, there are at least $h(C_n)$ vertices that cannot participate in any maximum matching. Therefore, $n - 2\nu(G_n) \geq h(C_n)$, and we see that

$$\nu(G_n) \leq \frac{n - h(C_n)}{2}$$

Next, let us define a new threshold graph $G'$ by removing the $h(C_n)$ zero-vertices of highest index from $G_n$. Since any induced subgraph of a threshold graph is another threshold graph, $G'$ is a threshold graph on $n - h(C_n)$ vertices. Then the creation sequence for $G'$, denoted $C'$, is acquired by removing the right-most $h(C_n)$ zeroes from $C_n$.

By Lemma 2.4.1, $h(C') = 0$, so by the Reilly and Scheinerman result, $G'$ has either a perfect matching or almost-perfect matching, depending upon whether $n - h(C_n)$ is even or odd, respectively. Thus,

$$\nu(G_n) \geq \left\lfloor \frac{n - h(C_n)}{2} \right\rfloor$$

And as $\nu(G_n)$ must be an integer, we see that

$$\nu(G_n) = \left\lfloor \frac{n - h(C_n)}{2} \right\rfloor$$

With this, we can calculate the distribution of $\nu(G_n)$ by finding that of $h(C_n)$. Note that for $h(C_n)$ to equal $k$, there must be a sequence of $n$ binary digits such that some tail has exactly $k$ more zeroes than ones, but in no tail is there ever $k + 1$ more zeroes than ones.

In order to count the number of such sequences, we instead interpret the binary sequences as moves on an integer lattice. Starting at the origin, we move a single unit upwards whenever we encounter a zero, and rightwards when we encounter a one. If our current position is the point $(a, b)$, then we have read $a + b$ digits, of which there
were $b - a$ more zeroes than ones. Such sequences of moves, known as “monotonic paths” or “staircase walks”, have been studied in great detail, and we can use existing results (Vilenkin [15]).

Lemma 2.4.3. Let $a_2(u, v, k)$, for $v - u \leq k$, denote the number of staircase walks from $(0, 0)$ to $(u, v)$ that touch, but do not cross, the line $y = x + k$. Then

$$a_2(u, v, k) = \binom{u + v}{u + k} - \binom{u + v}{u + k + 1}$$

Lemma 2.4.4. Let $a_1(n, k)$ denote the number of staircase walks of length $n$ from $(0, 0)$ that touch, but do not cross, the line $y = x + k$. Then

$$a_1(n, k) = \binom{n}{n + k - \lfloor \frac{n+k}{2} \rfloor}$$

We now apply the latter lemma to find the distribution of $h(C_n)$:

Proposition 2.4.5. For a random threshold graph $G_n$ with creation sequence $C_n$,

$$P(h(C_n) = k) = \left( \frac{1}{2} \right)^{n-1} \binom{n - 1}{\lfloor \frac{n}{2} \rfloor}$$

Proof. Using the uniformity of the distribution, the probability that $h(C_n) = k$ is proportional to the number of $(n - 1)$-long binary sequences that have a tail with $k$ more zeroes than ones, but no tail has $k + 1$ more zeroes than ones. By reading the creation sequences from right to left, we see that this is equal to the number of staircase walks of length $n - 1$ that touch, but do not cross, the line $y = x + k$. So by Lemma 2.4.4,

$$P(h(C_n) = k) = \left( \frac{1}{2} \right)^{n-1} \binom{n - 1}{n - 1 + k - \lfloor \frac{n+k}{2} \rfloor}$$

Note that in the special case of $k = 0$, we get

$$P(h(C_n) = 0) = \left( \frac{1}{2} \right)^{n-1} \binom{n - 1}{n - 1 - \lfloor \frac{n-1}{2} \rfloor} = \left( \frac{1}{2} \right)^{n-1} \binom{n - 1}{\lfloor \frac{n-1}{2} \rfloor}$$
This concurs with the prior result of Reilly and Scheinerman. But we now return to the more general case:

**Theorem 2.4.6.** For a random threshold graph $G_n$,

$$P(\nu(G_n) = k) = \begin{cases} 
\left(\frac{1}{2}\right)^{n-1} \binom{n}{k} & k < \frac{n}{2} \\
\left(\frac{1}{2}\right)^{n-1} \binom{n-1}{\left\lfloor \frac{n-1}{2} \right\rfloor} & k = \frac{n}{2} 
\end{cases}$$

**Proof.** Recalling that

$$\nu(G_n) = \left\lfloor \frac{n - h(C_n)}{2} \right\rfloor,$$

we see that since $h(C_n)$ can only assume integer values,

$$P(\nu(G_n) = k) = P(n - 2k - 1 \leq h(C_n) \leq n - 2k)$$

$$= P(h(C_n) = n - 2k - 1) + P(h(C_n) = n - 2k)$$

And as $h(C_n)$ must be non-negative, we see that for $0 \leq k < n/2$,

$$P(\nu(G_n) = k) = \left(\frac{1}{2}\right)^{n-1} \left(\binom{n-1}{\left\lfloor \frac{n-1}{2} \right\rfloor} + \binom{n-1}{\left\lfloor \frac{2n-2k+1}{2} \right\rfloor}\right)$$

$$= \left(\frac{1}{2}\right)^{n-1} \left(\binom{n-1}{k} + \binom{n-1}{k-\left\lfloor \frac{2n-2k+1}{2} \right\rfloor}\right)$$

$$= \left(\frac{1}{2}\right)^{n-1} \left(\binom{n-1}{k} + \binom{n-1}{k-1}\right)$$

$$= \left(\frac{1}{2}\right)^{n-1} \binom{n}{k}$$

As for $k = n/2$, which can occur only when $n$ is even, this corresponds to there being a perfect matching, the probability of which has already been computed. 

\[\square\]
2.5 Longest Cycle Length

For a threshold graph $G$, let $\psi(G)$ denote the length of the longest cycle in $G$. Reilly and Scheinerman found the probability that $\psi(G_n) = n$, which corresponds to the event in which $G_n$ is Hamiltonian, through the following result:

**Theorem 2.5.1** (Reilly, Scheinerman). For a threshold graph $G_n$ where $n \geq 3$, $G_n$ is Hamiltonian if and only if $t^u_k(C_n) > t^z_k(C_n)$ for all $1 \leq k \leq n - 1$.

(This requirement can be equivalently restated as

$$\max_{1 \leq i \leq n-1} \{t^u_i(C_n) - t^z_i(C_n)\} < 0,$$

which has certain parallels to the previous section.) Here, we generalize to find the full distribution of $\psi(G_n)$.

We define, for threshold graphs with at least two one-vertices, a particular way of constructing a cycle. This greedy algorithm constructs a path in a single pass through the creation sequence, joins the two ends, and creates a cycle which will later be shown to have maximum length.

Speaking informally, the path will extend through all of the one-vertices, ordered from right to left, beginning with $u_{U_n}$ and ending with $u_1$. (As the path is built, the endpoints will always be one-vertices.) This path attempts to include as many zero-vertices as possible, depending upon available space.

As an example, consider Figure 2.2, which contains the creation sequence $C$ for a ten-vertex threshold graph $G$, each digit labelled with the name of the corresponding vertex. We will walk through the construction of one such preliminary path, and its subsequent extension into a maximal cycle.

As the example has four one-vertices, the path begins with $u_4$. From the position of its corresponding digit at index 8, we move leftwards in the creation sequence. We
are unable to insert \( z_4 \), as it has but one neighbor in the path, and no subsequent vertex would permit extension. On the other hand, \( u_3 \), like all one-vertices, is able to be inserted, so we have a path containing two vertices. Unlike \( z_4 \), \( z_3 \) does have two neighbors in the path, so we insert it between \( u_4 \) and \( u_3 \). While \( z_2 \) also claims \( u_4 \) and \( u_3 \) as neighbors, those two are no longer adjacent in the path, so there is no way to insert \( z_2 \). As one-vertices, both \( u_2 \) and \( u_1 \) are simply appended. The last digit corresponds to \( z_1 \), which is adjacent to \( u_1 \), \( u_2 \), and \( u_3 \); as there are two possible insertion positions, we choose the one closer to the beginning.

More formally, we describe the process as follows: We create a path in \( G_n \) by moving through the creation sequence from right to left, and for each digit, we attempt to add it to the path in the following manner:

- If the digit is 1, then add the corresponding one-vertex \( u_j \) to the cycle by adjoining it to \( u_{j+1} \); \( u_j \) now becomes to new endpoint of the path. In the event that no \( u_{j+1} \) exists, indicating that the path is currently empty, then \( u_j \) will (temporarily) be the sole vertex in the path.

- If the digit is 0, then attempt to add the corresponding zero-vertex \( z_i \) to the cycle by inserting it between some pair of consecutive one-vertices \( u_ju_{j+1} \). If
multiple such pairs exist, then choose the largest such \( j \). If no such pairs exist, then we omit \( z_i \) from the path.

At the end of construction, we have a path from \( u_{U_n} \) to \( u_1 \) that goes through every single one-vertex and possibly some zero-vertices. We then convert this path into a cycle by adjoining both \( u_1 \) and \( u_{U_n} \) to base vertex \( v_0 \), as \( v_0 \) is adjacent to all one-vertices. (Recall that we required \( U_n \geq 2 \), ensuring that the two endpoints are distinct.) Let us denote this cycle by \( Y = Y(G_n) \):

\[
Y = u_{U_n} \rightarrow u_{U_n-1} \rightarrow \cdots \rightarrow u_1 \rightarrow v_0 \rightarrow u_{U_n}
\]

To continue with our example, the extension of our path into a maximum cycle would be

**Proposition 2.5.2.** For a threshold graph \( G_n \) with \( U_n \geq 2 \), \( Y(G_n) \) is a cycle of maximum length.
Proof. Assume, for purposes of contradiction, that there exists cycle $X$ in $G_n$ such that $|X| > |Y|$. Let us define $m$ by

$$m = \max\{ j : |X \cap T_j(C_n)| > |Y \cap T_j(C_n)| \}$$

(That is, working from right to left, $m$ is the first position in which cycle $X$ has more vertices than cycle $Y$ amongst those vertices of index greater than or equal to $m$.) Since $T_1(C_n) = C_n$ and $X \cap T_1(C_n) = X$, $m$ is well-defined and at least 1. As an upper bound, we know that $m \leq n - 2$: if $c_{n-1} = 1$, then $v_{n-1} \in Y$, and if $c_{n-1} = 0$, then $v_{n-1}$ is an isolated vertex, unable to be part of any cycle.

Now, by maximality of $m$ and since $|T_{i+1}(C_n)| - |T_i(C_n)| = 1$ for all $1 \leq i \leq n - 2$,

$$|X \cap T_{m+1}(C_n)| = |Y \cap T_{m+1}(C_n)| = |Y \cap T_m(C_n)|$$

and furthermore,

$$|X \cap T_m(C_n)| = |Y \cap T_m(C_n)| + 1$$

Since cycle $Y$ does not include vertex $v_m$, it must be a zero-vertex; all one-vertices are automatically included in $Y$ by construction. So the only way that $v_m$ could be included in cycle $X$ is to adjoin it to two one-vertices of higher index. Therefore $t_m^u(C_n) = t_{m+1}^u(C_n) \geq 2$. 

In order for $Y$ to exclude $v_m$, it must have already included the maximum number of zero-vertices possible, which is one less than the number of previously-included
one-vertices. As each of the $t^u_m(C_n)$ one-vertices has already been added, this means that $Y$ has, prior to index $m$, added exactly $t^u_m(C_n) - 1$ zero-vertices. Thus $Y$ contains exactly $2t^u_m(C_n) - 1$ of the vertices in tail $T_m(C_n)$. By definition of $m$, cycle $X$ must contain at least $2t^u_m(C_n)$ vertices in that same tail.

Of the vertices in $X \cap T_m(C_n)$, at least $t^u_m(C_n)$ must be zero-vertices. As any zero-vertex is adjacent only to one-vertices of higher index, all of the zero-vertices in $X \cap T_m(C_n)$ must be adjacent (in $X$) to one-vertices in $X \cap T_m(C_n)$. But as there are at least as many of the former as the latter, the only possible arrangement is to have a closed cycle of length $2t^u_m(C_n)$, alternating between one- and zero-vertices.

Since $X$ closes with the inclusion of $v_m$, we have $X = X \cap T_m(C_n)$, and thus

$$|X| = |X \cap T_m(C_n)| = |Y \cap T_m(C_n)| + 1$$

But by construction, cycle $Y$ contains the base vertex $v_0$ which is not in any tail $T_i(C_n)$, and therefore

$$|Y| \geq |Y \cap T_m(C_n)| + 1 = |X \cap T_m(C_n)| = |X|,$$

which contradicts the assumption that $|X| > |Y|$. Thus, cycle $Y$ is of maximum length. \[\square\]

Let us define a function $r$ on the set of all binary sequences as, for any $n$-long binary sequence $S = s_1s_2\ldots s_n$,

$$r(S) = \max (\{0\} \cup \{i : s_i = 1\})$$

That is, $r(S)$ returns the index of the right-most one in $S$ in the event that such exists, and zero otherwise. (So in the event that $U_n \geq 2$, $r(C_n) \geq 2$.) As such, every digit of the creation sequence with index greater than $r(C_n)$ must be zero.

Furthermore, we use $r(S)$ to define a subsequence $R(S)$ as follows:

$$R(S) = s_1s_2\ldots s_{r(S)-1}$$
Thus, whenever $r(S)$ is at least 2, $R(S)$ is a non-empty subsequence. For $r(S) \leq 1$, on the other hand, $R(S)$ is the empty sequence.

**Proposition 2.5.3.** For a threshold graph $G_n$ with $U_n \geq 2$,

$$\psi(G_n) = r(C_n) + 1 - h(R(C_n))$$

**Proof.** Essentially, we show that the given formula expresses the length of the cycle $Y$, which was already shown to be optimal. As with finding $\nu(G_n)$, it is helpful to think in terms of excluded vertices. We can restate the above claim as

$$n - \psi(G_n) = (n - 1 - r(C_n)) + h(R(C_n)),$$

which says that the number of vertices not present in any cycle of maximum length is equal to the number of consecutive zeroes at the end of the creation sequence plus the maximum number of excess zeroes amongst all tails of $R(C_n)$.

We now look at how many zero-vertices are omitted by the construction of the greedy cycle $Y$. First, any zero-vertices of higher index than $r(C_n)$ are discarded, as there exist no pairs of one-vertices between which to insert them. Thus, $(n - 1 - r(C_n))$ vertices are automatically excluded from $Y$.

Second, let $m$ denote a maximizing index for $h$ such that

$$h(R(C_n)) = t_m^u(R(C_n)) + t_m^u(R(C_n))$$

Including the one-vertex at index $r(C_n)$, there are $t_m^u(R(C_n)) + 1$ one vertices in the tail. Since any zero-vertex is adjacent only to one-vertices of higher index, we can insert at most $t_m^u(R(C_n))$ of the zero-vertices in the tail; the other $h(c_1 \cdots c_{r(C_n)-1})$ will not be present in $Y$. Thus, combining these two sources of excluded vertices,

$$n - \psi(G_n) \geq (n - 1 - r(C_n)) + h(R(C_n)),$$
or equivalently
\[
\psi(G_n) \leq r(C_n) + 1 - h(R(C_n))
\]

Third, let us construct a new creation sequence \( C' \) by removing the \((n-1-r(C_n))\) zeroes of index higher than \( r(C_n) \) from \( C_n \), and also the right-most \( h(R(C_n)) \) zeroes of index less than \( r(C_n) \). That is, we start with the sequence \( R(C_n) \), and then remove the right-most \( h(R(C_n)) \) zeroes from that sequence. Then by Lemma 2.4.1, \( h(C') = 0 \), so every tail of \( C' \) has at least as many ones as zeroes. If we now append the one-vertex at index \( r(C_n) \), the sequence \( C'1 \) has strictly more ones than zeroes in every tail, and therefore has a Hamiltonian cycle by the Reilly and Scheinerman result. Thus,

\[
\psi(G_n) \geq n - (n-1-r(C_n)) - h(R(C_n)) = r(C_n) + 1 - h(R(C_n))
\]

Theorem 2.5.4. For a random threshold graph \( G_n \),

\[
P(\psi(G_n) = k) = \left( \frac{1}{2} \right)^{n-1} \left[ \binom{n-1}{\left\lfloor \frac{k}{2} \right\rfloor} - \binom{k-2}{\left\lfloor \frac{k}{2} \right\rfloor} \right]
\]

for all \( k \geq 3 \).

Proof. By Proposition 2.5.3, in order for \( \psi(G_n) \) to equal \( k \),

\[
r(C_n) - k + 1 = h(c_1 \cdots c_{r(C_n)-1}),
\]

so for \( 3 \leq k \leq n \),

\[
P(\psi(G_n) = k) = \sum_{j=1}^{n-1} P(r(C_n) = j, \psi(G_n) = k)
= \sum_{j=1}^{n-1} P(r(C_n) = j, r(C_n) - k + 1 = h(c_1 \cdots c_{r(C_n)-1}))
= \sum_{j=k-1}^{n-1} P(r(C_n) = j, j - k + 1 = h(c_1 \cdots c_{j-1}))
\]
Note that the intersecting events are independent: the first only uses information about the final digits of the creation sequence, the second only concerns the initial digits, and there is no overlap. Since each digit is determined independently, we can then decompose the probability into the product of simpler events:

\[ P(\psi(G_n) = k) = \sum_{j=k-1}^{n-1} P(r(C_n) = j)P(h(C_j) = j - k + 1) \]

\[ = \sum_{j=k-1}^{n-1} \left( \frac{1}{2} \right)^{n-j} \left( \frac{1}{2} \right)^{j-1} \left( j - 1 \right) \left( 2j - k - \left\lfloor \frac{2j-k}{2} \right\rfloor \right) \]

\[ = \left( \frac{1}{2} \right)^{n-1} \sum_{j=k-1}^{n-1} \left( j - 1 \right) \left( j - k - \left\lfloor \frac{k}{2} \right\rfloor \right) \]

\[ = \left( \frac{1}{2} \right)^{n-1} \left[ \left( n - 1 \right) - \left( \frac{n}{2} - 1 \right) \right] \]

\[ \square \]

Note that this yields the previously published result by Reilly and Scheinerman:

**Corollary 2.5.5.** For a random threshold graph \( G_n \),

\[ P(G_n \text{ is Hamiltonian}) = \left( \frac{1}{2} \right)^{n-1} \left( \frac{n-2}{\left\lfloor \frac{n}{2} \right\rfloor} - 1 \right) \]

**Proof.** Using the distribution for \( \psi(G_n) \), above, we have

\[ P(\psi(G_n) = n) = \left( \frac{1}{2} \right)^{n-1} \left[ \left( n - 1 \right) - \left( \frac{n}{2} - 1 \right) \right] \]

For even values of \( n = 2m \), this becomes

\[ P(\psi(G_n) = n) = \left( \frac{1}{2} \right)^{n-1} \left[ \left( 2m - 1 \right) - \left( 2m - 2 \right) \right] \]

\[ = \left( \frac{1}{2} \right)^{n-1} \left( 2m - 2 \right) \left( m - 1 \right) \]

\[ = \left( \frac{1}{2} \right)^{n-1} \left( \frac{n}{2} - 1 \right) \]

\[ = \left( \frac{1}{2} \right)^{n-1} \left( \frac{n}{2} - 1 \right) \]
Whereas for odd values of $n = 2m + 1$,

\[
P(\psi(G_n) = n) = \left(\frac{1}{2}\right)^{n-1} \left[ \binom{2m}{m} - \binom{2m - 1}{m} \right]
\]

\[
= \left(\frac{1}{2}\right)^{n-1} \left(2m - 1\right)
\]

\[
= \left(\frac{1}{2}\right)^{n-1} \left(n - 2\right)
\]

\[
= \left(\frac{1}{2}\right)^{n-1} \left[\frac{n}{2} - 1\right]
\]

\[\blacksquare\]

### 2.6 Planarity

A graph is called \textit{planar} if it can be drawn in the plane without any of the edges intersecting. It is simple to show that a graph is planar, as all one has to do is provide the appropriate drawing. However, showing that a graph is not planar requires a little more effort.

Given a graph $G$ with edge $e = \{x, y\} \in E(G)$, a \textit{subdivision} of $e$ is formed by replacing $e$ with the pair of edges $\{x, v\}$ and $\{v, y\}$, where $v$ is a new vertex that is not present in $V(G)$. A \textit{subdivision} of $G$ is any graph formed by iteratively applying the subdivision process to edges of $G$.

\textbf{Theorem 2.6.1} (Kuratowski [11]). \textit{A finite graph is planar if and only if it does not contain a subgraph that is a subdivision of $K_5$ or $K_{3,3}$}.

We can now use Kuratowski’s theorem, along with the following lemma, to determine whether or not a threshold graph is planar.

\textbf{Lemma 2.6.2.} \textit{Let $H$ be a subdivision of $G$. Then for all $v \in V(G) \cap V(H)$, $\deg_G(v) = \deg_H(v)$}.

\textit{Proof.} As $H$ is a subdivision of $G$, it can be obtained via a series of subdivisions of edges. Thus, it suffices to show that subdividing an edge of $G$ does not change the degree of any vertices in $G$.
When edge \( \{x, y\} \) is subdivided, no other edge is altered. So all vertices in \( G \) other than \( x \) and \( y \) are untouched, and their degrees remain constant. On the other hand, the degrees of vertices \( x \) and \( y \) are each reduced by one with the removal of \( \{x, y\} \), but are also each increased by one with the addition of \( \{x, v\} \) and \( \{v, y\} \). Thus, even though the total degree is increased by a positive net change of one edge, the degrees of each of the original vertices remain the same in \( H \) as they were in \( G \).

Thus, when we start hunting for subdivisions of \( K_5 \), we can restrict our attention to subgraphs that have at least five vertices of degree four or more. Similarly, any subgraph with fewer than six vertices of degree three or more cannot contain a subdivision of \( K_{3,3} \).

**Theorem 2.6.3.** For a random threshold graph \( G_n \),

\[
P(G_n \text{ is planar}) = \begin{cases} 
1, & \text{for } n \leq 4 \\
\frac{3n^2 - 13n + 20}{2^n}, & \text{for } n \geq 4
\end{cases}
\]

*Proof.* Note that any subdivision of either \( K_5 \) or \( K_{3,3} \) must have at least five vertices. Thus, any graph on four or fewer vertices is automatically planar, as it cannot contain a subgraph larger than itself. We now consider several cases, based on the number of one-vertices in \( G_n \):

Reilly and Scheinerman showed that for a threshold graph \( G_n \), the clique number \( \omega(G_n) \) is equal to \( U_n + 1 \). That is, there will always be a set of \( U_n + 1 \) vertices such that every vertex in the set is adjacent to every other vertex therein. So if \( U_n \geq 4 \), \( G_n \) has a clique of size at least 5, and thus a subgraph isomorphic to \( K_5 \). So in this case, \( G_n \) is non-planar.

On the other hand, if \( U_n \leq 2 \), then the base vertex and the zero-vertices are adjacent to at most two vertices. So the only vertices which can have degree of three or more are the one-vertices. Thus, since the number of vertices with degree at least
three is at most two, no subgraph can be a subdivision of either $K_5$ or $K_{3,3}$. So in this case, $G_n$ is planar.

Now consider the case when $U_n = 3$. Since the base vertex and zero-vertices are adjacent only to one-vertices, none of them have degree exceeding three. Then only the one-vertices can have degree of more than three, so there there are at most three vertices with degree of four or more. So we can again discard the possibility of having a subgraph that is a subdivision of $K_5$, and restrict ourselves to detecting the presence of a subdivided $K_{3,3}$.

Here there are two subcases, determined by the initial digits of the creation sequence. If $U_n = 3$ and $c_1 = c_2 = 0$, then the three one-vertices are each of higher index than $v_0, z_1,$ and $z_2$. So each of those is adjacent to each of the three one-vertices, and $G_n$ contains a subgraph isomorphic to $K_{3,3}$, with no subdivision necessary. Alternatively, if either $c_1$ or $c_2$ is equal to 1, then there exists a one-vertex which is adjacent to at most two non-one-vertices, and thus there are at most two non-one-vertices with degree of three or more. Thus, there are at most five vertices of degree three or above, and no subgraph can be isomorphic to $K_{3,3}$.

Thus, in order for $G_n$ to be planar, we need one of the following mutually-exclusive events:

\[
\{G_n \text{ is planar}\} = \{U_n = 0\} \cup \{U_n = 1\} \cup \{U_n = 2\}
\]
\[
\cup (\{U_n = 3\} \cap \{c_1 = c_2 = 0\})^c
\]
\[
= \{U_n = 0\} \cup \{U_n = 1\} \cup \{U_n = 2\}
\]
\[
\cup \{U_n = 3, c_1 = 1\} \cup \{U_n = 3, c_1 = 0, c_2 = 1\}
\]

Since our probability space is uniform on the set of all binary sequences of length
$n - 1$, the probability of each event is proportional to the number of sequences therein. Thus,

$$P(G_n \text{ is planar}) = P(U_n = 0) + P(U_n = 1) + P(U_n = 2)$$

$$+ P(U_n = 3, c_1 = 1) + P(U_n = 3, c_1 = 0, c_2 = 1)$$

$$= \binom{n-1}{0} \cdot \frac{1}{2^{n-1}} + \binom{n-1}{1} \cdot \frac{1}{2^{n-1}} + \binom{n-1}{2} \cdot \frac{1}{2^{n-1}} + \binom{n-2}{2} \cdot \frac{1}{2^{n-1}} + \binom{n-3}{2} \cdot \frac{1}{2^{n-1}}$$

Simplifying the binomials, we see that

$$P(G_n \text{ is planar}) = \frac{1}{2^{n-1}} \left( \frac{3n^2 - 13n + 20}{2} \right) = \frac{3n^2 - 13n + 20}{2^n}$$

2.7 k-Core

A $k$-core of a graph $G$ is the maximum induced subgraph $H \subseteq G$ such that all vertices of $H$ have degree at least $k$, formed by iteratively deleting all vertices with degree less than $k$. The degeneracy of a graph $G$ is the largest $k$ such that the $k$-core of $G$ is non-empty.

An equivalent formulation for degeneracy of $G$ is the maximum, over all induced subgraphs $H \subseteq G$, of the minimum degree of a vertex in $H$. That is,

$$\text{degen}(G) = \max_{H \subseteq G} \min_{v \in V(H)} \deg(v)$$

Regardless of definition, degeneracy is a measure of the density of a graph; for a given number of vertices, larger degeneracies correspond to higher number of edges. Similarly, the resulting $k$-cores are useful for discussing the clustering within these random graphs.

**Proposition 2.7.1.** For a threshold graph $G_n$, $\text{degen}(G_n) \geq d$ if and only if $K_{d+1} \subseteq G_n$. 

29
Proof. First, assume that \(G_n\) contains a subgraph isomorphic to \(K_{d+1}\). Then every vertex in the subgraph not only has degree \(d\), but is adjacent to \(d\) vertices in the subgraph that are also of degree \(d\). As no vertex in the subgraph can be pruned away until one of the other vertices in the subgraph is removed first, we see that the subgraph must be part of the \(k\)-core for any \(k \leq d\). So the degeneracy must be at least \(d\).

Second, assume that \(G_n\) has degeneracy \(d\). Then \(G_n\) has a non-empty \(d\)-core, and therefore there exists a set \(V' \subseteq V(G_n)\) in which every vertex of \(V'\) is adjacent to at least \(d\) other members of \(V'\). Because of this degree requirement, we see that \(|V'| \geq d + 1\). Furthermore, \(V'\) must contain at least \(d\) one-vertices, as otherwise there could be at most \(d - 1\) vertices of \(V'\) with degree of \(d\) or greater. Therefore \(U_n \geq d\), so the clique number \(\omega(G_n) \geq d + 1\), and there exists a subgraph of \(G_n\) isomorphic to \(K_{d+1}\).

Corollary 2.7.2. For a threshold graph \(G_n\), \(\text{degen}(G_n) = d\) if and only if \(U_n = d\).

Corollary 2.7.3. For a random threshold graph \(G_n\),

\[
P(\text{degen}(G_n) = d) = P(U_n = d) = \frac{1}{2^{n-1}} \binom{n-1}{d}
\]

Proof. Since each of the \((n-1)\) digits of the creation sequence is independently chosen uniformly at random from \(\{0, 1\}\), the number of ones is binomially distributed with probability \(p = \frac{1}{2}\).

Proposition 2.7.4. For a threshold graph \(G_n\) with \(U_n \geq k\), a vertex \(v \in V(G_n)\) is in the \(k\)-core of \(G_n\) if and only if \(\deg_{G_n}(v) \geq k\).

Proof. First, if a vertex is in the \(k\)-core of \(G_n\), it must, by definition, have a degree of at least \(k\) in that subgraph, and thus a degree of at least \(k\) in \(G_n\). (Note that the restriction \(U_n \geq k\) serves to remove from consideration those cases in which \(G_n\) does not have a \(k\)-core.)
Second, consider some vertex $v$ such that $\operatorname{deg}_{G_n}(v) \geq k$; there are two possible cases, depending upon the classification of $v$. If $v$ is a one-vertex, then it is adjacent to all other one-vertices (of which there are at least $k - 1$) and to the base vertex. So $v$ is an vertex of some $K_k$, and therefore part of the $k$-core. On the other hand, if $v$ is a base or zero-vertex, then $v$ is adjacent to $k$ one-vertices in the $k$-core, and therefore part of the $k$-core itself.

Another way to phrase the above proposition is that if $G_n$ has a non-empty $k$-core, no vertex of degree at least $k$ will have its degree reduced below $k$ by the pruning process.

**Theorem 2.7.5.** For a random threshold graph $G_n$,

$$P(|k\text{-core}(G_n)| = j) = \begin{cases} \sum_{i=0}^{k-1} \left( \frac{1}{2} \right)^{n-1} \binom{n-1}{i} & j = 0 \\ \left( \frac{1}{2} \right)^{n+k-j} \binom{n + k - j - 1}{k - 1} & j \geq k + 1 \end{cases}$$

*Proof.* First, in order for the $k$-core of $G_n$ to be empty, the degeneracy of $G_n$ can be at most $k - 1$. Thus, by Corollary 2.7.3, the total probability of having $|k\text{-core}(G_n)| = 0$ is

$$\sum_{i=0}^{k-1} P(\text{degen}(G_n) = i) = \sum_{i=0}^{k-1} \left( \frac{1}{2} \right)^{n-1} \binom{n-1}{i}$$

Note that the event of the $k$-core having exactly $j$ vertices is the same as the event of having exactly $n - j$ vertices removed by the pruning process. Since $j \geq 1$, this means the pruning process will not remove all of the vertices, necessitating that a non-empty $k$-core exist; so this event is a sub-event of $U_n \geq k$. Thus the removed vertices are all zero-vertices of degree less than $k$, so they must have at most $k - 1$ one-vertices of higher index. Then $C_n$ has $n - j$ zeroes lying to the right of the $k$-th one from the right, denoted $u_{U_n-k+1}$. 31
To summarize, the $k$-core of $G_n$ has exactly $j$ vertices if and only if the right-most $n + k - j - 1$ digits of $C_n$ are exactly $k - 1$ ones and $n - j$ zeroes, and the $n + k - j$-th digit from the right is a one.

The initial $j - k - 1$ digits can be either zeroes or ones, and the final $n + k - j - 1$ digits can be in any order, so there are $2^{j-k-1} \binom{n + k - j - 1}{k - 1}$ such sequences. Therefore,

$$P(|k\text{-core}(G_n)| = j) = \frac{2^{j-k-1}}{2^{n-1}} \binom{n + k - j - 1}{k - 1}$$

\[ \square \]

### 2.8 $t$-Stability

A vertex subset $V' \subset V(G)$ is called $t$-stable if the subgraph induced by $V'$ has a maximum degree of at most $t$. The $t$-stability number $\alpha_t(G)$ is the size of the largest $t$-stable set in $G$. This is a generalization of the independence number $\alpha(G)$, which is the size of the largest set of vertices such that no two vertices in the set are adjacent; note that $\alpha(G) = \alpha_0(G)$, as any independent set induces a subgraph of maximum degree zero.

As the maximum $t$-stable set need not be unique, we begin by showing that for threshold graphs, there must always exist a $t$-stable set in an easily-analyzed form:

**Proposition 2.8.1.** For a threshold graph $G_n$, there exists a $t$-stable set of maximum size that contains the base vertex and all zero-vertices.

**Proof.** In the interest of brevity, we shall temporarily classify the base vertex as a zero-vertex; any claims made herein about the latter will also apply to the former.

Take any $t$-stable set $S$ of maximum size, and let $k$ denote the number of zero-vertices of $G_n$ that are not in $S$. Suppose that $k > 0$; let $z$ denote any such excluded zero-vertex.
As $S$ is of maximum size, we cannot add $z$ to $S$ without raising the degree of some element of $S$ above $t$. Since $z$ is a zero-vertex, and thus only adjacent to one-vertices, this means that $S$ must contain at least one one-vertex $u$. As $u$ is a one-vertex, it is adjacent to all other elements of $S$, so if we define $S' = S \setminus \{u\}$, all of the vertices must have degree (in $S'$) of at most $t - 1$. Then adding $z$ to $S'$, which raises the degree of any vertex therein by at most 1, will still result in a $t$-stable set.

Then we have a $t$-stable set $S' \cup \{z\}$ that is the same size as $S$, but excludes only $k - 1$ zero-vertices of $G_n$. Repeating this process another $k - 1$ times produces a $t$-stable set that includes the base vertex and all zero-vertices.

**Proposition 2.8.2.** For a threshold graph $G_n$, a $t$-stable set of maximum size is given by

$$S = \{v_0, v_1, \ldots, v_t\} \cup Z$$

*Proof.* The set $S$ is clearly maximal; the only vertices of $G_n$ that are not included are one-vertices, and since $S$ already contains $\{v_0, \ldots, v_t\}$, any of the excluded one-vertices would be adjacent to at least $t + 1$ elements of $S$.

Suppose, for the sake of contradiction, that there exists some $t$-stable set $S'$ such that $|S| < |S'|$. By Proposition 2.8.1, we can assume that $S'$ contains both $Z$ and $v_0$. So $|S' \cap U| > |S \cap U|$, and the set difference $S' \setminus S$ is a non-empty subset of $U$. Let $u$ denote the element of $S' \setminus S$ with greatest index; by construction of $S$, said index must be strictly greater than $t$. Then $u$ is adjacent to all other one-vertices in $S'$, as well as all zero-vertices of lower index.

Let $T = \{v_0, \ldots, v_t\} \setminus U$, the number of base and zero-vertices amongst the $t + 1$ vertices of least index. Then we see that $|T| + |S \cap U| = t + 1$, and therefore $|T| + |S' \cap U| \geq t + 2$. Since $u$ is adjacent to all of $T$ and all but one of the vertices in $S' \cap U$, the $\text{deg}_{S'}(u) \geq t + 1$, which contradicts the assumption that $S'$ is $t$-stable. □
Corollary 2.8.3. For a random threshold graph $G_n$,

$$P(\alpha_t(G_n) = k) = \left(\frac{1}{2}\right)^{n-t-1} \binom{n-t-1}{k-t-1}$$

**Proof.** By Proposition 2.8.2, the event $\alpha_t(G_n) = k$ is equivalent to the event that the union of the set of zero-vertices and the set of $t+1$ vertices of least index has size $k$. That is, that there are exactly $k - t - 1$ zeroes amongst the right-most $n - t - 1$ digits of the creation sequence. \qed
CHAPTER 3
DIFFERENCE GRAPH MODELS

3.1 Background

The mechanics of threshold graphs can be slightly changed to produce a very similar class of graphs. We retain the vertex set \( V \) and weighting function \( w \), but extend the codomain from \([0, 1]\) to \([0, 1] \times \{0, 1\}\). Instead of adding edge \( \{v_i, v_j\} \) if and only if \( w(v_i) + w(v_j) > 1\), we require that \( w_1(v_i) + w_1(v_j) > 1 \) and \( w_2(v_i) \neq w_2(v_j) \). Those graphs which result from this process are known as difference graphs.\(^1\)

As a consequence of the second requirement for edge creation, all difference graphs are bipartite; thus, they are also known as bipartite threshold graphs. So for a difference graph \( D \), we label the parts as follows:

\[
Z = \{v \in V(D) : w_2(v) = 0\} \quad \text{and} \quad U = \{v \in V(D) : w_2(v) = 1\}
\]

Since the only role of \( w_2(v) \) is to determine in which part vertex \( v \) lies, whenever we know the part containing \( v \), we shall simply use \( w(v) \) instead of \( w_1(v) \).

Another parallel between threshold graphs and difference graphs is their shared

\(^1\)The derivation of the name comes from the following property: define \( f : V \rightarrow [0, 1] \) by

\[
f(v) = w_1(v)(2w_2(v) - 1)
\]

Then \( v_i \) and \( v_j \) are adjacent if and only if \(|f(v_i) - f(v_j)| > 1\).
ability to encode their structure into a binary sequence. Before we show how to do so, we require a preliminary result:

**Proposition 3.1.1.** Let $D$ be a difference graph on at least one vertex, with vertex parts $Z$ and $U$. Then exactly one of the following must be true: either $U$ has a vertex that dominates all of $Z$, or $Z$ has an isolated vertex.

**Proof.** First, we assume that $Z$ is non-empty, as otherwise $U$ has a vertex which (trivially) dominates all of $Z$. Similarly, if $U$ is empty, then all vertices of $Z$ are isolated, as $D$ is bipartite. Having handled these trivial cases, we now suppose that $|Z|, |U| \geq 1$.

Let $z_{\min}$ be a vertex of $Z$ such that for all $z \in Z$, $w(z_{\min}) \leq w(z)$. Similarly, let $u_{\max}$ be a vertex of $U$ such that for all $u \in U$, $w(u) \leq w(u_{\max})$. We now consider two disjoint cases:

If $w(z_{\min}) + w(u_{\max}) > 1$, then there exists an edge between $z_{\min}$ and $u_{\max}$. Moreover, since $z_{\min}$ has the least weight of all vertices of $Z$, for any $z \in Z$, we have

$$ w(u_{\max}) + w(z) \geq w(u_{\max}) + w(z_{\min}) > 1 $$

Therefore, $u_{\max}$ dominates every vertex of $Z$.

On the other hand, if $w(z_{\min}) + w(u_{\max}) \leq 1$, then there does not exist an edge between $z_{\min}$ and $u_{\max}$. And as $u_{\max}$ has the greatest weight of all vertices of $U$, for any $u \in U$, we have

$$ w(u) + w(z_{\min}) \leq w(u_{\max}) + w(z_{\min}) \leq 1 $$

Therefore, $z_{\min}$ is an isolated vertex of $Z$.

So $D$ has either a dominating $U$-vertex or an isolated $Z$-vertex, but not both.  

With this result, and the fact that any induced subgraph of a difference graph is another difference graph, we can extend the notion of a creation sequence to difference
graphs. That is, we encode the entire structure of an $n$-vertex difference graph $D_n$ as an $n$-digit binary sequence, denoted $\text{seq}(D_n)$. We do so recursively, constructing the sequence from right to left as we process nested (induced) subgraphs of $D_n$:

Let $D$ be a difference graph, with vertex parts $Z$ and $U$ as defined above. If $|Z| = |U| = 0$, then $\text{seq}(D) = \emptyset$, the trivial sequence of length 0. Otherwise, there must exist some vertex $v$ such that $v$ is either an isolated vertex of $Z$ or a vertex of $U$ that dominates all of $Z$. Then, letting $||$ denote the operation of concatenation, we let

$$\text{seq}(D) = \text{seq}(D \setminus \{v\}) || d,$$

where $d = \begin{cases} 0 & \text{if } v \in Z \\ 1 & \text{if } v \in U \end{cases}$

So the creation sequence $\text{seq}(D)$ has $|Z|$ zeroes and $|U|$ ones, and two vertices $z_i$ and $u_j$ are adjacent in $D$ if and only if the digit corresponding to $z_i$ lies to the left of $u_j$'s in $\text{seq}(D)$.

**Proposition 3.1.2.** Let $S$ be a binary sequence. Then there exists a difference graph $D$ such that $\text{seq}(D) = S$.

**Proof.** We proceed by induction, and begin by considering three base cases. If $S$ is the empty sequence of length zero, then we let $D$ be the empty graph on zero vertices. If $S = 0$, then take $D$ to be the difference graph where $U$ is empty and $Z$ consists of a single vertex of weight $1/2$. Similarly, if $S = 1$, then let $D$ be the difference graph where $Z$ is empty and $U$ is a single vertex of weight $1/2$.

Now, suppose that $S$ has length $k$, where $k \geq 2$. Let $S'$ be the sequence consisting of the first $k - 1$ digits of $S$, and let $s$ be the final digit, so that $S = S' || s$. By induction, there exists a difference graph $D'$, with parts $Z'$ and $U'$, such that $\text{seq}(D') = S'$.

Using $D'$, we shall define a difference graph $D$ that corresponds to $S$ in one of two ways, depending upon $s$. In either case, we begin by adding a new vertex $v$ to $D'$. 37
If \( s = 0 \), then \( D \) must have an isolated vertex in \( Z \). We define the weight of \( v \) by
\[
w(v) = \left( \frac{1}{2} \right) \min \{1 - w(u) : u \in U'\}
\]
This renders the weight of \( v \) too low to form an edge with any vertex of \( U' \), as for all \( u \in U' \), \( w(v) + w(u) < 1 \). Then we define \( D \) by \( Z = Z' \cup \{v\} \) and \( U = U' \).

If \( s = 1 \), then \( D \) must have a vertex in \( U \) that dominates all of \( Z \). We define the weight of \( v \) by
\[
w(v) = \left( \frac{1}{2} \right) (1 + \max \{1 - w(z) : z \in Z'\})
\]
This causes the weight of \( v \) to be high enough to form edges with every vertex of \( Z' \), as for all \( z \in Z' \), \( w(z) + w(v) > 1 \). Then we define \( D \) by \( Z = Z' \) and \( U = U' \cup \{v\} \).

Thus, \( \text{seq}(D) = S \).

(Please note that the above process will, for most creation sequences, produce difference graphs with duplicate weights in each part. In subsequent sections, there are times when we need \( w(u_i) \neq w(u_j) \) for all \( u_i, u_j \in U \), and similar restrictions for vertices in \( Z \). In such cases, we can simply amend the above definitions of \( w(v) \) to
\[
w(v) = \left( \frac{1}{2} \right) \min (\{w(z) : z \in Z'\} \cup \{1 - w(u) : u \in U'\})
\]
or
\[
w(v) = \left( \frac{1}{2} \right) (1 + \min (\{w(u) : u \in U'\} \cup \{1 - w(z) : z \in Z'\}))
\]
depending upon whether \( s = 0 \) or \( s = 1 \), respectively.)

Thus, we have a bijection between the set of \( k \)-long binary sequences and the set of difference graphs on \( k \) vertices.

### 3.2 Comparison of Models

There are two models for the generation of difference graphs with fixed part sizes \( n_0 \) and \( n_1 \).
First, the *continuous model*, or weighting model, starts with an empty bipartite graph, whose parts we shall denote as $Z = \{z_1, \ldots, z_{n_0}\}$ and $U = \{u_1, \ldots, u_{n_1}\}$, and a random weighting vector $\vec{w}$, where

$$\vec{w} = \langle X_1, \ldots, X_{n_0}, Y_1, \ldots, Y_{n_1} \rangle,$$

and each $X_i$ and $Y_j$ is independent and identically uniform on $[0, 1]$. We add an edge $\{z_i, u_j\}$ if and only if $X_i + Y_j > 1$. We then remove the labels from the vertices and let $D(\vec{w})$ denote the random graph resulting from this process. (Alternatively, we can view $D(\vec{w})$ as the isomorphism class of the graph on vertex set $Z \cup U$.)

Second, the *discrete model*, or creation sequence model, starts by selecting some sequence $C_{n_0, n_1}$ uniformly at random from $\mathbb{C}_{n_0, n_1}$, the set of all binary sequences with exactly $n_0$ zeroes and $n_1$ ones. We start with an empty bipartite graph, with neither vertices nor edges, but with the parts denoted $U$ and $Z$. We then begin reading $C_{n_0, n_1}$ from left to right. Every time we encounter a zero, we add a vertex to part $Z$, but no edges. Every time we encounter a one, we add a vertex to part $U$, and add edges from that vertex to every vertex in $Z$. Let $D(C_{n_0, n_1})$ denote the unique graph resulting from this process.

Our goal is to show that these two methods are equivalent. That is, for a given difference graph $D$, the probability of a randomly-chosen weighting vector generating $D$ is equal to the probability of a randomly-chosen creation sequence generating $D$.

### 3.3 Defining the Regions

Let $\vec{w}$ denote the weighting vector used in the first model:

$$\vec{w} = \langle x_1, \ldots, x_{n_0}, y_1, \ldots, y_{n_1} \rangle \in [0, 1]^{n_0+n_1}$$

We then divide the space of possible vectors with the following hyperplanes:
\[
\text{\begin{itemize}
\item $\forall i, j \in [n] \text{ with } i < j, \alpha_{i,j} = \{ \vec{v} \in \mathbb{R}^{n_0+n_1} : x_i = x_j \}$
\item $\forall i, j \in [n] \text{ with } i < j, \beta_{i,j} = \{ \vec{v} \in \mathbb{R}^{n_0+n_1} : y_i = y_j \}$
\item $\forall i \in [n_0], \forall j \in [n_1], \gamma_{i,j} = \{ \vec{v} \in \mathbb{R}^{n_0+n_1} : x_i + y_j = 1 \}$
\end{itemize}}
\]

Let $\mathbb{P}_{n_0,n_1}$ denote the space $(0,1)^{n_0+n_1}$ without the above hyperplanes, and be called the space of proper representations. Then for all $\vec{w} \in \mathbb{P}_{n_0,n_1}$, the first $n_0$ coordinates are all distinct, as are the last $n_1$. As only a set of measure zero was removed, it suffices, for probabilistic purposes, to consider $\mathbb{P}_{n_0,n_1}$ instead of $[0,1]^{n_0+n_1}$. Let $\mathcal{R}$ denote the set of connected regions of $\mathbb{P}_{n_0,n_1}$.

We now show that within any given region $R$, all of the weight vectors within give rise to the same difference graph.

**Proposition 3.3.1.** For any $R \in \mathcal{R}$ and any $\vec{w}, \vec{w}' \in R$, $D(\vec{w}) \cong D(\vec{w}')$.

**Proof.** First, we can assume that both $n_0$ and $n_1$ are positive, for were one of them to equal zero, the resulting graphs would both be empty, and thus automatically isomorphic.

Next, as both $\vec{w}$ and $\vec{w}'$ are on the same side of all hyperplanes of the form $\gamma$, we see that $x_i + y_j > 1$ if and only if $x'_i + y'_j > 1$. As such, there is an edge between the vertices corresponding to weights $x_i$ and $y_j$ exactly when there exists an edge between the vertices corresponding to $x'_i$ and $y'_j$. Thus, the two graphs lie in the same isomorphism class, under the function that maps the vertex corresponding to $x_i$ to the vertex corresponding to $x'_i$. Thus, $D(\vec{w}) = D(\vec{w}')$.

\[\square\]

Thus, for any $C_{n_0,n_1}$, the set of $\vec{w} \in \mathbb{P}_{n_0,n_1}$ such that $D(C_{n_0,n_1}) \cong D(\vec{w})$ is a disjoint union of connected regions of $\mathbb{P}_{n_0,n_1}$, and therefore a subset of $\mathcal{R}$.
3.4 Counting the Regions

To count the regions, we establish a bijection between $\mathcal{R}$ and $S_{n_0} \times S_{n_1} \times C_{n_0,n_1}$.

Let $R$ be an element of $\mathcal{R}$, and let $\vec{w}$ be an element of $R$. Let $\sigma_R$ be the permutation of $[n_0]$ such that
\[ x_{\sigma_R(1)} < x_{\sigma_R(2)} < \cdots < x_{\sigma_R(n_0)} \]
That is, $\sigma_R$ is the permutation that sorts the first $n_0$ coordinates of $\vec{w}$ into increasing order. Note that the $\alpha$-hyperplanes cause this permutation to be uniquely defined, and moreover, consistent over all of $R$. Similarly, let $\tau_R \in S_{n_1}$ be the permutation that increasingly sorts the last $n_1$ elements of $\vec{w}$, such that
\[ y_{\tau_R(1)} < y_{\tau_R(2)} < \cdots < y_{\tau_R(n_1)} \]
Finally, let $\text{seq}(R) \in C_{n_0,n_1}$ be the creation sequence such that $D(\text{seq}(R)) \cong D(\vec{w})$. By Proposition 3.3.1, above, this sequence is uniquely defined and independent of the choice of $\vec{w} \in R$.

Thus, we can define a mapping
\[ f : \mathcal{R} \to S_{n_0} \times S_{n_1} \times C_{n_0,n_1} \text{ by } f : R \mapsto (\sigma_R, \tau_R, \text{seq}(R)). \]

**Proposition 3.4.1.** The mapping $f$ is a bijection.

**Proof.** First, we show that $f$ is injective. Take $R, R' \in \mathcal{R}$ such that $R \neq R'$. Then there must exist some hyperplane $\phi$, where $\phi$ is of the form $\alpha_{i,j}$, $\beta_{i,j}$, or $\gamma_{i,j}$, such that $R$ and $R'$ are on opposite sides of $\phi$.

If $\phi = \alpha_{i,j}$, for appropriate choice of $i$ and $j$, then $R$ and $R'$ have a different ordering of the first $n_0$ coordinates of their elements. In particular, either $x_i > x_j$ for all $\vec{w} \in R$ and $x'_i < x'_j$ for all $\vec{w}' \in R'$, or the opposite is true. Thus, $\sigma_R$ and $\sigma_{R'}$ are different permutations, and $f(R)$ and $f(R')$ differ in the first coordinate.
Similarly, if $\phi$ is of the form $\beta_{i,j}$, then $f(R)$ and $f(R')$ differ in the second coordinate, as $\tau_R \neq \tau_{R'}$. Finally, if $\phi$ is of the form $\gamma_{i,j}$, then elements of $R$ and elements of $R'$ give rise to different graphs, and thus $f(R)$ and $f(R')$ differ in the third coordinate.

Second, to show that $f$ is surjective, we take an arbitrary $(\sigma, \tau, C_{n_0,n_1}) \in \mathbb{S}_{n_0} \times \mathbb{S}_{n_1} \times \mathbb{C}_{n_0,n_1}$. Since $D(C_{n_0,n_1})$ is a difference graph, there must exist some weight vector $\vec{w}$ such that $D(\vec{w}) \cong D(C_{n_0,n_1})$ by Proposition 3.1.2. (By using the secondary weighting functions, one can construct a weight vector that does not intersect any of the hyperplanes, and it therefore in $\mathbb{P}_{n_0,n_1}$.)

Now let us define $\sigma_w \in \mathbb{S}_{n_0}$ and $\tau_w \in \mathbb{S}_{n_1}$ to be the permutations that sort the first $n_0$ and last $n_1$ coordinates of $\vec{w}$, respectively. That is,

$$x_{\sigma_w(1)} < \cdots < x_{\sigma_w(n_0)} \quad \text{and} \quad y_{\tau_w(1)} < \cdots < y_{\tau_w(n_1)}$$

Using $\vec{w}$, as well as all four permutations, we define a final weight vector $\vec{v}$ by permuting the coordinates of $\vec{w}$. Specifically,

$$\vec{v} = \langle x_{\sigma^{-1}(\sigma_w(1))}, \ldots, x_{\sigma^{-1}(\sigma_w(n_0))}, y_{\tau^{-1}(\tau_w(1))}, \ldots, y_{\tau^{-1}(\tau_w(n_1))} \rangle \in \mathbb{P}_{n_0,n_1}$$

Let $R_v$ be the element of $\mathcal{R}$ that contains $\vec{v}$. Then, since the permutations $\sigma$ and $\tau$ sort the coordinates of $\vec{v}$, they must be the first and second coordinates of $f(R_v)$. And since the resulting sorted vector produces a difference graph whose creation sequence is $C_{n_0,n_1}$, we have $f(R_v) = (\sigma, \tau, C_{n_0,n_1})$. Thus, $f$ is surjective.

**Corollary 3.4.2.** There are $(n_0 + n_1)!$ connected regions of $\mathbb{P}_{n_0,n_1}$.

**Proof.** From the above,

$$|\mathcal{R}| = |\mathbb{S}_{n_0} \times \mathbb{S}_{n_1} \times \mathbb{C}_{n_0,n_1}| = n_0! n_1! \binom{n_0 + n_1}{n_0} = (n_0 + n_1)!$$

\[\square\]
3.5 Congruence of the Regions

Having found the number of connected regions of $\mathbb{P}_{n_0,n_1}$, we now show that these regions are congruent. Since the weight vectors are drawn uniformly from $[0,1]^{n_0+n_1}$, this congruence will prove that each has the same measure in the probability space.

**Proposition 3.5.1.** For any region $R \in \mathcal{R}$, any reflection across a hyperplane of type $\alpha$, $\beta$, or $\gamma$ produces another element of $\mathcal{R}$.

*Proof.* Since every such reflection $\phi$ is a continuous mapping, the image of any connected set is connected. Furthermore, since $\phi$ is an involution and

$$\phi \left( (0,1)^{n_0+n_1} \right) = (0,1)^{n_0+n_1},$$

we need only check that $\phi(R)$ does not intersect any of the excluded hyperplanes.

Take some $R \in \mathcal{R}$ and $\vec{w} \in R$, where

$$\vec{w} = (x_1, \ldots, x_{n_0}, y_1, \ldots, y_{n_1})$$

By definition of $\mathcal{R}$, we know that all of the $x_i$ are distinct, as are each of the $y_j$, and for any $i$ and $j$, $x_i + y_j \neq 1$.

As reflections of the $\alpha$ and $\beta$ types only transpose two coordinates, the resulting $\alpha(\vec{w})$ and $\beta(\vec{w})$ also lie in $\mathbb{P}_{n_0,n_1}$. Now consider the effect of $\gamma_{i,j}$:

$$\gamma_{i,j}(\vec{w}) = (x_1, \ldots, x_{i-1}, 1 - y_j, x_{i+1}, \ldots, x_{n_0}, y_1, \ldots, y_{j-1}, 1 - x_i, y_{j+1}, \ldots, y_{n_1})$$

The first $n_0$ coordinates are all distinct, as $x_k = 1 - y_j$ implies that $\vec{w} \in \gamma_{k,j}$, a contradiction. Similarly, the last $n_1$ coordinates are distinct. And the sum of any of the first $n_0$ with any of the last $n_1$ cannot equal 1, because:

- $x_k + y_l = 1$ implies $\vec{w} \in \gamma_{k,l}$
- $x_k + (1 - x_i) = 1$ implies $\vec{w} \in \alpha_{i,k}$
• $(1 - y_j) + y_l = 1$ implies $\vec{w} \in \beta_{j,l}$

• $(1 - y_j) + (1 - x_i) = 1$ implies $\vec{w} \in \gamma_{i,j}$

Thus, by connectedness of $\phi(R)$, we have $\phi(R) \in \mathcal{R}$. \hfill \Box

Having shown that these reflections act as maps of $\mathcal{R}$ to itself, we now demonstrate how to map any region to any other.

**Proposition 3.5.2.** Let $1_{n_0}$ and $1_{n_1}$ denote the identity permutations on $[n_0]$ and $[n_1]$, respectively, and $S_0$ the creation sequence of the empty difference graph on parts of size $n_0$ and $n_1$:

$$S_0 = \underbrace{1 \cdots 1}_{n_0} \underbrace{0 \cdots 0}_{n_1}$$

Then for any $R \in \mathcal{R}$, there exists a composition of reflections of the form $\alpha$, $\beta$, and $\gamma$ that maps $R$ to $f^{-1}(1_{n_0}, 1_{n_1}, S_0)$.

**Proof.** Specifically, we produce an algorithm that constructs the mapping. Take any region $R \in \mathcal{R}$, and any $\vec{w} \in R$. Then apply the following process to $\vec{w}$:

1. Apply a composition of $\alpha$-reflections to (increasingly) sort the first $n_0$ coordinates.

2. If there exists some index $j$ such that $x_{n_0} + y_j > 1$, then reflect about $\gamma_{n_0,j}$ and return to Step 1. (This replaces $x_{n_0}$ with $1 - y_j$ and $y_j$ with $1 - x_{n_0}$, decreasing both.) Otherwise, proceed to Step 3.

3. Apply a composition of $\beta$-reflections to (increasingly) sort the last $n_1$ coordinates.

The resulting vector already has all $x_i$ and $y_j$ listed in increasing order, and thus its region corresponds to the identity permutations $1_{n_0}$ and $1_{n_1}$. As for the
corresponding difference graph, note that courtesy of this sorting and Step 2, for any \( i \in [n_0], j \in [n_1], \)
\[
x_i + y_j < x_{n_0} + y_j < 1,
\]
and therefore there are no edges in the graph.

Thus, the resulting vector lies in \( f^{-1}(1_{n_0}, 1_{n_1}, S_0) \), so applying the same composition of reflections to \( R \) will produce \( f^{-1}(1_{n_0}, 1_{n_1}, S_0) \).

Thus, since every region can be mapped onto this “base region” by some composition of invertible rigid transformations, all of the regions are congruent.

### 3.6 Conclusion

Having shown that \( \mathbb{P} \), which has measure 1, consists of \( (n_0 + n_1)! \) regions, each of which is congruent, we see that each individual region has measure \(( (n_0 + n_1)! )^{-1}\).

Now let us fix an arbitrary difference graph \( D_{n_0, n_1} \), and consider the probability of generating that particular graph via each of the two random models.

First, when using the creation sequence model \( D(C_{n_0, n_1}) \), each of the \( \dbinom{n_0 + n_1}{n_0} \) sequences is equally likely. So the probability of generating \( \text{seq}(D_{n_0, n_1}) \) is:
\[
P(D(C_{n_0, n_1}) = D_{n_0, n_1}) = \frac{1}{\binom{n_0 + n_1}{n_0}} = \frac{n_0!n_1!}{(n_0 + n_1)!}
\]

On the other hand, when using the random weighting model \( D(\vec{w}) \), a random vector \( \vec{w} \) generates \( D_{n_0, n_1} \) if and only if it lies in a region \( R \) such that the third coordinate of \( f(R) \) is the creation sequence of \( D_{n_0, n_1} \). Thus,
\[
P(D(\vec{w}) = D_{n_0, n_1}) = n_0!n_1!P(\vec{w} \in f^{-1}(1_{n_0}, 1_{n_1}, \text{seq}(D_{n_0, n_1}))) = \frac{n_0!n_1!}{(n_0 + n_1)!}
\]

Therefore, the two models of generating difference graphs have the exact same distribution.
CHAPTER 4
DIFFERENCE GRAPH PROPERTIES

4.1 Background

The power of the previous chapter is that we can now discard the continuous variables of the traditional difference graph model. As in Chapter 2, we need only consider the creation sequence.

For example, in Proposition 4.3.1, we show that a difference graph is connected if and only if no vertices are isolated. If we restricted ourselves to the continuous model, calculating the probability of connectedness would require us to find the measure of some set $R \subseteq [0, 1]^{n_0+n_1}$, where for all weight vectors $\vec{w} \in R$, the threshold graph corresponding to $\vec{w}$ has no isolated vertices. Which in turn means that the maximum of the first $n_0$ coordinates of $\vec{w}$ is greater than 1 minus the minimum of the last $n_1$ coordinates, and vice-versa. So even determining the likelihood of one of the simplest of graph properties requires delving into order statistics; not impossible, but comparatively awkward.

Instead, for each of the graph properties, we begin by determining those characteristics of the creation sequence which are necessary and sufficient for the resulting difference graph to have said property. There are only a finite number of sequences that encode difference graphs of a given size, so by using combinatorial arguments to enumerate those which fit the specified characteristics, we can invoke the uniformity
of the distribution to determine the probability of a random difference graph having this property.

The major difference between the creation sequence models of threshold and difference graphs lies in the construction of their sequences. In the former case, we dealt with graphs with a fixed number of vertices, denoted $n$. In the latter, not only is the number of vertices fixed, but so are the number of vertices in each part of the bipartite graph. This means that the composition of the creation sequence is fixed, with prescribed numbers of zeroes and ones. So whereas the digits of the creation sequence $C_n$ are independent and chosen uniformly at random from $\{0, 1\}$, the digits of the $C_{n_0,n_1}$ are not independent.

4.2 Fundamentals

Please note that although we are, as usual, dealing with unlabeled graphs, it is nevertheless useful to refer to specific vertices in the course of a proof. Thus, the name $z_i$ is used instead of the more unwieldy “vertex in Z corresponding to the $i$-th zero from the left in the creation sequence”. In this manner, $Z = \{z_1, \ldots, z_{n_0}\}$, with each $z_i$ lying to the left of $z_{i+1}$ in the creation sequence. Similarly, $U = \{u_1, \ldots, u_{n_1}\}$, again counting from left to right. The important distinction is that the assignment of these labels is a consequence of the structure of the graph, rather than being an independent property.

Given a vertex in $D_{n_0,n_1}$, we will often refer to its index, which is its position in the creation sequence. Indices run from 1 to $n_0 + n_1$ and increase from right to left. As a consequence of this enumeration and the vertex labels, note that if there exists an edge $\{z_{i+1}, u_j\}$, then there exist edges $\{z_{i+1}, u_{j+1}\}$ and $\{z_i, u_j\}$. This follows because the zero corresponding to $z_i$ has lower index than the zero corresponding to $z_{i+1}$, so any one-vertex adjacent to $z_{i+1}$ is also adjacent to $z_i$. 47
It will also prove useful to analyze the “tail” of a sequence $S$, which consists of the subsequence of $S$ that includes all digits after a certain index. In particular, if $S$ is the $n$-long sequence of the form

$$S = s_1s_2 \ldots s_{n-1}s_n,$$

then the $i$-th tail of $S$ is the sequence $s_is_{i+1} \ldots s_n$. In the same vein, let $t_i^z(S)$ and $t_i^u(S)$ denote the number of zeroes and ones, respectively, in the $i$-th tail of $S$. Then for any $i$ such that $1 \leq i \leq n_0 + n_1$, $t_i^u(C_{n_0,n_1}) + t_i^z(C_{n_0,n_1}) = n_0 + n_1 - i + 1$.

Given a binary sequence $S$, we let $D(S)$ denote the difference graph whose creation sequence is $S$. Conversely, given a difference graph $D$, let $\text{seq}(D)$ denote the creation sequence of $D$.

### 4.3 Connectivity

**Proposition 4.3.1.** Let $D$ be a difference graph with at least two vertices. Then $D$ is connected if and only if no vertex of $D$ is isolated.

**Proof.** Note that the vertex requirement serves to remove from consideration all difference graphs with fewer than two vertices; as such graphs are trivially connected, they are safely ignored.

By Proposition 3.1.1, we know that $Z$ lacks isolated vertices if and only if $U$ has a vertex that dominates all of $Z$. A similar argument shows that $U$ lacks isolated vertices if and only if some vertex of $Z$ dominates all of $U$. Thus, we have two mutually exclusive possibilities: either there exists an isolated vertex of $D$, or $U$ and $Z$ each contain a vertex that dominates the other.

In this second case, let $u \in U$ and $z \in Z$ denote the dominating vertices. Then for any two vertices of $D$, there exists a path between them through $u$ or $z$, and therefore $D$ is connected. 

\[\square\]
With this, we can use the uniform distribution on the set of creation sequences to determine the probability of connectivity:

**Proposition 4.3.2.** Let $D_{n_0,n_1}$ be a random difference graph such that $n_0 + n_1 \geq 2$. Then

$$P(D \text{ is connected}) = \frac{n_0n_1}{(n_0 + n_1)(n_0 + n_1 - 1)}$$

**Proof.** As shown in Proposition 4.3.1, $D$ is connected if and only if no vertex is isolated. By the construction of the creation sequence, a zero-vertex is isolated when there are no one-vertices of higher index; that is, there are no ones lying to the right of the corresponding zero in the creation sequence. Similarly, a one-vertex is isolated when there are no zero vertices of lower index.

Thus, to ensure that there are no isolated vertices of either part, we require that a zero-vertex have the lowest index and a one-vertex has the highest index. The remaining vertices, arranged between these two extremes, can have any configuration possible, as that will not impact the connectivity. Therefore,

$$P(D \text{ is connected}) = \frac{(n_0+n_1-2)}{(n_0+1)} = \frac{n_0n_1}{(n_0 + n_1)(n_0 + n_1 - 1)}$$

We can generalize the above results to broader statements about the connectivity of $D_{n_0,n_1}$, considering both vertex-connectivity and edge-connectivity. Starting with the former, a graph is *k*-vertex-connected if it remains connected after the removal of up to $k - 1$ vertices. (As vertex-connectivity is a more common concern than edge-connectivity, such a graph is usually simply said to be *k*-connected).

**Proposition 4.3.3.** Let $D_{n_0,n_1}$ be a difference graph. Then $D_{n_0,n_1}$ is *k*-vertex-connected if and only if $C_{n_0,n_1}$ begins with at least $k$ consecutive zeroes and ends with at least $k$ consecutive ones.
Proof. First, suppose that $D_{n_0,n_1}$ is $k$-connected. Then every vertex $v$ in $D_{n_0,n_1}$ must have degree at least $k$, as otherwise the set of $v$’s neighbors constitutes a set of size less than $k$ whose removal disconnects the graph. So every zero-vertex must have at least $k$ one-vertices of higher index, and every one-vertex must have at least $k$ zero-vertices of lower index. Therefore, the creation sequence must begin with at least $k$ consecutive zeroes, and end with at least $k$ consecutive ones.

Second, suppose that $C_{n_0,n_1}$ begins with at least $k$ consecutive zeroes and ends with at least $k$ consecutive ones. Let $V = V(D_{n_0,n_1})$ denote the vertex set of $D_{n_0,n_1}$, and let $A = \{a_1, \ldots, a_{k-1}\} \subset V$ be a subset of $k-1$ vertices of $V$; we shall show that the induced subgraph on $V \setminus A$ remains connected.

Since $C_{n_0,n_1}$ begins with $k$ zeroes, every element of $\{z_1, \ldots, z_k\}$ dominates all of $U$. Similarly, every element of $\{u_{n_1-k+1}, \ldots, u_{n_1}\}$ dominates all of $Z$. Since $A$ consists of only $k-1$ elements, there must exist some $z_i \in V \setminus A$ that dominates all of $U \setminus A$ and some $u_j \in V \setminus A$ that dominates all of $Z \setminus A$. Thus, the subgraph on $V \setminus A$ is connected, and therefore $D_{n_0,n_1}$ is $k$-connected.

Corollary 4.3.4. Let $D_{n_0,n_1}$ be a random difference graph. Then

$$P(D_{n_0,n_1} \text{ is } k\text{-vertex-connected}) = \frac{\binom{n_0+n_1-2k}{n_0-k}}{\binom{n_0+n_1}{n_0}} = \frac{(n_0)_k(n_1)_k}{(n_0+n_1)_2k},$$

where $(n)_k$ is the falling factorial $(n)(n-1)\cdots(n-k+1)$.

This, in turn, enables us to calculate the vertex connectivity $\kappa(D_{n_0,n_1})$, the maximum $k$ such that $D_{n_0,n_1}$ is $k$-vertex-connected.

Corollary 4.3.5. Let $D_{n_0,n_1}$ be a random difference graph. Then

$$P(\kappa(D_{n_0,n_1}) = k) = \binom{n_0+n_1}{n_0}^{-1}\left(\binom{n_0+n_1-2k}{n_0-k} - \binom{n_0+n_1-2k-2}{n_0-k-1}\right)$$
Proof. For the connectivity of $D_{n_0,n_1}$ to equal $k$, $D_{n_0,n_1}$ must be $k$-connected but not $(k+1)$-connected. Since the event of $D_{n_0,n_1}$ being a $(k+1)$-connected graph necessitates that it is also $k$-connected,

$$P(\kappa(D_{n_0,n_1}) = k) = P(D_{n_0,n_1} \text{ is } k\text{-conn.}) - P(D_{n_0,n_1} \text{ is } (k+1)\text{-conn.})$$

$$= \binom{n_0 + n_1 - 2k}{n_0 - k} - \binom{n_0 + n_1 - 2(k+1)}{n_0 - (k+1)}$$

$$= \binom{n_0 + n_1}{n_0} \left( \frac{n_0 + n_1 - 2k}{n_0 - k} - \frac{n_0 + n_1 - 2(k+1)}{n_0 - k - 1} \right)$$

Now we can look at edge connectivity, denoted $\lambda(D_{n_0,n_1})$, which is the minimum number of edges whose removal disconnects $D_{n_0,n_1}$. Clearly, $\lambda(D_{n_0,n_1})$ is less than or equal to the minimum degree of all vertices in $D_{n_0,n_1}$, which we denote by $\delta(D_{n_0,n_1})$; removing all edges incident to a vertex will disconnect that vertex. Before we can strengthen the relationship between those functions, however, we need another known result concerning $\lambda$:

Lemma 4.3.6. For a complete bipartite graph $K_{n,n}$, $\lambda(K_{n,n}) = n$.

Proof. We proceed by induction. In the base case of $n = 1$, $K_{1,1}$ is definitely 1-edge-connected, but not 2-edge-connected, as the removal of the single edge disconnects the graph.

Next, assume that $\lambda(K_{n-1,n-1}) = n - 1$, and consider $K_{n,n}$. We denote the two parts of $K_{n,n}$ as $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$. Now remove $n - 1$ edges from $K_{n,n}$.

Case 1: None of the removed edges are incident to $a_n$ or $b_n$. Then $K_{n,n}$ is still connected, as each vertex of $A$ is adjacent to $b_n$, and each vertex of $B$ is adjacent to $a_n$. 

51
Case 2: At least one of the removed edges is incident to \(a_n\) or \(b_n\). Then at most \(n - 2\) of the removed edges are from the \(K_{n-1,n-1}\) induced by \(A \setminus \{a_n\}\) and \(B \setminus \{b_n\}\), so by induction, those \(2n - 2\) vertices remain part of a single connected component. Furthermore, both \(a_n\) and \(b_n\) are connected to said component: even if \(a_n\)’s only remaining neighbor is \(b_n\), then all \(k - 1\) of the removed edges must be amongst those incident to \(a_n\), and therefore \(b_n\) must be adjacent to every member of \(A\).

Either way, we see that \(K_{n,n}\) remains connected after the removal of \(n - 1\) edges, and therefore \(\lambda(K_{n,n}) = n\).

Proposition 4.3.7. Let \(D_{n_0,n_1}\) be a difference graph. Then \(\lambda(D_{n_0,n_1}) = \delta(D_{n_0,n_1})\).

Proof. Above, we saw that \(\lambda(D_{n_0,n_1}) \leq \delta(D_{n_0,n_1})\). For the opposite direction, let us suppose that \(\delta(D_{n_0,n_1}) = k\). Since every vertex has degree at least \(k\), \(D_{n_0,n_1}\) must contain a complete bipartite graph \(K_{k,k}\). Specifically, every vertex in \(\{z_1, \ldots, z_k\}\) is adjacent to every vertex in \(\{u_{n_1 - k + 1}, \ldots, u_{n_1}\}\). Let \(H\) denote the subgraph of \(D_{n_0,n_1}\) induced by those \(2k\) vertices.

From Lemma 4.3.6, \(H\) remains connected after the removal of up to \(k - 1\) edges. Furthermore, since every vertex in \(D_{n_0,n_1}\) is adjacent to at least \(k\) vertices of \(H\), there exists a path from any vertex to some vertex of \(H\) after the removal of \(k - 1\) edges. Thus, \(D_{n_0,n_1}\) is still connected after removing \(k - 1\) edges, and so \(\lambda(D_{n_0,n_1}) \geq \delta(D_{n_0,n_1})\).

Proposition 4.3.8. Let \(D_{n_0,n_1}\) be a random difference graph. Then

\[
P(\lambda(D_{n_0,n_1}) = k) = P(\delta(D_{n_0,n_1}) = k) = \binom{n_0 + n_1}{n_0}^{-1} \left[ \binom{n_0 + n_1 - 2k}{n_0 - k} - \binom{n_0 + n_1 - 2k - 2}{n_0 - k - 1} \right]
\]

Proof. As \(D_{n_0,n_1}\) is a bipartite graph, its minimum degree must be the smaller of the minimum degrees of each part. The vertex \(z_{n_0}\) has the minimum degree of all vertices.
in $Z$, and $u_1$ the minimum of all vertices in $U$. As $z_{n_0}$ corresponds to the right-most zero in $C_{n_0,n_1}$, its degree is equal to the number of consecutive ones at the end of the sequence. Similarly, the degree of $u_1$ equals the number of consecutive zeroes at the beginning.

Letting $i$ denote the number of initial zeroes and $j$ the number of terminal ones,

$$P(\delta(D_{n_0,n_1}) = k) = P(i = k \text{ and } j \geq k) + P(j = k \text{ and } i \geq k) - P(i = j = k),$$

which raises the following question: How many binary sequences of $a$ zeroes and $b$ ones do not start with a zero or end with a one?

If $a = b = 0$, the answer is one. If exactly one of the two is zero, then there are no such sequences. But if both $a$ and $b$ are positive, then the answer is $\binom{a+b-2}{a-1}$. Because of the awkwardness at $a = b = 0$, we must temporarily divide our calculation into cases.

If $k < \min(n_0,n_1)$, then $P(\delta(D_{n_0,n_1}) = k)$ becomes

$$\binom{n_0 + n_1}{n_0}^{-1} \left[ \sum_{j=k}^{n_1-1} \binom{n_0 - k + n_1 - j - 2}{n_0 - k - 1} + \sum_{i=k}^{n_0-1} \binom{n_0 - i + n_1 - k - 2}{n_1 - k - 1} - \binom{n_0 + n_1 - 2k - 2}{n_0 - k - 1} \right],$$

which simplifies to

$$\binom{n_0 + n_1}{n_0}^{-1} \left[ \binom{n_0 + n_1 - 2k}{n_0 - k} - \binom{n_0 + n_1 - 2k - 2}{n_0 - k - 1} \right].$$

On the other hand, if $k = \min(n_0,n_1)$, then there is only one sequence that can produce a minimum degree of $k$, namely

$$C_{n_0,n_1} = \underbrace{0 \cdots 0}_{n_0} \underbrace{1 \cdots 1}_{n_1}.$$
so $P(\delta(D_{n_0,n_1}) = k) = \binom{n_0+n_1}{n_0}^{-1}$.

However, this is consistent with taking $k = \min(n_0, n_1)$ in the previous case, so therefore

$$P(\delta(D_{n_0,n_1}) = k) = \binom{n_0+n_1}{n_0}^{-1} \left[ \binom{n_0+n_1-2k}{n_0-k} - \binom{n_0+n_1-2k-2}{n_0-k-1} \right]$$

\[ \square \]

### 4.4 Degrees and Coloring

As part of Proposition 4.3.8, we found the distribution of the minimum degree $\delta(D_{n_0,n_1})$ over all vertices in a random difference graph $D_{n_0,n_1}$. Here, we start by finding the distribution of the maximum degree, denoted $\Delta(D_{n_0,n_1})$. This is a rather more troublesome concept than the minimum degree, and requires us to introduce a construction which will be used both here and in subsequent sections.

Recall that a difference graph $D$ is connected if and only if $\text{seq}(D)$ begins with a 0 and ends with a 1. In this vein, we define the interior of creation sequence $S$, denoted $\text{int}(S)$, to be the subsequence that lies after the left-most 0 and before the right-most 1. That is, if $i$ is the index of the left-most 0 and $j$ the index of the right-most 1, then $\text{int}(S) = c_{i+1}c_{i+2}\ldots c_{j-2}c_{j-1}$.

Essentially, $\text{int}(D)$ contains those digits corresponding to the non-trivial connected component of $D$, with the exception of a single dominating vertex from each part. Note that there are two cases in which $\text{int}(D)$ is the empty sequence: either $D$ is the empty graph, or the non-trivial connected component is exactly two vertices.

Since $\text{int}$ removes the right-most one and the left-most zero, $\text{int}(C_{n_0,n_1})$ has at most $n_0 - 1$ zeroes and $n_1 - 1$ ones. Furthermore, for every binary sequence $S$ with
i zeroes and j ones, where $0 \leq i \leq n_0 - 1$ and $0 \leq j \leq n_1 - 1$, there exists a $C_{n_0,n_1}$ such that $\text{int}(C_{n_0,n_1}) = S$. To see this, we define $C_{n_0,n_1}$ as follows:

$$C_{n_0,n_1} = \underbrace{0 \cdots 0}_{n_0-i-1} \underbrace{1 \cdots 1}_{n_1-j-1}$$

If we attempt to count the number of sequences $S$ that are obtainable as some $\text{int}(C_{n_0,n_1})$, we find that

$$\sum_{i=0}^{n_0-1} \sum_{j=0}^{n_1-1} \binom{i+j}{i} = \sum_{i=0}^{n_1-1} \binom{i + n_1}{i + 1} = \binom{n_0 + n_1}{n_0} - 1.$$

Recalling that there are two distinct creation sequences with empty interior, this proves that every possible non-empty sequence occurs exactly once as the interior of some $C_{n_0,n_1}$.

**Theorem 4.4.1.** For a difference graph $D_{n_0,n_1}$ with $n_0 + n_1 \geq 1$, the probability that $\Delta(D_{n_0,n_1}) = k$ equals, for $1 \leq k \leq \max(n_0, n_1)$,

$$\begin{cases} 
\binom{n_0 + n_1}{n_0}^{-1} & k = 0 \\
\binom{n_0 + n_1}{n_0}^{-1} \left( 2 \binom{2k-1}{k-1} - \binom{2k-2}{k-1} \right) & 1 \leq k \leq \min(n_0, n_1) \\
\binom{n_0 + n_1}{n_0}^{-1} \binom{\min(n_0, n_1) + k - 1}{k} & \min(n_0, n_1) < k \leq \max(n_0, n_1)
\end{cases}$$

**Proof.** Note that as $D_{n_0,n_1}$ is bipartite, the maximum degree of any one-vertex is the number of zero-vertices, and vice-versa. Thus, the maximum degree can never be larger than the maximum of $n_0$ and $n_1$. On the other extreme, the only way for the maximum degree to be zero is for $D_{n_0,n_1}$ to be the empty graph, which happens in exactly one way, thus having probability $\binom{n_0 + n_1}{n_0}^{-1}$.

The maximum degree amongst the zero-vertices is equal to the number of one-vertices whose digits have higher index than that of the left-most zero in $C_{n_0,n_1}$. That is, the maximum degree of any zero-vertex is $n_1$ minus the number of initial ones in $C_{n_0,n_1}$. Similarly, the maximum degree of any one-vertex is $n_0$ minus the number of
terminal zeroes in $C_{n_0,n_1}$, the number of zeroes that have a higher index than any one-vertex.

Therefore, $\Delta(D_{n_0,n_1}) \leq k$ if and only if $C_{n_0,n_1}$ starts with $\max(0,n_1-k)$ consecutive ones and ends with $\max(0,n_0-k)$ consecutive zeroes. Equivalently, $\Delta(D_{n_0,n_1}) \leq k$ if and only if $\text{int}(C_{n_0,n_1})$ has at most $k-1$ zeroes and at most $k-1$ ones; this corresponds to the primary connected component of $D_{n_0,n_1}$ consisting of at most $k$ zero-vertices and at most $k$ one-vertices.

As such, for $k \geq 1$, $\Delta(D_{n_0,n_1}) = k$ if and only if the minimum of the number of zeroes and the number of ones in $\text{int}(C_{n_0,n_1})$ is $k-1$. Decomposing this event, we see that the corresponding probability is given by:

$$P(\Delta(D_{n_0,n_1}) = k) = P(\text{int}(C_{n_0,n_1}) \text{ has } k-1 \text{ zeroes, } \leq k-1 \text{ ones})$$

$$+ P(\text{int}(C_{n_0,n_1}) \text{ has } k-1 \text{ ones, } \leq k-1 \text{ zeroes})$$

$$- P(\text{int}(C_{n_0,n_1}) \text{ has } k-1 \text{ zeroes and } k-1 \text{ ones})$$

Here we must exercise caution, as one or more of those probabilities may be zero. For example, if $n_0 < k < n_1$, it would be impossible to have $k-1$ zeroes in the interior of $C_{n_0,n_1}$. But in the event that $k \leq \min(n_0,n_1)$, the total probability can be written as:

$$\binom{n_0+n_1}{n_0}^{-1} \left[ \sum_{j=0}^{k-1} \binom{j+k-1}{j} + \sum_{i=0}^{k-1} \binom{i+k-1}{i} - \binom{2k-2}{k-1} \right]$$

On the other hand, for $\min(n_0,n_1) < k \leq \max(n_0,n_1)$, at two of the three components are zero, and the third is equal to

$$\binom{n_0+n_1}{n_0}^{-1} \sum_{j=0}^{\min(n_1,n_0)-1} \binom{j+k-1}{j}$$

After simplification through the identity

$$\sum_{i=0}^{m} \binom{i+a}{i} = \binom{m+a+1}{a+1},$$

we arrive at the distribution claimed above. □
Now we turn to the distribution of the coloring numbers of $D_{n_0, n_1}$.

First we consider the distribution of $\chi(D_{n_0, n_1})$, the chromatic number of $D_{n_0, n_1}$. Also known as the vertex-coloring number, this is the smallest number of colors necessary to assign a color each vertex under the restriction that there are no edges between vertices of similar color.

**Proposition 4.4.2.** Let $D_{n_0, n_1}$ be a difference graph with $n_0 + n_1 \geq 1$. Then

$$P(\chi(D_{n_0, n_1}) = k) = \begin{cases} 
1 - \left( \frac{n_0 + n_1}{n_0} \right)^{-1} & \text{for } k = 2 \\
\left( \frac{n_0 + n_1}{n_0} \right)^{-1} & \text{for } k = 1 \\
0 & \text{otherwise}
\end{cases}$$

**Proof.** As $D_{n_0, n_1}$ is a bipartite graph, $\chi(D_{n_0, n_1})$ is at most equal to 2: simply assign one color to all vertices in $U$, and another color to all of $Z$. For $\chi$ to return a value less than 2, there can be no edges, and $D_{n_0, n_1}$ must be the empty graph. As there is only one creation sequence to generate such a graph, $P(\chi(D_{n_0, n_1}) = 1) = \frac{1}{(n_0 + n_1)}$, and all other creation sequences produce graphs with a chromatic number of 2.

Next we turn to the chromatic index, or edge-coloring number, denoted $\chi'(D_{n_0, n_1})$. This is the least number of colors required to color every edge, under the restriction that no vertex is an endpoint of two like-colored edges. By König’s Bipartite Theorem [10] (also known as König’s Line Coloring Theorem), the chromatic index of $D_{n_0, n_1}$ equals the maximum degree amongst all vertices of $D_{n_0, n_1}$.

**Corollary 4.4.3.** Let $D_{n_0, n_1}$ be a random difference graph with $n_0 + n_1 \geq 2$. Then the probability that $\chi'(D_{n_0, n_1}) = k$ is, for $1 \leq k \leq \max(n_0, n_1)$, given by

$$\left( \frac{n_0 + n_1}{n_0} \right)^{-1} \left[ \left( \frac{\min(n_1, k) + k - 1}{k} \right) + \left( \frac{\min(n_0, k) + k - 1}{k} \right) - \left( \frac{2k - 2}{k - 1} \right) \right]$$

**Proof.** Simply apply König’s Bipartite Theorem to Proposition 4.4.1. □
4.5 Matching Number

Using techniques similar to those employed in the threshold graph analogue, here we shall find the distribution of the size of the maximum matching of a difference graph, denoted \( \nu(D_{n_0,n_1}) \).

We once again employ function \( h \), defined above for any binary sequence \( S \) as

\[
h(S) = \max_{0 \leq i \leq |S|+1} \{ t^z_i(S) - t^u_i(S) \}
\]

That is, \( h(S) \) is the maximum number of excess zeroes across all tails of sequence \( S \).

**Proposition 4.5.1.** For a difference graph \( D_{n_0,n_1} \) with creation sequence \( C_{n_0,n_1} \) and matching number \( \nu(D_{n_0,n_1}) \),

\[
n_0 - \nu(D_{n_0,n_1}) = h(C_{n_0,n_1})
\]

*Proof.* First, since \( D_{n_0,n_1} \) is a bipartite graph with parts of size \( n_0 \) and \( n_1 \), the maximum size of any possible matching cannot exceed the minimum of the size of either part. As \( h \) is always non-negative, our claim satisfies the restriction that \( \nu(D_{n_0,n_1}) \leq n_0 \). Furthermore, if \( n_1 \leq n_0 \), then \( h(C_{n_0,n_1}) \geq n_0 - n_1 \), which would ensure that \( \nu(D_{n_0,n_1}) \leq n_1 \). But regardless of these constraints, we can still interpret the left-hand side as the number of unmatched zero-vertices in any maximum matching.

Let \( j \) denote a maximizing index for \( h \) such that \( h(C_{n_0,n_1}) = t^z_j(C_{n_0,n_1}) - t^u_j(C_{n_0,n_1}) \). Then the tail beginning at \( j \) has \( h(C_{n_0,n_1}) \) more zeroes than ones. Since any zero-vertex is adjacent only to one-vertices of higher index, the zero-vertices in the tail can only be matched to one-vertices in the same tail. So at least \( h(C_{n_0,n_1}) \) of them must remain unmatched in any matching, and therefore \( n_0 - \nu(D_{n_0,n_1}) \geq h(C_{n_0,n_1}) \).

For the other direction, let us create a new binary sequence \( S \) by removing the right-most \( h(C_{n_0,n_1}) \) zeroes from creation sequence \( C_{n_0,n_1} \). Then \( S \) has \( n_0 - h(C_{n_0,n_1}) \)
zeroes and $n_1$ ones, and by construction, $h(S) = 0$, so every tail of $S$ contains at least as many ones as zeroes. So within $S$, for all $k$ such that $1 \leq k \leq n_0 - h(C_{n_0,n_1})$, the $k$-th zero from the right lies to the left of the $k$-th one from the right. (Equivalently, we could say that the index of $z_{n_0-k+1}$ is less than the index of $u_{n_1-k+1}$.)

Thus there are $n_0 - h(C_{n_0,n_1})$ mutually disjoint pairs of zeroes and ones in $S$ where in each pair, the zero lies to the left of the one. Since $S$ was created by removing digits from $C_{n_0,n_1}$, the original sequence must also have at least that many pairs, and $\nu(D_{n_0,n_1}) \geq n_0 - h(C_{n_0,n_1})$.

Having found the properties of the creation sequence that correspond to the matching number $\nu(D_{n_0,n_1})$, we can calculate the distribution.

**Theorem 4.5.2.** For a random difference graph $D_{n_0,n_1}$, the distribution of the matching number $\nu(D_{n_0,n_1})$ is given by

$$P(\nu(D_{n_0,n_1}) = k) = \frac{(n_0 + n_1 - 2k + 1)n_0!n_1!}{k!(n_0 + n_1 - k + 1)!},$$

for $0 \leq k \leq \min(n_0, n_1)$.

**Proof.** We begin by recalling that $\nu(D_{n_0,n_1}) = n_0 - h(C_{n_0,n_1})$. Then $\nu(D_{n_0,n_1}) = k$ if and only if $h(C_{n_0,n_1}) = n_0 - k$. Therefore, the probability that $\nu(D_{n_0,n_1}) = k$ is proportional to the number of binary sequences of $n_0$ zeroes and $n_1$ ones where some tail contains exactly $n_0 - k$ more zeroes than ones, but no tail contains a difference greater than that.

Reading the sequence from right to left, there is a clear bijection between the number of such sequences and the number of staircase walks from $(0, 0)$ to $(n_1, n_0)$ that touch, but do not cross, the line $y = x + n_0 - k$. From Proposition 2.4.3, this number is exactly

$$\binom{n_1 + n_0}{k} - \binom{n_1 + n_0}{k-1}.$$
At this point, we can remove the assumption that $n_0 \leq n_1$; counting the number of such sequences under the assumption $n_1 \leq n_0$ would, by the interchangability of $n_0$ and $n_1$ in the above result, change nothing. Therefore,

$$P(\nu(D_{n_0,n_1}) = k) = \frac{1}{\binom{n_0+n_1}{n_0}} \left( \binom{n_1+n_0}{n_1+n_0-k} - \binom{n_1+n_0}{n_1+n_0-k+1} \right)$$

$$= \frac{(n_0 + n_1 - 2k + 1)n_0!n_1!}{k!(n_0 + n_1 - k + 1)!}$$

4.6 Coverings

A vertex cover of $D_{n_0,n_1}$ is a set of vertices such that every edge of $D_{n_0,n_1}$ is incident to at least one vertex in the set. The vertex covering number, denoted $\tau(D_{n_0,n_1})$, is the minimum size amongst all vertex covers.

By König’s Theorem, the matching number of a bipartite graph equals the vertex covering number.

**Proposition 4.6.1.** For a random difference graph $D_{n_0,n_1}$,

$$P(\tau(D_{n_0,n_1}) = k) = \frac{(n_0 + n_1 - 2k + 1)n_0!n_1!}{k!(n_0 + n_1 - k + 1)!}$$

**Proof.** By König’s Theorem and Proposition 4.5.2. □

An edge cover of $D_{n_0,n_1}$ is a set of edges such that every vertex of $D_{n_0,n_1}$ is incident to at least one edge in the set. The edge covering number, denoted $\rho(D_{n_0,n_1})$, is the minimum size amongst all edge covers.

Awkwardly, if a graph contains any vertices of degree zero, then no edge coverings exist; vertices that lack edges cannot be covered by any edge. Thus, we generalize to a different concept, the the independence number.
An independent set of $D_{n_0, n_1}$ is a set of vertices such that no two vertices in the set are adjacent. The *independence number*, denoted $\alpha(D_{n_0, n_1})$, is the size of the largest independent set.

**Proposition 4.6.2.** For a random difference graph $D_{n_0, n_1}$ and $k \geq \max(n_0, n_1)$,

$$P(\alpha(D_{n_0, n_1}) = k) = \frac{(2k - n_0 - n_1 + 1)n_0!n_1!}{(n_0 + n_1 - k)!(k + 1)!}$$

*Proof.* First, since $D_{n_0, n_1}$ is a bipartite graph, the sets $U$ and $Z$ are each independent, and therefore $\alpha(D_{n_0, n_1}) \geq \max(n_0, n_1)$. Hence the restriction on $k$.

Note that since there are no edges between vertices in an independent set, every edge must be incident to some vertex in the complement of said set. Thus, the complement of an independent set is a vertex cover, and the independence number plus the vertex covering number equals the number of vertices.

So $\alpha(D_{n_0, n_1}) = k$ if and only if $\tau(D_{n_0, n_1}) = n_0 + n_1 - k$, and therefore by Proposition 4.6.1,

$$P(\alpha(D_{n_0, n_1}) = k) = P(\tau(D_{n_0, n_1}) = n_0 + n_1 - k)$$

$$= \frac{(n_0 + n_1 - 2(n_0 + n_1 - k) + 1)n_0!n_1!}{(n_0 + n_1 - k)!(n_0 + n_1 - (n_0 + n_1 - k) + 1)!}$$

$$= \frac{(2k - n_0 - n_1 + 1)n_0!n_1!}{(n_0 + n_1 - k)!(k + 1)!}$$

Using the independence number, we can now return to the edge covering number. For a connected bipartite graph, the edge covering number equals the size of the maximum independent set.

For a connected bipartite graph, the edge covering number equals the number of vertices minus the vertex covering number.
Proposition 4.6.3. For a random connected bipartite graph \( D_{n_0,n_1} \) with \( n_0 + n_1 \geq 2 \) and \( k \geq \max(n_0, n_1) \), the probability that \( \rho(D_{n_0,n_1}) = k \) is given by

\[
\left(\frac{n_0 + n_1 - 2}{n_0 - 1}\right)^{-1} \left(\frac{n_0 + n_1 - c(k)}{n_1 + k - \max(n_0, n_1)}\right) - \left(\frac{n_0 + n_1 - c(k)}{n_1 + k - \max(n_0, n_1)}\right),
\]

where \( c(k) \) is given by

\[
c(k) = \begin{cases} 
0 & \text{if } k = n_0 = n_1 \\
2 & \text{if } \max(n_0, n_1) < k < 2n_0 - n_1 \\
1 & \text{otherwise}
\end{cases}
\]

Proof. Untangling the various theorems, we see that

\[
P(\rho(D_{n_0,n_1}) = k) = P(\alpha(D_{n_0,n_1}) = k | D_{n_0,n_1} \text{ connected})
\]

\[
= P(\tau(D_{n_0,n_1}) = n_0 + n_1 - k | D_{n_0,n_1} \text{ connected})
\]

\[
= P(\nu(D_{n_0,n_1}) = n_0 + n_1 - k | D_{n_0,n_1} \text{ connected})
\]

\[
= P(h(C_{n_0,n_1}) = k - \max(n_0, n_1) | D_{n_0,n_1} \text{ connected})
\]

Recalling that \( D_{n_0,n_1} \) is connected if and only if its creation sequence \( C_{n_0,n_1} \) both begins with a 0 and ends with a 1, we now count the number of such sequences \( S \) with the additional property that \( h(S) = k - \max(n_0, n_1) \). To do this, we use the same bijection between sequences and walks that we did when computing the distribution of the matching number.

We count the number of staircase walks from \((0,0)\) to \((n_1, n_0)\) that begin with a move to the right, end with a move upwards, and touch, but do not cross, the line \( y = x + k - \max(n_0, n_1) \). If \((0,0)\) is on said line, meaning that \( k = \max(n_0, n_1) \), then the first move would have had to be rightwards anyway, to prevent the walk from crossing the line. Similarly, if \((n_1, n_0)\) is on the line \( y = x + k - \max(n_0, n_1) \), then the final move has to be upwards because a rightwards move would necessitate the
walk having already crossed the line. However, note that as \( k \geq \max(n_0, n_1) \), for the
point \((n_1, n_0)\) to be on the line means that \( n_0 \geq n_1 \), and thus \( k = 2n_0 - n_1 \).

To summarize, for \( k \geq \max(n_0, n_1) \): if \( k = \max(n_0, n_1) \), then the walk starts at
(0,0); otherwise, it begins at (1,0). If \( k = 2n_0 - n_1 \), then the walk ends at \((n_1, n_0)\);
otherwise it ends at \((n_1, n_0 - 1)\). We can now make extensive use of Lemma 2.4.3:

If \( k = \max(n_0, n_1) = 2n_0 - n_1 \), then \( k = n_0 = n_1 \), so the number of walks is
\[
\binom{2n_0}{n_0} - \binom{2n_0}{n_0 + 1}
\]

If \( k = \max(n_0, n_1) \neq 2n_0 - n_1 \), then the number of walks is
\[
\binom{n_0 + n_1 - 1}{n_1 + k - \max(n_0, n_1)} - \binom{n_0 + n_1 - 1}{n_1 + k - \max(n_0, n_1) + 1}
\]

If \( k > \max(n_0, n_1) \) and \( k = 2n_0 - n_1 \), then \( k = n_0 > n_1 \), so we count the number
of walks from \((1,0)\) to \((n_1, n_0)\) that merely touch \( y = x + k - n_0 \). Shifting the
\( x \)-coordinate by one unit,
\[
\binom{n_0 + n_1 - 1}{n_1 + k - n_0} - \binom{n_0 + n_1 - 1}{n_1 + k - n_0 + 1}
\]

If \( k > \max(n_0, n_1) \) and \( k \neq 2n_0 - n_1 \), then walking from \((1,0)\) to \((n_1, n_0 - 1)\) while
touching \( y = x + k - \max(n_0, n_1) \) becomes, after a small shift in each coordinate,
a walk from \((0,0)\) to \((n_1 - 1, n_0 - 1)\) that touches, but does not cross, the line
\( y = x + k - \max(n_0, n_1) + 1 \).
\[
\binom{n_0 + n_1 - 2}{n_1 + k - \max(n_0, n_1)} - \binom{n_0 + n_1 - 2}{n_1 + k - \max(n_0, n_1) + 1}
\]

Finally, since we are conditioning on the probability that \( D_{n_0, n_1} \) is connected, we
get
\[
\left(\binom{n_0 + n_1 - 2}{n_0 - 1}\right)^{-1} \left(\binom{n_0 + n_1 - 2}{n_0 + n_1 - k} - \binom{n_0 + n_1 - 2}{n_0 + n_1 - k + 1}\right)
\]
4.7 Cycles

When looking at cycles within a difference graph $D_{n_0,n_1}$, it is important to remember that, as a bipartite graph, $D_{n_0,n_1}$ cannot have any cycles of odd length.

First, we determine the probability that $D_{n_0,n_1}$ has a cycle.

**Proposition 4.7.1.** For a difference graph $D_{n_0,n_1}$, there exists a cycle in $D_{n_0,n_1}$ if and only if the creation sequence $C_{n_0,n_1}$ contains 0011 as a subsequence.

**Proof.** First, suppose that $C_{n_0,n_1}$ contains a subsequence of the form 0011. Then there are two zero-vertices and two one-vertices such that each of the former is adjacent to each of the latter. So $D_{n_0,n_1}$ contains $K_{2,2}$ as a subgraph, and therefore has a cycle of length four.

Next, suppose that $D_{n_0,n_1}$ contains a cycle, which we denote by $X$. Then as $D_{n_0,n_1}$ is bipartite, $X$ is of length $2k$, where $k \geq 2$. Then we can decompose the vertices of $X$ into $X \cap Z$ and $X \cap U$ as follows:

$$X \cap Z = \{z_{i_1}, z_{i_2}, \ldots, z_{i_k}\} \text{ and } X \cap U = \{u_{j_1}, u_{j_2}, \ldots, u_{j_k}\},$$

where each $i_m < i_{m+1}$ and $j_m < j_{m+1}$. As every vertex in $X \cap U$ is adjacent to at least two vertices in $X \cap Z$, they must also be adjacent to both $z_{i_1}$ and $z_{i_2}$, since those are of lesser index. Similarly, every vertex in $X \cap Z$ must be adjacent to both $u_{j_{k-1}}$ and $u_{j_k}$. Thus, since both $z_{i_1}$ and $z_{i_2}$ are adjacent to both $u_{j_{k-1}}$ and $u_{j_k}$, the zeroes corresponding to the former must lie before the ones corresponding to the latter. Thus, $C_{n_0,n_1}$ contains a subsequence of the form 0011. \hfill \Box

With this result,

**Proposition 4.7.2.** For a random difference graph $D_{n_0,n_1}$,

$$P(D_{n_0,n_1} \text{ has a cycle}) = 1 - \frac{n_0n_1 + 1}{\binom{n_0+n_1}{n_0}}$$

64
Proof. In order for $D_{n_0,n_1}$ to have a cycle, there must be at least two ones that appear after the second zero in $C_{n_0,n_1}$. So if we let $j$ denote the number of ones that appear after the second zero, the total number of sequences is

$$\sum_{j=2}^{n_1} \binom{n_1-j+1}{1} \binom{n_0+j-2}{j} = \binom{n_0+n_1}{n_0} - (n_0n_1 + 1)$$

This leads to a simple calculation of the distribution of the girth of $D_{n_0,n_1}$, denoted $g(D_{n_0,n_1})$, which is the length of the shortest cycle in the graph.

**Corollary 4.7.3.** For a random difference graph $D_{n_0,n_1}$,

$$P(g(D_{n_0,n_1}) = k) = \begin{cases} 
\frac{n_0n_1 + 1}{(n_0+n_1)} & \text{if } k = 0 \\
1 - \frac{n_0n_1 + 1}{(n_0+n_1)} & \text{if } k = 4 \\
0 & \text{otherwise}
\end{cases}$$

Now we can turn to finding the length of the longest cycle.

**Proposition 4.7.4.** If $D_{n_0,n_1}$ contains a cycle of length $2k$, then it contains a $2k$-cycle of the form

$$z_1 \rightarrow u_{n_1-k+1} \rightarrow z_2 \rightarrow u_{n_1-k+2} \rightarrow z_3 \rightarrow \cdots \rightarrow z_{k-1} \rightarrow u_{n_1-1} \rightarrow z_k \rightarrow u_{n_1} \rightarrow z_1$$

Proof. First, note that by virtue of the labelling system, if there exists an edge $\{z_i, u_j\}$, then for all $i' \leq i$ and $j' \geq j$, there must also exist an edge $\{z_{i'}, u_{j'}\}$.

Now let us suppose, for the sake of contradiction, that the claim is false. So there exists some cycle $Y$ of size $2k$, but the following cycle, which we shall denote by $X$, does not exist:

$$z_1 \rightarrow u_{n_1-k+1} \rightarrow z_2 \rightarrow u_{n_1-k+2} \rightarrow z_3 \rightarrow \cdots \rightarrow z_{k-1} \rightarrow u_{n_1-1} \rightarrow z_k \rightarrow u_{n_1} \rightarrow z_1$$
Notice that within $X$, each $z_m$ is adjacent to both $u_{n_1-k+m-1}$ and $u_{n_1-k+m}$, with the exception of $z_1$.

By the existence of $Y$, there must exist $k$ one-vertices of positive degree, and therefore the right-most $k$ ones all have higher index than the left-most zero. Thus, $z_1$ is adjacent to both $u_1$ and $u_{n_1-k+1}$. Since $X$ does not exist, there must exist some minimum $m \geq 2$ such that $z_m$ is not adjacent to $u_{n_1-k+m-1}$. (If $z_m$ is not adjacent to $u_{n_1-k+m}$, then it cannot be adjacent to $u_{n_1-k+m-1}$, as the latter has lower index.)

Since $z_m$ is not adjacent to $u_{n_1-k+m-1}$, all of $z_m$’s neighbors must lie in the set \{ $u_{n_1-k+m}, \ldots, u_{n_1}$ \}, so $z_m$ has at most $(k - m + 1)$ neighbors. Furthermore, for all $r > m$, $z_r$ is also not adjacent to $u_{n_1-k+m-1}$; letting $A = \{ z_m, \ldots, z_k \}$, we see that $|N(A)| \leq k - m + 1 = |A|$. By Hall’s Marriage Theorem, since the restriction of $Y$ to $A \cup N(A)$ contains a matching, $|N(A)| \geq |A|$.

Because $|A| = |N(A)|$, any cycle containing $A$ must have length exactly $2|A| < 2k$, which contradicts the existence of $Y$. Thus, we see that $X$ must exist.

We now have the machinery required to look at $\psi(D_{n_0,n_1})$, the length of the longest cycle in $D_{n_0,n_1}$.

**Proposition 4.7.5.** If $D_{n_0,n_1}$ is a difference graph, then for all $k \geq 2$, $\psi(D_{n_0,n_1}) = 2k$ if and only if $\nu(D(int(C_{n_0,n_1}))) = k - 1$.

**Proof.** First, suppose that the longest cycle is of length $2k$. Then by Proposition 4.7.4, the following cycle, which we denote $Y$, must exist:

$$z_1 \rightarrow u_{n_1-k+1} \rightarrow z_2 \rightarrow u_{n_1-k+2} \rightarrow z_3 \rightarrow \cdots \rightarrow z_{k-1} \rightarrow u_{n_1-1} \rightarrow z_k \rightarrow u_{n_1} \rightarrow z_1$$

As all of the above vertices are non-isolated, their corresponding digits all lie in $int(C_{n_0,n_1})$, with the exception of $z_1$ and $u_{n_1}$. Then for $2 \leq m \leq k$, edges of the form \{ $z_m, u_{n_1-k+m-1}$ \} comprise a matching of size $k - 1$, and therefore $\nu(D(int(C_{n_0,n_1}))) \geq k - 1$. 

66
Second, to show the converse, suppose that \( \nu(D(int(C_{n_0,n_1}))) = k - 1 \). Then there must exist edges in \( D_{n_0,n_1} \) of the form

\[
\{ z_{\sigma(1)}, u_{\tau(1)} \}, \{ z_{\sigma(2)}, u_{\tau(2)} \}, \ldots, \{ z_{\sigma(k-1)}, u_{\tau(k-1)} \},
\]

where both \( \sigma \) and \( \tau \) are injective, and for all \( i \), \( \sigma(i) \geq 2 \) and \( \tau(i) \leq n_1 - 1 \). Furthermore, there must exist edges

\[
\{ z_2, u_{\pi(2)} \}, \{ z_3, u_{\pi(3)} \}, \ldots, \{ z_k, u_{\pi(k)} \},
\]

as the existence of edge \( \{ z_{i+1}, u_j \} \) implies the existence of edge \( \{ z_i, u_j \} \). (The function \( \pi \) obeys the same constraints as \( \tau \).) And therefore, the following cycle exists in \( D_{n_0,n_1} \):

\[
z_1 \to u_{\pi(2)} \to z_2 \to u_{\pi(3)} \to z_3 \to \cdots \to u_{\pi(k-1)} \to z_{k-1} \to u_{\pi(k)} \to z_k \to u_{n_1} \to z_1
\]

Thus, \( \psi(D_{n_0,n_1}) \geq 2k \).

To better illustrate this construction, consider Figure 4.1, which demonstrates the construction for the difference graph with creation sequence \( C = 001001011 \). Then \( int(C) = 0100101 \), which has a matching number of three:

![Figure 4.1: Difference Graph Cycles, Step 1](image-url)

By the above arguments, however, we know that there must exist some maximum matching on the interior that is of the form

\[
\{ z_2, u_{\pi(1)} \}, \{ z_3, u_{\pi(2)} \}, \{ z_4, u_{\pi(3)} \}, \text{ where } \pi(1) < \pi(2) < \pi(3) < 4
\]

67
Such a matching is given in Figure 4.2:

Of course, by properties of difference graphs, there must also exist several other edges, which are shown as dotted lines in Figure 4.3. Then by taking the edge between $z_1$ and $u_4$, we have a cycle with length $8 \ (= 2 \times 3 + 2)$.

Since both $\psi$ and $\nu$ are, by definition, functions that return maxima, the composition of the above results proves that $\psi(D_{n_0,n_1}) = 2k$ if and only if $\nu(D(int(C_{n_0,n_1}))) = k - 1$. 

Having shown that $\psi(D_{n_0,n_1})$ depends solely on $int(C_{n_0,n_1})$, we take a deeper look at the latter. Recalling that there are two distinct creation sequences with empty interior, and that every possible non-empty sequence occurs exactly once as the interior of some $C_{n_0,n_1}$,
Theorem 4.7.6. If $D_{n_0,n_1}$ is a difference graph and $2 \leq k$, then $P(\psi(D_{n_0,n_1}) = k)$ is given by

$$\binom{n_0+n_1}{n_0}^{-1} \left[ \binom{n_0+n_1}{k+1} + \binom{n_0+k-1}{k} + \binom{n_1+k-1}{k} + \binom{2k-2}{k+1} 
- \binom{n_0+n_1}{k} - \binom{n_0+k-1}{k+1} - \binom{n_1+k-1}{k+1} - \binom{2k-2}{k} \right]$$

Proof. Note that

$$P(\psi(D_{n_0,n_1}) = 2k) = \frac{1}{\binom{n_0+n_1}{n_0}} \sum_{D_{n_0,n_1} \in D_{n_0,n_1}} 1_{\{\psi(D_{n_0,n_1}) = 2k\}}$$

$$= \binom{n_0+n_1}{n_0}^{-1} \sum_{C_{n_0,n_1} \in \mathbb{C}_{n_0,n_1}} 1_{\{\psi(D(C_{n_0,n_1})) = 2k\}}$$

$$= \binom{n_0+n_1}{n_0}^{-1} \sum_{C_{n_0,n_1} \in \mathbb{C}_{n_0,n_1}} 1_{\{\nu(D(int(C_{n_0,n_1}))) = k-1\}}$$

$$= \binom{n_0+n_1}{n_0}^{-1} \sum_{i=0}^{n_0-1} \sum_{j=0}^{n_1-1} \sum_{c_{i,j} \in \mathbb{C}_{i,j}} 1_{\{\nu(D(c_{i,j})) = k-1\}}$$

$$= \binom{n_0+n_1}{n_0}^{-1} \sum_{i=0}^{n_0-1} \sum_{j=0}^{n_1-1} \binom{i+j}{i} P(\nu(D_{i,j}) = k-1)$$

Using Proposition 4.5.2, this can be further simplified to

$$P(\psi(D_{n_0,n_1}) = 2k) = \binom{n_0+n_1}{n_0}^{-1} \sum_{i=0}^{n_0-1} \sum_{j=0}^{n_1-1} \binom{i+j}{k-1} - \binom{i+j}{k-2} \right)$$

By making extensive use of the identity

$$\sum_{m=k-1}^{n-1} \binom{m+a}{b} = \binom{a+n}{b+1} - \binom{a+k-1}{b+1},$$

$P(\psi(D_{n_0,n_1}) = 2k)$ becomes

$$\binom{n_0+n_1}{n_0}^{-1} \sum_{i=k-1}^{n_0-1} \left[ \binom{i+n_1}{k} - \binom{i+k-1}{k} - \binom{i+n_1}{k-1} + \binom{i+k-1}{k-1} \right],$$

which simplifies to

$$\binom{n_0+n_1}{n_0}^{-1} \left[ \binom{n_0+n_1}{k+1} + \binom{n_0+k-1}{k} + \binom{n_1+k-1}{k} + \binom{2k-2}{k+1} 
- \binom{n_0+n_1}{k} - \binom{n_0+k-1}{k+1} - \binom{n_1+k-1}{k+1} - \binom{2k-2}{k} \right]$$

69
Corollary 4.7.7. If $D_{n,n}$ is a difference graph with parts of equal size and $n \geq 2$, then

$$P(D_{n,n} \text{ is Hamiltonian}) = \frac{1}{4n - 2}$$

4.8 Planarity

Proposition 4.8.1. A difference graph $D_{n_0,n_1}$ is planar if and only if its creation sequence $C_{n_0,n_1}$ does not contain a subsequence of the form 000111.

Proof. By Kuratowski’s theorem, $D_{n_0,n_1}$ is planar if and only if it does not contain a subdivision of $K_5$ or $K_{3,3}$.

First, suppose that $C_{n_0,n_1}$ contains 000111 as a subsequence. Then as each of those three zeroes lies to the left of those three ones in $C_{n_0,n_1}$, their corresponding vertices are adjacent in $D_{n_0,n_1}$. So there exist three zero-vertices and three one-vertices in $D_{n_0,n_1}$ such that each of the former is adjacent to each of the latter, forming a complete bipartite graph on parts of size three and three. Thus $D_{n_0,n_1}$ contains a subgraph isomorphic to $K_{3,3}$, and is non-planar.

Second, suppose that $D_{n_0,n_1}$ is non-planar and contains a subdivision of $K_5$. Let $A$ denote the set of vertices that correspond to the vertices of $K_5$; since all of those vertices have degree exceeding two, they could not have been introduced via subdivision, and must correspond to five original vertices of $D_{n_0,n_1}$. As $D_{n_0,n_1}$ is bipartite, we know by the pigeonhole principle that one of the parts must contain at least three elements of $A$; without loss of generality, we can assume that at least three members of $A$ are vertices in $Z$. Let us denote these vertices by $z_{i_1}$, $z_{i_2}$, and $z_{i_3}$, where $i_1 < i_2 < i_3$.

As the subdivision process does not increase the degree of any (existing) vertices,
we know that \( \deg_{D_{n_0,n_1}}(z_{i_3}) \geq 4 \). Thus, there exist at least four elements in \( U \) that are adjacent (in \( D_{n_0,n_1} \)) to \( z_{i_3} \). But as both \( z_{i_1} \) and \( z_{i_2} \) are of lower index than \( z_{i_3} \), they are adjacent to all neighbors of \( z_{i_3} \). Thus we have three zero-vertices, each of which is adjacent to each of four one-vertices. So \( C_{n_0,n_1} \) contains a subsequence of the form 0001111, and therefore one of the form 0001111.

Third, suppose that \( D_{n_0,n_1} \) is non-planar and contains a subdivision of \( K_{3,3} \). The same argument as above holds, although the result is slightly weaker; \( C_{n_0,n_1} \) does not necessarily contain 0001111, but it does contain some 0001111.

**Proposition 4.8.2.** For a random difference graph \( D_{n_0,n_1} \) with \( n_0, n_1 \geq 3 \), the probability of \( D_{n_0,n_1} \) being planar is

\[
\left( \frac{n_0 + n_1}{n_0} \right)^{-1} \left( \frac{4 + n_0n_1(n_0n_1 - n_0 - n_1 + 5)}{4} \right)
\]

**Proof.** By the above, we can exclude those cases where \( n_0 \) or \( n_1 \) is less than three, as those cases will always produce a planar graph. Let \( j \) denote the number of ones in \( C_{n_0,n_1} \) of higher index than that of \( z_3 \). In order for \( D_{n_0,n_1} \) to be planar, \( j \) can be no larger than two.

Then the initial segment of \( C_{n_0,n_1} \), from the beginning up to the third zero, has exactly three zeroes and \( n_1 - j \) ones. The rest of the sequence has exactly \( n_0 - 3 \) zeroes and \( j \) ones. Counting the number of such sequences,

\[
P(D_{n_0,n_1} \text{ is planar}) = \left( \frac{n_0 + n_1}{n_0} \right)^{-1} \sum_{j=0}^{2} \binom{n_1 - j + 2}{2} \binom{n_0 - 3 + j}{j}
\]

\[
= \left( \frac{n_0 + n_1}{n_0} \right)^{-1} \left( \binom{n_1 + 2}{2} + \binom{n_1 + 1}{2} \binom{n_0 - 2}{j} + \binom{n_1}{2} \binom{n_0 - 1}{2} \right)
\]

\[
= \left( \frac{n_0 + n_1}{n_0} \right)^{-1} \left( 1 + \frac{n_0n_1(n_0n_1 - n_0 - n_1 + 5)}{4} \right)
\]

71
Thus, the probability of planarity is

\[
\left( \frac{n_0 + n_1}{n_0} \right)^{-1} \left( \frac{4 + n_0 n_1 (n_0 n_1 - n_0 - n_1 + 5)}{4} \right)
\]

\[\square\]

4.9 \textit{k-Core}

A \textit{k-core} of a graph \(G\) is the maximal induced subgraph \(H \subseteq G\) such that all vertices of \(H\) have degree at least \(k\), formed by iteratively deleting all vertices with degree less than \(k\). The degeneracy of a graph \(G\) is the largest \(k\) such that the \(k\)-core of \(G\) is non-empty.

An equivalent formulation for degeneracy of \(G\) is the maximum, over all induced subgraphs \(H \subseteq G\), of the minimum degree of a vertex in \(H\). That is,

\[
degen(G) = \max_{H \subseteq G} \min_{v \in V(H)} \deg(v)
\]

**Proposition 4.9.1.** For a difference graph \(D_{n_0,n_1}\), \(degen(D_{n_0,n_1}) \geq d\) if and only if \(K_{d,d} \subseteq D_{n_0,n_1}\).

**Proof.** First, suppose that \(K_{d,d}\) is an induced subgraph of \(D_{n_0,n_1}\). Then there exists a subgraph of \(D_{n_0,n_1}\) in which all degrees are exactly \(d\), and therefore the degeneracy is at least \(d\).

Next, suppose that the degeneracy of \(D_{n_0,n_1}\) is at least \(d\). Then there exists some subgraph \(H \subseteq D_{n_0,n_1}\) such that the degree of every vertex in \(H\) is at least \(d\). As \(D_{n_0,n_1}\) is bipartite, \(H\) must contain at least \(d\) vertices from each part. Therefore \(D_{n_0,n_1}\) must contain \(d\) zero-vertices and \(d\) one-vertices of degree at least \(d\).

Denoting the aforementioned vertices as \(\{z_1, \ldots, z_d\}\) and \(\{u_{j_1}, \ldots, u_{j_d}\}\), respectively, recall that for \(s < t\), \(z_s\) is adjacent to all neighbors of \(z_t\), and \(u_t\) is adjacent to all neighbors of \(u_s\). Thus, every vertex in the sets \(\{z_1, \ldots, z_d\}\) and \(\{u_{n_1-d+1}, \ldots, u_{n_1}\}\)
has degree at least \( d \), and moreover, every vertex in each set is adjacent to all vertices in the other. So \( K_{d,d} \subseteq D_{n_0,n_1} \).

**Proposition 4.9.2.** For a random \( D_{n_0,n_1} \),

\[
P(\text{degen}(D_{n_0,n_1}) = d) = \binom{n_0 + n_1}{n_0}^{-1} \binom{n_0}{d} \binom{n_1}{d}
\]

**Proof.** From the above, we see that

\[
\text{degen}(D_{n_0,n_1}) = \max \{ d : K_{d,d} \subseteq D_{n_0,n_1} \}
\]

Now, \( D_{n_0,n_1} \) contains \( K_{d,d} \) if and only if the creation sequence \( C_{n_0,n_1} \) contains \( 0^d1^d \) as a subsequence. So in order for \( \text{degen}(D_{n_0,n_1}) \) to equal \( d \), \( C_{n_0,n_1} \) must contain a substring of the form \( 0^d1^d \), but cannot contain \( 0^d1^{d+1} \).

The requirement that \( C_{n_0,n_1} \) contain \( 0^d1^d \) means that there must be at least \( d \) ones after the \( d \)th zero; equivalently, there must be at most \( n_1 - d \) ones before the \( d \)th zero. The requirement that \( C_{n_0,n_1} \) not contain \( 0^{d+1}1^{d+1} \) can be restated as requiring that there be at most \( d \) ones after the \((d+1)\)st zero, or that there be at least \( n_1 - d \) ones before the \((d+1)\)st zero.

Letting \( i \) equal the number of ones before the \( d \)th zero, we have the range \( 0 \leq i \leq n_1 - d \). Letting \( j \) denote the number of ones between \( z_d \) and \( z_{d+1} \), the requirement \( i + j \geq n_1 - d \) becomes \( n_1 - d - i \leq j \leq n_1 - i \). Thus,

\[
P(\text{degen}(D_{n_0,n_1}) = d) = \binom{n_0 + n_1}{n_0}^{-1} \sum_{i=0}^{n_1-d} \sum_{j=n_1-d-i}^{n_1-i} \binom{i + d - 1}{i} \binom{n_1 - i - j + n_0 - d - 1}{n_1 - i - j}
\]

which simplifies to

\[
P(\text{degen}(D_{n_0,n_1}) = d) = \binom{n_0 + n_1}{n_0}^{-1} \sum_{i=0}^{n_1-d} \binom{i + d - 1}{i} \binom{n_0}{d}
= \binom{n_0 + n_1}{n_0}^{-1} \binom{n_0}{d} \binom{n_1}{d}
\]
CHAPTER 5
ASYMPTOTICS

5.1 Introduction

In chapters two and four, we found the exact distributions for many graph properties, expressed in terms of \( n, n_0, \) and \( n_1. \) For many purposes, these are ideal, despite any potential unwieldiness (e.g., Theorem 4.7.6). However, it can be equally important to look beyond the minutiae, to know the limiting distributions as these parameters approach infinity. So as not to disrupt the combinatorial nature of the previous chapters, we include those asymptotics here.

With respect to the difference graph \( D_{n_0,n_1}, \) rather than allowing \( n_0 \) and \( n_1 \) to grow independently, we fix a constant ratio \( a = \frac{n_0}{n_1}. \) This greatly simplifies the process, and we can now consider \( D_{an,n} \) as \( n \) approaches infinity. (Note that by symmetry, \( D_{an,n} \) and \( D_{n,an} \) have the same properties, so we can assume that \( a \geq 1. \))

Some of these asymptotics are based upon cumulative distribution functions, whereas others stem from density functions, depending upon which form would best capture the limiting behavior. In the latter case, please note that all of the random variables under consideration assume only integer values, and hence the probability of any of them having a non-integral value is exactly zero.
5.2 Threshold Graphs: Matching Number

We begin with the matching number $\nu(G_n)$, whose distribution was given in Theorem 2.4.6 as:

$$P(\nu(G_n) = k) = \begin{cases} \left(\frac{1}{2}\right)^{n-1} \binom{n}{k} & k < \frac{n}{2} \\ \left(\frac{1}{2}\right)^{n-1} \binom{n-1}{\left[\frac{n-1}{2}\right]} & k = \frac{n}{2} \end{cases}$$

Having found the exact distribution of $\nu(G_n)$, we are now in position to discuss the asymptotic behavior. First, we consider the expected value:

**Proposition 5.2.1.** For a random threshold graph $G_n$,

$$E[\nu(G_n)] = \frac{n}{2} - \frac{\sqrt{n}}{\sqrt{2\pi}} + O\left(\frac{1}{\sqrt{n}}\right)$$

**Proof.** Due to the piecewise nature of $\nu$’s distribution, we decompose into two cases, depending upon the parity of $n$.

For odd values of $n = 2m + 1$,

$$E[\nu(G_n)] = \sum_{k=0}^{\left\lfloor \frac{n}{2} \right\rfloor} k \left(\frac{1}{2}\right)^{n-1} \binom{n}{k} = \left(\frac{1}{2}\right)^{n-1} \sum_{k=1}^{m} n \binom{n-1}{k-1} = \frac{n}{2^{m-1}} \sum_{j=0}^{m-1} \binom{2m-1}{j} = \frac{n}{2^{2m}} \left(2^{2m-1} - \frac{1}{2} \binom{2m}{m}\right) = \frac{n}{2} - \frac{n}{2^{2m+1}} \binom{2m}{m}$$

Applying Stirling’s approximation, we see that

$$\frac{n}{2^{2m+1}} \binom{2m}{m} = \sqrt{\frac{n}{2\pi}} + O\left(\frac{1}{\sqrt{n}}\right)$$

Similar results hold for even values of $n = 2m$. \qed
Theorem 5.2.2. For a random threshold graph $G_n$,

$$P\left(\frac{n}{2} - \nu(G_n) = x\right) \sim \sqrt{\frac{8}{\pi n}} e^{-2x^2},$$

where $0 < x = o(\sqrt{n})$.

**Proof.** Noting that the expected value of $\nu(G_n)$ is very close to its maximum value, we look at the size of its fluctuations around that point. Using Stirling, the probability that $\nu(G_n)$ is equal to $n/2 - x\sqrt{n}$ is asymptotic to

$$\sqrt{\frac{8}{\pi n}} \left(1 - \frac{2x}{\sqrt{n}}\right)^{x\sqrt{n}} \left(1 + \frac{2x}{\sqrt{n}}\right)^{-x\sqrt{n}}$$

For $x = o(\sqrt{n})$, we note that as $n \to \infty$,

$$\left(1 - \frac{2x}{\sqrt{n}}\right)^{x\sqrt{n}} \left(1 + \frac{2x}{\sqrt{n}}\right)^{-x\sqrt{n}} \to e^{-2x^2}$$

and similarly,

$$\left(1 - \frac{2x}{\sqrt{n}}\right)^{n/2} \left(1 + \frac{2x}{\sqrt{n}}\right)^{n/2} \to 1$$

Thus,

$$P\left(\nu(G_n) = \frac{n}{2} - x\sqrt{n}\right) \sim \sqrt{\frac{8}{\pi n}} e^{-2x^2}$$

5.3 Threshold Graphs: Longest Cycle Length

We now turn to the length of the longest cycle, $\psi(G_n)$:

**Proposition 5.3.1.** For a random threshold graph $G_n$,

$$E[\psi(G_n)] = n - \sqrt{\frac{2n}{\pi}} + O(1)$$
Proof.

\[ E[\psi(G_n)] = \left( \frac{1}{2} \right)^{n-1} \sum_{k=3}^{n} k \left[ \left( \frac{n-1}{k} \right) - \left( \frac{k-2}{k} \right) \right] \]

Breaking apart the sum, we consider each component individually. Supposing odd values of \( n = 2m + 1 \),

\[ \sum_{k=3}^{n} k \left( \frac{n-1}{k} \right) = 3 \binom{n-1}{1} + \sum_{k=4}^{2m+1} k \left( \frac{2m}{k} \right) \]

\[ = 3(n-1) + \sum_{j=2}^{m} ((2j + 2j + 1)) \left( \frac{2m}{j} \right) \]

\[ = 3(2m) + 4 \sum_{j=2}^{m} j \left( \frac{2m}{j} \right) + \sum_{j=2}^{m} \left( \frac{2m}{j} \right) \]

\[ = (4m + 1)(2^{2m-1} - 1) + \frac{1}{2} \binom{2m}{m} \]

\[ = 2^{2m+1}m + O(2^{2m}) \]

\[ = 2^{n-1}n + O(2^n) \]

Similarly,

\[ \sum_{k=3}^{n} k \left( \frac{k-2}{k} \right) = 3 + \sum_{k=4}^{2m+1} k \left( \frac{k-2}{k} \right) \]

\[ = 3 + \sum_{j=2}^{m} \left( 2j \left( \frac{2m-2}{j} \right) + (2j + 1) \left( \frac{2j-1}{j} \right) \right) \]

\[ \rightarrow \left( 2m \left( \frac{2m-2}{m} \right) + (2m + 1) \left( \frac{2m-1}{m} \right) \right) \left( \frac{1}{1-\frac{1}{4}} \right) \]

\[ = \left( \frac{3}{2} \sqrt{\frac{m}{\pi}} 2^{2m} + O(2^{2m}) \right) \left( \frac{4}{3} \right) \]

\[ = \sqrt{\frac{m}{\pi}} 2^{2m+1} + O(2^{2m}) \]

\[ = \sqrt{\frac{n}{2\pi}} 2^n + O(2^n) \]

Thus,

\[ E[\psi(G_n)] = 2^{1-n} \left( 2^{n-1}n - \sqrt{\frac{n}{2\pi}} 2^n + O(2^n) \right) = n - \sqrt{\frac{2n}{\pi}} + O(1) \]
As for the limiting behavior, rather than taking the direct approach, we instead examine the relationship between \( \nu(G) \) and \( \psi(G) \).

**Proposition 5.3.2.** For any threshold graph \( G \),

\[
2 \nu(G) - 1 \leq \psi(G) \leq 2 \nu(G) + 1
\]

*Proof.* First, note that given a cycle of length \( \psi(G) \), one can form a matching simply by removing every other edge. (In the case when \( \psi(G) \) is odd, one additional edge must also be removed.) In this fashion,

\[
\nu(G) \geq \left\lfloor \frac{\psi(G)}{2} \right\rfloor, \text{ so } 2 \nu(G) \geq \psi(G) - 1
\]

For the lower bound, let \( C \) denote the creation sequence of \( G \), and let \( M \) be a maximum matching of \( G \); that is, \( |V(M)| = 2\nu(G) \) and no two edges of \( M \) share a vertex. (So for \( \nu(G) \geq 2 \), \( M \) is not a threshold graph.) Let \( H \) be the induced subgraph of \( G \) by the vertices \( V(M) \), and let \( C_H \) be its creation sequence.

Since \( H \) has a perfect matching, we know by Proposition 2.4.2 that \( h(C_H) = 0 \). Let us define subsequence \( C'_H \) of \( C_H \) by removing the right-most zero from \( C_H \), if such exists, or by taking the entire \( C_H \) otherwise. Then every tail of \( C'_H \) must have strictly more ones than zeroes, and therefore the corresponding threshold graph, which is a graph on at least \( 2\nu(G) - 1 \) vertices, must contain a Hamilton cycle. Since this graph is an induced subgraph of \( H \) and thus of \( G \), we see that

\[
2 \nu(G) - 1 \leq \psi(G)
\]

\[\square\]

**Corollary 5.3.3.** For a random threshold graph \( G_n \),

\[
P\left(\frac{n - \psi(G_n)}{\sqrt{n}} = x\right) \sim \sqrt{\frac{8}{\pi n}} e^{-\frac{x^2}{4}},
\]

where \( 0 < x = o(\sqrt{n}) \).
5.4 Threshold Graphs: Planarity

**Proposition 5.4.1.** For a random threshold graph $G_n$,

$$P(G_n \text{ is planar}) = \frac{3n^2}{2^n} \left(1 + O\left(\frac{1}{n}\right)\right)$$

*Proof.* Recall from Theorem 2.6.3 that

$$P(G_n \text{ is planar}) = \frac{3n^2 - 13n + 20}{2^n}$$

From here, the nearly exponentially-small asymptotic probability of planarity follows with minimal effort. \(\square\)

5.5 Threshold Graphs: Degeneracy

**Proposition 5.5.1.** For a random threshold graph $G_n$,

$$P\left(\frac{\text{degen}(G_n) - \frac{n-1}{2}}{\sqrt{n}} = x\right) \sim \sqrt{\frac{2}{\pi n}} e^{-2x^2},$$

where $x = o(\sqrt{n})$.

*Proof.* Notice that, by Theorem 2.7.3, the degeneracy of a random threshold graph $G_n$ follows the binomial distribution exactly:

$$P(\text{degen}(G_n) = d) = \left(\frac{1}{2}\right)^{n-1} \binom{n-1}{d}$$

So by applying the same techniques used in Theorem 5.2.2, we see that

$$P\left(\text{degen}(G_n) = \frac{n-1}{2} + x\sqrt{n}\right) \sim \sqrt{\frac{2}{\pi n}} e^{-2x^2}$$

\(\square\)
5.6 Threshold Graphs: $k$-Core

This section becomes slightly more complicated than the previous, due to the additional parameter $k$. However, we keep $k$ fixed, and consider the asymptotics as $n$ approaches infinity.

**Proposition 5.6.1.** For a random threshold graph $G_n$ and a fixed $k$,

$$E[|k\text{-core}(G_n)|] = n - k + o(1)$$

as $n \to \infty$.

**Proof.** Recall, from Theorem 2.7.5, that the probability of the $k$-core being empty is

$$\sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^{n-1} \binom{n-1}{i}$$

Working crudely, we implement the following bound,

$$\sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^{n-1} \binom{n-1}{i} \leq \sum_{i=0}^{k-1} \frac{(n-1)^i}{2^{n-1}} = O\left(\frac{n^k}{2^n}\right),$$

and see that for any fixed value of $k$, a non-empty $k$-core exists with probability approaching 1.

We now express the size of the $k$-core as $n$ minus the number of “pruned” vertices. That is, given a threshold graph $G$ with $n$ vertices, after one iteratively removes all of the vertices with degree less than $k$, the result is the $k$-core. As such, by linearity of expectation,

$$E[|k\text{-core}|] = n - E[\#\text{pruned}]$$

To find the distribution of the number of pruned vertices, we return again to Theorem 2.7.5:

$$P(\#\text{pruned} = i) = P(|k\text{-core}(G_n)| = n-i) = \begin{cases} \left(\frac{1}{2}\right)^{k+i} \binom{k+i-1}{k-1} & i \leq n - k - 1 \\ \sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^{n-1} \binom{n-1}{i} & i = n \end{cases}$$
Above, we showed that the maximum contribution from the lower case to the expected value would be $O\left(\frac{n^{k+1}}{2^n}\right)$, which again approaches zero as $n$ approaches infinity. Thus,

$$E[|k\text{-core}(G_n)|] = n - \sum_{i=0}^{n-k-1} \left(\frac{1}{2}\right)^{k+i} \binom{k+i-1}{k-1} i + O\left(\frac{n^{k+1}}{2^n}\right)$$

Now, as $n$ approaches infinity,

$$\sum_{i=0}^{n-k-1} \left(\frac{1}{2}\right)^{k+i} \binom{k+i-1}{k-1} i \to \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{k+i} \binom{k+i-1}{k-1} i = k \left(\frac{1}{2}\right)^{k+1} 2^{k+1} = k$$

with an error within that given above. As such, we see that $E[|k\text{-core}(G_n)|] = n - k + o(1)$.

\[\square\]

**Proposition 5.6.2.** For a random threshold graph $G_n$ and fixed $k$,

$$n - |k\text{-core}(G_n)| \Rightarrow NB\left(k, \frac{1}{2}\right)$$

where $NB(k,1/2)$ is the negative binomial distribution, as $n \to \infty$.

**Proof.** Recall, from our work in Proposition 5.6.1, that for $k + 1 \leq n - i$,

$$P(|k\text{-core}| = n - i) = \left(\frac{1}{2}\right)^{k+i} \binom{k+i-1}{k-1}$$

That is, in the event that a non-empty $k$-core exists, the number of “pruned” zero-vertices obeys the negative binomial distribution. This works on an intuitive level, as one way to interpret the size of the $k$-core is by reading the creation sequence from right to left: the number of zero-vertices with degree less than $k$ is given by the number of zeroes before encountering the $k$-th “1” from the right.
As $k$ is fixed, the probability of the $k$-core being empty approaches zero as $n$ approaches infinity. So we see that $n - |k\text{-core}(G_n)|$ converges in distribution to $\text{NB}\left(k, \frac{1}{2}\right)$.

5.7 Threshold Graphs: $t$-Stability

**Proposition 5.7.1.** For a random threshold graph $G_n$,

$$P\left(\alpha_t(G_n) = k\right) = \left(\frac{1}{2}\right)^{n-t-1} \binom{n-t}{k-t-1},$$

where $t$ is fixed and $x = o(\sqrt{n})$.

**Proof.** While the distribution of $\alpha_t(G_n)$ was computed in Corollary 2.8.3 as

$$P(\alpha_t(G_n) = k) = \left(\frac{1}{2}\right)^{n-t-1} \binom{n-t}{k-t-1},$$

notice that the distribution of the number of vertices not present in a maximum $t$-stable set is more familiar:

$$P(n - \alpha_t(G_n) = k) = \left(\frac{1}{2}\right)^{n-t-1} \binom{n-t}{k}$$

As the latter distribution is exactly that of a binomial, we omit the now-familiar calculations showing that

$$P\left(n - \alpha_t(G_n) = \frac{n-t-1}{2} + x\sqrt{n}\right) = \sqrt{\frac{2}{\pi(n-t-1)}} e^{-2x^2}$$

5.8 Difference Graphs: Connectivity

**Proposition 5.8.1.** For a random difference graph $D_{an,n}$,

$$P(\kappa(D_{an,n}) \geq x) \sim \left(\frac{a}{(a+1)^2}\right)^x$$

where $0 < x = o(\sqrt{n})$.  

82
Proof. Recall from Corollary 4.3.5 that

\[ P(\kappa(D_{an,n}) = k) = \binom{an+n}{n}^{-1} \left( \binom{an+n-2k}{n-k} - \binom{an+n-2(k+1)}{n-(k+1)} \right) \]

To best exploit the obvious telescoping, we consider not the usual form of the cumulative distribution function, but its opposite:

\[ P(\kappa(D_{an,n} \geq k) = P(D_{an,n} \text{ is } k\text{-conn.}) = \binom{an+n}{n}^{-1} (an+n-2k) \]

Applying Stirling, we see that this probability can be rewritten, for \( k = x \), as

\[ \sqrt{\frac{a(a+1-\frac{2x}{n})}{(a+1)(1-\frac{x}{n})(a-\frac{1}{n})}} \left( \frac{a^a}{(a+1)^{a+1}} \right)^n \frac{(a+1-\frac{2x}{n})^{an+n-2x}}{(1-\frac{x}{n})^{n-x}(a-\frac{x}{n})^{an-x}} \]

And for \( x = o(\sqrt{n}) \), the above is asymptotic to

\[ \left( \frac{a}{(a+1)^2} \right)^x \]

suggesting a sharp concentration about 0.

\[ \square \]

Corollary 5.8.2. For a random difference graph \( D_{an,n} \),

\[ P(\lambda(D_{an,n}) \geq x) \sim \left( \frac{a}{(a+1)^2} \right)^x \]

where \( 0 < x = o(\sqrt{n}) \).

Proof. By Proposition 4.3.8, \( \lambda(D_{an,n}) \) has the same distribution as \( \kappa(D_{an,n}) \).

\[ \square \]

5.9 Difference Graphs: Extreme Degrees

Proposition 5.9.1. For a random difference graph \( D_{an,n} \),

\[ P(an - \Delta(D_{an,n}) = x) \sim c \left( \frac{a^a}{(a+1)^{a+1}} \right)^n \]

where \( 0 < x = o(n) \) and

\[ c = \begin{cases} 3 & \text{if } a = 1 \\ 4 & \text{if } a > 1 \\ \frac{a}{a+1} & \text{if } a > 1 \end{cases} \]
Proof. While we never explicitly computed the cumulative distribution function for the maximum degree, recall that $\Delta(D_{an,n}) \leq k$ if and only if $C_{an,n}$ starts with $\max(0, n - k)$ consecutive ones and ends with $\max(0, an - k)$ consecutive zeroes.

Recall from Theorem 4.4.1 that the probability that $\Delta(D_{an,n}) = k$ equals

$$P(\Delta(D_{an,n}) = k) = \begin{cases} \binom{an + n}{n}^{-1} & k = 0 \\ \binom{an + n}{n}^{-1} \left( \binom{2k - 1}{k} - \binom{2k - 2}{k - 1} \right) & 1 \leq k \leq n \\ \binom{an + n}{n}^{-1} \left( \binom{n + k - 1}{k} \right) & n < k \leq an \end{cases}$$

We consider the case $a = 1$ first, where for $1 \leq k \leq n$,

$$P(\Delta(D_{n,n}) = k) = \binom{2n}{n}^{-1} \left( \binom{2k - 1}{k - 1} - \binom{2k - 2}{k - 1} \right) = \binom{2n}{n}^{-1} \left( \frac{2k}{k} \frac{3k - 2}{4k - 2} \right)$$

Then for $0 < x = o(n)$,

$$P(\Delta(D_{n,n}) = n - x) \sim \frac{3}{4} \left( \frac{1}{4} \right)^x$$

Now we consider the case where $a > 1$. If we write $k$ as $an - x$, for $0 < x = o(n)$, then we use the third component of the piecewise distribution,

$$P(\Delta(D_{n,n}) = k) = \binom{an + n}{n}^{-1} \left( \binom{n + k}{n} \frac{n}{n + k} \right) \sim \frac{1}{a + 1} \left( \frac{a^x}{(a + 1)^{a+1}} \right)$$

So in both cases, the maximum degree is sharply concentrated about the size of the largest part. \hfill \square

**Proposition 5.9.2.** For a random difference graph $D_{an,n}$,

$$P(\delta(D_{an,n}) \geq x) \sim \left( \frac{a}{(a + 1)^2} \right)^x$$

where $0 < x = o(\sqrt{n})$.

Proof. By Proposition 4.3.8, $\delta(D_{an,n})$ has the same distribution as $\kappa(D_{an,n})$. \hfill \square
5.10 Difference Graphs: Matching Number

As a reminder, when considering asymptotics for $D_{an,n}$, we assume by symmetry that $a \geq 1$.

Let us first recall from Theorem 4.5.2 that

$$P(\nu(D_{an,n}) = j) = \left(\frac{an + n}{n}\right)^{-1} \left(\binom{an + n}{j} - \binom{an + n}{j - 1}\right)$$

Proposition 5.10.1. For a random difference graph $D_{an,n}$,

$$E[\nu(D_{an,n})] = n - O(\sqrt{n})$$

Proof. We begin by using the formula

$$E[\nu(D_{an,n})] = n - E[h(C_{an,n})],$$

which follows from Proposition 4.5.1.

Continuing our assumption that $a \geq 1$, we see that

$$E[h(C_{an,n})] = \sum_{k \geq 1} P(h(C_{an,n}) \geq k)$$

$$= \sum_{k=1}^{n} \left(\frac{an + n}{n}\right)^{-1} \left(\frac{an + n}{n + k}\right)$$

$$= \left(\frac{an + n}{n}\right)^{-1} \sum_{j=0}^{n-1} \left(\frac{an + n}{j}\right)$$

For all $a > 1$, the ratio between consecutive summands is both greater than and bounded away from 1, and therefore

$$\sum_{j=0}^{n-1} \left(\frac{an + n}{j}\right) = O\left(\left(\frac{an + n}{n} \right)\right),$$

so $E[h(C_{an,n})] = O(1)$.  

85
On the other hand, if $a = 1$,

$$E[h(C_{an,n})] = \binom{2n}{n}^{-1} \sum_{j=0}^{n-1} \binom{2n}{j}$$

$$= \binom{2n}{n}^{-1} \left( \frac{1}{2} \right) \left( 2^{2n} - \binom{2n}{n} \right)$$

$$= O(\sqrt{n})$$

Thus, $E[h(C_{an,n})] = O(\sqrt{n})$, and therefore

$$E[\nu(D_{an,n})] = n + O(\sqrt{n})$$

Proposition 5.10.2. For a random difference graph $D_{n,n}$,

$$P \left( \frac{n - \nu(D_{n,n})}{\sqrt{n}} \geq x \right) \sim e^{-x^2},$$

and for a random difference graph $D_{an,n}$ with $a > 1$,

$$P \left( n - \nu(D_{an,n}) \geq x \right) \sim a^{-x},$$

where $0 \leq x = o(\sqrt{n})$.

Proof. Consider the cumulative distribution function for $\nu(D_{an,n})$. Applying Stirling,

$$P(\nu(D_{an,n}) \leq k) = \binom{an + n}{n}^{-1} \binom{an + n}{k}$$

$$\sim \sqrt{\frac{an^2}{k(an + n - k)}} \frac{n^{an+n} a^{an}}{k^k (an + n - k)^{an+n-k}}$$

Writing $k$ as $k = n - x\sqrt{n}$,

$$P(\nu(D_{an,n}) \leq k) \sim \frac{a^{-x\sqrt{n}}}{\sqrt{\left(1 - \frac{x}{\sqrt{n}}\right) \left(1 + \frac{x}{a\sqrt{n}}\right)}} \frac{\left(1 - \frac{x}{\sqrt{n}}\right)^{x\sqrt{n}} \left(1 + \frac{x}{a\sqrt{n}}\right)^{-x\sqrt{n}}}{\left(1 - \frac{x}{\sqrt{n}}\right)^n \left(1 + \frac{x}{a\sqrt{n}}\right)^{an}}$$
Considering the limit of each exponential in turn, as \( n \to \infty \),

\[
\left(1 - \frac{x}{\sqrt{n}}\right)^{x\sqrt{n}} \to e^{-x^2} \\
\left(1 + \frac{x}{a\sqrt{n}}\right)^{-x\sqrt{n}} \to e^{-\frac{x^2}{a}} \\
\left(1 - \frac{x}{\sqrt{n}}\right)^n \left(1 + \frac{x}{a\sqrt{n}}\right)^{an} \to e^{-\frac{x^2}{a} \left(1 + \frac{1}{a}\right)} \\
\left(1 - \frac{x}{\sqrt{n}}\right) \left(1 + \frac{x}{a\sqrt{n}}\right) \to 1
\]

Thus,

\[
P(\nu(D_{an,n}) \leq n - x\sqrt{n}) a^x \sqrt{n} \sim e^{-\frac{x^2}{a} \left(1 + \frac{1}{a}\right)}
\]

Note that the presence of the factor \( a^x \sqrt{n} \) implies that, for the case \( a > 1 \), the variance of \( \nu(D_{an,n}) \) is significantly less than \( n \). (The error terms from our expectation calculation give further weight to this hypothesis.) In fact, the concentration about \( n \) is so sharp that, for \( x = o(n^{1/2}) \),

\[
P(\nu(D_{an,n}) \leq n - x) \sim \frac{n^{an+n} a^{an}}{(n-x)^{n-x}(an + x)^{an+x}} \\
= a^{-x} \left(1 - \frac{x}{n}\right)^{-n+x} \left(1 + \frac{x}{an}\right)^{-an-x} \\
\sim a^{-x}
\]

5.11 Difference Graphs: Longest Cycle Length

Now we examine the limiting distribution of \( \psi(D) \), but we do so by relating it to \( \nu(D) \).

**Proposition 5.11.1.** Let \( D \) be a difference graph with \( \psi(D) > 0 \). Then

\[
2\nu(D) - 2 \leq \psi(D) \leq 2\nu(D)
\]
Proof. Let us examine first the point of failure. Suppose that $\psi(D) = 0$, so that $D$ contains no cycles at all. It is nevertheless possible that $\nu(D) = 2$, as shown by the graph induced by the following creation sequence:

\[
\begin{array}{c}
\vdots \cdot 101010\cdots 0 \\
\hline
n_1 \quad n_2
\end{array}
\]

For such a graph, our claim simplifies to the absurd $2 \leq 0$; hence our requirement that $\psi(D) > 0$. Hereafter, we assume that $\psi(D) > 0$, and there exists some cycle $X$ of length $\psi(D)$.

Since $D$ is a bipartite graph, $\psi(D)$ is an even number. Taking every other edge in $X$ results in a matching of size $\frac{\psi(D)}{2}$, and therefore $\psi(D) \leq 2\nu(D)$.

For the other direction, recall from Proposition 4.7.5 that $\psi(D) = 2k$ if and only if $\nu(D(int(C))) = k - 1$, and consider the relationship between $\nu(D)$ and $\nu(D(int(C)))$. Specifically, note that $D$ has one more non-isolated zero-vertex and one more non-isolated one-vertex than $D(int(C))$. Matching these two vertices together shows that

$$\nu(D) \geq \nu(D(int(C))) + 1,$$

but it is sometimes possible to instead match them with vertices in the interior. Specifically, if the interior contains both an unmatched zero-vertex and an unmatched one-vertex, then

$$\nu(D) = \nu(D(int(C))) + 2$$

This is actually the upper bound for $\nu(D)$, as any extra edges would violate the maximality of $\nu(D(int(C)))$. Thus,

$$\nu(D) \leq \nu(D(int(C))) + 2 = \frac{\psi(D)}{2} + 1,$$

and therefore $2\nu(D) - 2 \leq \psi(D)$.

\[\square\]

**Corollary 5.11.2.** For a random difference graph $D_{an,n}$,

$$E[\psi(D_{an,n})] = 2n + O(\sqrt{n})$$
Corollary 5.11.3. For a random difference graph $D_{n,n}$,

$$P \left( \frac{2n - \psi(D_{n,n})}{\sqrt{n}} \geq x \right) \sim e^{-\frac{x^2}{4}},$$

and for a random difference graph $D_{an,n}$ with $a > 1$,

$$P \left( 2n - \psi(D_{an,n}) \geq x \right) = \Theta \left( a^{-\frac{x}{2}} \right)$$

where $0 \leq x = o(\sqrt{n})$.

Proof. In the first claim, the constant difference between $\psi(D_{n,n})$ and $2\nu(D_{n,n})$ is subsumed by the scaling factor of $\sqrt{n}$. For the second claim, note that

$$P(2\nu(D_{an,n}) \leq 2n - x) \leq P(\psi(D_{an,n}) \leq 2n - x) \leq P(2\nu(D_{an,n}) - 2 \leq 2n - x),$$

and therefore by Proposition 5.10.2,

$$a^{-\frac{x}{2}} \leq P(2n - \psi(D_{an,n}) \geq x) \leq a^{-\frac{x}{2} + 1}$$

Thus, we see that $\psi(D_{an,n})$ is bounded in probability, and the cumulative distribution of $2n - \psi(D_{an,n})$ is within a factor of $a$ from $a^{-x/2}$.

If we desire more precision, it is possible to find the joint distribution of $\nu(D_{an,n})$ and $\psi(D_{an,n})$ through analysis of the interior sequence, but be warned that the result is (understandably) even more complex than the distribution of $\psi(D_{an,n})$.

5.12 Difference Graphs: Planarity

Proposition 5.12.1. For a random difference graph $D_{an,n}$,

$$P(D_{an,n} \text{ is planar}) = n^{9/2} \left( \frac{a^a}{(a + 1)^{a+1}} \right)^n \sqrt{\frac{\pi a^5}{8(a + 1)}} \left( 1 + O \left( \frac{1}{n} \right) \right)$$
Proof. From Theorem 4.8.2, we know that

\[ P(D_{\text{an},n} \text{ is planar}) = \left( \frac{an + n}{n} \right)^{-1} \left( \frac{4 + an^2(an^2 - an - n + 5)}{4} \right) \]

\[ = \left( \frac{an + n}{n} \right)^{-1} O(n^4) \]

Applying Stirling to the binomial coefficient, we receive the asymptotics claimed above.

\[ \square \]

5.13 Difference Graphs: Degeneracy

Proposition 5.13.1. For a random difference graph \( D_{\text{an},n} \),

\[ P\left( \frac{\text{degen}(D_{\text{an},n}) - \left( \frac{a}{a+1} \right) n}{\sqrt{n}} = x \right) \sim (a + 1)^{3/2} e^{-\frac{x^2(a+1)^3}{2a^2}}, \]

where \( x = o(\sqrt{n}) \).

Proof. Note that the distribution function \( P(\text{degen}(D_{\text{an},n}) = d) \),

\[ P(\text{degen}(D_{\text{an},n}) = d) = \left( \frac{an + n}{n} \right)^{-1} \left( \frac{an}{d} \right) \left( \frac{n}{d} \right), \]

achieves its maximum near \( d = \left( \frac{a}{a+1} \right) n \). For simplicity of notation, we let \( b = \frac{a}{a+1} \).

For \( d = bn + x\sqrt{n} \), where \( x = o(\sqrt{n}) \),

\[ \left( \frac{an}{d} \right) \sim \sqrt{\frac{1}{2\pi nb(a-b)}} \left( b + \frac{x}{\sqrt{n}} \right)^{bn+x\sqrt{n}} \left( a - b - \frac{x}{\sqrt{n}} \right)^{(a-b)n-x\sqrt{n}} \]

\[ = \sqrt{\frac{1}{2\pi nb^2}} \left( b + \frac{x}{\sqrt{n}} \right)^{bn+x\sqrt{n}} \left( ab - \frac{x}{\sqrt{n}} \right)^{abn-x\sqrt{n}} \]

\[ \sim \sqrt{\frac{1}{2\pi nb^2}} \left( \frac{a^n}{b^n(ab)^{ab}} \right)^n \left( x^n a^x e^{-\frac{x^2(a+1)^3}{2a^2}} \right) \]

\[ \sim \sqrt{\frac{1}{2\pi nb^2}} \left( \frac{a^n}{b^n(ab)^{ab}} \right)^n a^n x^n e^{-\frac{x^2(a+1)^3}{2a^2}} \]
Similarly,
\[
\binom{n}{d} \sim \sqrt{\frac{1}{2\pi nb(1-b)}} \left( b + \frac{x}{\sqrt{n}} \right)^{bn+x\sqrt{n}} \left( 1 - b - \frac{x}{\sqrt{n}} \right)^{(1-b)n-x\sqrt{n}}
\]
\[
= \sqrt{\frac{a}{2\pi nb^2}} \left( \frac{1}{b^{b/a}} \right)^n \left( 1 - \frac{a}{b\sqrt{n}} \right)^{bn+x\sqrt{n}} \left( 1 - \frac{a}{b\sqrt{n}} \right)^{bn-x\sqrt{n}}
\]
\[
\sim \sqrt{\frac{a}{2\pi nb^2}} \left( \frac{a^{b/a}}{b^{b/a}} \right)^n e^{-\frac{x^2(1+a)}{2b}}
\]
And since
\[
\binom{an+n}{n}^{-1} \sim \sqrt{\frac{2\pi an}{a+1}} \left( \frac{a^a}{(a+1)^{a+1}} \right)^n
\]
we see that
\[
\binom{an+n}{n}^{-1} \binom{an}{d} \binom{n}{d} \sim \sqrt{\frac{a}{2\pi nb^2}} \sqrt{\frac{2\pi an}{a+1}} \left( \frac{a^{2a-ab+b/a}}{b^{2b+ab+b/a}(a+1)^{a+1}} \right)^n e^{-\frac{x^2(a+1)^2}{2ab}}
\]
\[
= \frac{(a+1)^2}{a} \sqrt{\frac{1}{2\pi(a+1)n}} \left( \frac{a}{b(a+1)} \right)^{n(a+1)} e^{-\frac{x^2(a+1)^2}{2ab}}
\]
\[
= \frac{(a+1)^{3/2}}{a\sqrt{2\pi n}} e^{-\frac{x^2(a+1)^3}{2a^2}}
\]
\[
\square
\]
91
CHAPTER 6
OPEN PROBLEMS AND GENERALIZATIONS

The deceptive simplicity of threshold graph generation lends itself easily to generalization. However, the true beauty of threshold and difference graphs is their ability to harness the infinite freedom of continuous weights through the discrete representations of creation sequences.

Extensions to the above research break down into two categories: new applications of the current models, or entirely new models.

6.1 Unexplored Issues

Given the sheer numbers of graph properties that have been defined, it should not be surprising that many of them are still open for investigation. While we tried to sample invariants from many different areas, the vast majority are still untouched.

For example, it has been shown that the probability of a threshold graph being planar approaches zero as the order approaches infinity. (Likewise for difference graphs wherein the size of each part approaches infinity.) So what is the distribution of the genus of a random threshold or difference graph?

Alternatively, consider the arboricity of a graph, the minimum number of forests into which its edges can be partitioned. Although the arboricity is always within a factor of two of the degeneracy, it remains to find the precise distribution.
But the challenge inherent to these, and to many other properties, is to find some relationship with the creation sequence. For while it is not impossible to express a characteristic in terms of vertex weights, it is the beauty of the binary encoding which permits the simplicity of the results we found above.

6.2 Difference Graphs of Undetermined Parts

When generating our random difference graphs, we fixed the sizes of the two parts at \( n_0 \) and \( n_1 \). Thus, the creation sequence was forced to contain specified numbers of zeroes and ones, so the digits were no longer generated independently.

In this variation, we instead generate the creation sequences by choosing each digit \( d_i \) independently and identically, with distribution

\[
P(d_i = d) = \begin{cases} 
\frac{n_0}{n_0 + n_1} & d = 0 \\
\frac{n_1}{n_0 + n_1} & d = 1
\end{cases}
\]

That is, if we let \( p = \frac{n_1}{n_0 + n_1} \), each digit equals 1 with probability \( p \) and 0 with probability \( 1 - p \).

The immediate negative consequence is that this model would diminish possible applications; most scenarios, such as employees and shifts, processors and tasks, or instructors and classes, have sets of prescribed size, and the elements of each part are not interchangable.

However, the benefits are twofold. First, it becomes easier to calculate the probability of certain sequences, given the newfound independence. Second, with proper conditioning on the number of zeroes and ones, it is possible to apply results from this model \( D_p \) to the standard model \( D_{n_0,n_1} \), much in the same way that results from the Erdős-Reyni model \( G(n,m) \) of traditional random graphs can be applied to the Bernoulli model \( G(n,p) \).
6.3 Hypergraphs

In light of the multiple motivating applications for threshold graphs, their generalizations to higher dimensions should not come as a surprise. What is surprising, however, are the difficulties involved in doing so.

Above, we mentioned that there are at least eight equivalent characterizations of threshold graphs. And while many of these characterizations offer intuitive extensions to hypergraphs, the results need not be equal. In particular, J. Reiterman et al. [14] showed that three particular generalizations by Golumbic [6] produced three distinct classes of hypergraphs.

That said, let us consider one of Golumbic’s proposals: given a set $V$ of $n$ vertices with weights $w(v_1), \ldots, w(v_n)$, we can define a $r$-uniform hypergraph on $V$ through the definition of edge set $E$ as

$$E = \left\{ S \subset V : |S| = r \text{ and } \sum_{v \in S} w(v) > t \right\}$$

How many possible threshold $r$-hypergraphs exist? Is it possible to encode them?

Interestingly, there exists a bijection between such graphs and certain classes of $r$-dimensional solid diagrams. Specifically, the number of $r$-uniform threshold hypergraphs on $n$ vertices is equal to the number of totally symmetric $r$-dimensional solid diagrams where each dimension has size at most $n - r + 1$. (This result was already discovered by Klivans [9], who found intermediate bijections with certain sets of shifted complexes.) While such numbers have already been found, at least for smaller values of $r$, and strongly suggest that simple encoding may not be an option, this does open new possibilities for applications and representations.
6.4 Multigraphs

Given a set $V$ of $n$ vertices with weights $w(v_1), \ldots, w(v_n)$, and a set $T$ of $k$ real numbers, we can define a multigraph on $V$ by adding edge ${v_i, v_j}$ with multiplicity $m_{i,j}$, where

$$m_{i,j} = |\{t \in T : w(v_i) + w(v_j) > t\}|$$

That is, each edge is added a number of times equal to the number of thresholds exceeded by the sum of the weights of its endpoints. As there are $k$ thresholds, we refer to such a graph as a $k$-threshold graph.

Viewed another way, these $k$-threshold graphs can be viewed as the product of superimposing $k$ threshold graphs, each on the same set of vertices. These preliminary graphs cannot be combined freely, however: if one graph implies that the weight of vertex $v_i$ exceeds that of $v_j$, all other graphs must respect that restriction.

These additional edges offer further possibilities to applications. For example, a double edge between a lecturer and a class may represent the ability to teach two sections of the course during the semester in question. Thus, we find ourselves once again asking the same questions about the graph properties.
BIBLIOGRAPHY


