Properties of Random Difference Graphs

Christopher Ross

The Ohio State University
Columbus, Ohio, USA

`cross@math.osu.edu`

Abstract

Generate a bipartite graph on a partitioned set of vertices by randomly assigning to each vertex v some weight $w(v) \in [0,1]$ and adding an edge between vertices u and v (in distinct parts) if and only if $w(u) + w(v) > 1$; the results of such processes are known as *difference graphs*.

Random difference graphs of a given size can be produced either by uniformly random generation of weights or by choosing a graph uniformly at random from the set of all such graphs. We prove that these two methods give rise to the same distribution, and use this equivalence to find exact results for the likelihood of connectivity and Hamiltonicity. We also find the distribution of other properties, such as matching number and degeneracy.

1 Introduction

We say that a bipartite graph is a *difference graph* if we can assign weights to the vertices such that two vertices in distinct parts are adjacent if and only if the sum of their weights exceeds some fixed threshold value. Without loss of generality, we can restrict these weights to the interval $[0,1]$ and set the threshold to 1.

To generate a random difference graph of a particular size, there are two natural options. The simplest is to choose a difference graph uniformly at random from the set of all difference graphs with the appropriate number of vertices. The alternative is to determine the vertex weights according to some distribution, and let the weights generate the difference graph. We shall show that, for certain weight distributions, both of these methods are equivalent, and then use this equivalence to determine the distribution of such properties as connectivity, matching number, and the length of the longest cycle.

Difference graphs were formally defined by Hammer, Peled, and Sun [5] in 1990, although they had been independently explored prior to that point [3], a result of the multiple equivalent characterizations for this class of graphs [6]. They are closely related

to threshold graphs, which can be viewed as difference graphs without the bipartitioning; this similarity enables us to adapt the techniques used in the study of random threshold graphs, notably those of Reilly and Scheinerman [7]. Here, we focus on difference graphs with a specified bipartition, also known as *bipartite threshold graphs* (Diaconis, Holmes, Janson [4]).

2   Fundamentals

Before showing the equivalence of models, we introduce some essential information and known results about difference graphs.

Definition 1. A bipartite graph G with vertex parts Z and U is a *difference graph* if and only if there exists a function $w : V(G) \to [0,1]$ such that for all $x \in Z$ and $y \in U$,

$$\{x, y\} \in E(G) \iff w(x) + w(y) > 1$$

An equivalent formulation assigns negative weights to vertices in one of the parts, in which case adjacency depends upon the difference of the weights, motivating the name. The choice to label the parts Z and U - for "zero" and "one", respectively - will shortly become more clear. As a consequence of this definition, all difference graphs exhibit certain structural similarities:

Proposition 2. *Let G be a difference graph on at least one vertex, with vertex sets 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. We can 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.

Suppose that $|Z|$, $|U| \geqslant 1$. Let z_{min} be a vertex of Z such that for all $z \in Z$, $w(z_{min}) \leqslant w(z)$, and let u_{max} be a vertex of U such that for all $u \in U$, $w(u) \leqslant w(u_{max})$. We consider two disjoint cases:

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

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

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

On the other hand, if $w(z_{min}) + w(u_{max}) \leqslant 1$, then there does not exist an edge between z_{min} and u_{max}. So for any $u \in U$,

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

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

This property, coupled with the fact that any induced subgraph of a difference graph is another difference graph, allows us to encode such graphs as binary strings. These strings, known as *creation sequences*, completely describe the structure of a difference graph, and allow a convenient way to reference particular graphs.

Specifically, for a difference graph G on n vertices, the corresponding creation sequence, denoted $\text{seq}(G)$, is a binary string of length n that is constructed from right to left as follows: from Proposition 2, there exists some vertex v of G such that v is either an isolated vertex of Z or a vertex of U that dominates all of Z. We remove this vertex, and add a single corresponding digit to our sequence: in the former case, the digit corresponding to v is 0, whereas in the latter case it is 1. We then remove v from G and repeat this process, continuing until all n vertices have been removed.

More formally, we can recursively construct $\text{seq}(G)$ by defining the creation sequence of the empty graph as the empty sequence. Then taking vertex v to be the vertex of G necessitated by Proposition 2,

$$\text{seq}(G) = \text{seq}(G \setminus \{v\}) \, d, \text{ where } d = \begin{cases} 0 & \text{if } v \in Z \\ 1 & \text{if } v \in U \end{cases},$$

and $\text{seq}(G \setminus \{v\}) \, d$ represents the appending of digit d to the creation sequence of the subgraph induced by the removal of vertex v.

To recreate the graph, we read the sequence from left to right. Every time we encounter a zero, we add a vertex to Z. On the other hand, every time we encounter a one, we add a vertex to U as well as edges from that new vertex to all existing vertices of Z. In this fashion we reverse the process, at each step adding either an isolated vertex of Z or a vertex of U that dominates all of Z.

There are several immediate consequences of this construction. First, if G is a difference graph with vertex parts Z and U, then $\text{seq}(G)$ has exactly $|Z|$ zeroes and $|U|$ ones. Furthermore, two vertices z_i and u_j are adjacent in G if and only if the digit corresponding to z_i lies to the left of u_j's in $\text{seq}(G)$.

To prove that this process forms a bijection, it remains to show surjectivity. But it is actually simpler to prove a slight strengthening:

Proposition 3. *Let S be a binary sequence of length k. Then there exists a difference graph G such that $\text{seq}(G) = S$ and for all $v \in G$,*

$$2^{-k} \leqslant w(v) \leqslant 1 - 2^{-k}$$

Proof. If S is the empty sequence of length zero, then we let G be the empty graph on zero vertices. We then proceed by induction, supposing that S has length $k \geqslant 1$. Let S' be the sequence consisting of the first $k-1$ digits of S, and d the final digit, such that $S = S'd$. By induction, there exists some graph G', with vertex parts Z' and U', such that $\text{seq}(G') = S'$.

If $d = 0$, then let Z be formed by adding a new vertex to Z' of weight 2^{-k}. By induction, all vertices of G' have weight at most $1 - 2^{-k+1}$, so no edges are formed by this addition. Thus, the difference graph G with parts Z and U' has creation sequence S.

If $d = 1$, then let U be formed by adding a new vertex to U' of weight $1 - 2^{-k}$. By induction, all vertices of Z' have weight at least 2^{-k+1}, so all of them are adjacent to this new addition. Thus, the difference graph G with parts Z' and U has creation sequence S. $\qquad\square$

(Please note that the above process will produce a difference graph where each of the vertices has a distinct weight, and furthermore no pair of vertices has weights that add to exactly 1.)

Thus, we have a bijection between the set of k-long binary sequences and the set of difference graphs on k vertices. In particular, this shows that there are exactly $\binom{n_0 + n_1}{n_0}$ difference graphs with part sizes $|Z| = n_0$ and $|U| = n_1$.

3 Comparison of Models

The previous section suggests two models for the random 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

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

each random variable X_i and Y_j being 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 let $G(\vec{w})$ denote the unlabeled random graph resulting from this process. (Equivalently, we can view $G(\vec{w})$ as the isomorphism class of the graph on vertex set $Z \cup U$.) Thus, for any difference graph G of appropriate size, the probability of generating G under this model is the measure of the set $\{\vec{w} \in [0, 1]^{n_0 + n_1} : G(\vec{w}) = G\}$.

Second, the *discrete model*, or creation sequence model, begins 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 an isolated vertex to part Z. Every time we encounter a one, we add a vertex to part U and edges from that vertex to every vertex in Z. Let $G(C_{n_0, n_1})$ denote the unique graph resulting from this process. So for any difference graph G, the probability of generating G under this model is exactly $\binom{n_0 + n_1}{n_0}^{-1}$.

We shall show that these two methods are equivalent. That is, for a given difference graph G, the probability of a randomly-chosen weighting vector generating G is equal to the probability of a randomly-chosen creation sequence generating G. To do this, we have to examine those $\vec{w}$ such that $G(\vec{w}) = G$.

3.1 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:

- $\forall i, j \in [n_0]$ with $i < j$, $\alpha_{i,j} = \{\vec{v} \in \mathbb{R}^{n_0 + n_1} : x_i = x_j\}$

- $\forall i, j \in [n_1]$ with $i < j$, $\beta_{i,j} = \{\vec{v} \in \mathbb{R}^{n_0 + n_1} : y_i = y_j\}$

- $\forall i \in [n_0], \forall j \in [n_1]$, $\gamma_{i,j} = \{\vec{v} \in \mathbb{R}^{n_0 + n_1} : x_i + y_j = 1\}$

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 for any given region R, all of the weight vectors within give rise to the same difference graph.

Proposition 4. *For any $R \in \mathcal{R}$ and any $\vec{w}, \vec{w}' \in R$, $G(\vec{w}) = G(\vec{w}')$.*

Proof. As both $\vec{w}$ and $\vec{w}'$ are on the same side of all hyperplanes of the form γ, 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'. So 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'. Therefore $G(\vec{w}) = G(\vec{w}')$. $\qquad \square$

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

3.2 Counting the Regions

To count the regions, we establish a bijection between $\mathcal{R}$ and $S_{n_0} \times S_{n_1} \times \mathbb{C}_{n_0, n_1}$, where S_k is the set of all permutations on $[k] = \{1, \ldots, k\}$.

Let R be an element of $\mathcal{R}$, and let $\vec{w}$ be an element of R. Let σ_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, σ_R is the permutation that sorts the first n_0 coordinates of $\vec{w}$ into increasing order. Note that the α-hyperplanes cause this permutation to be well-defined and constant over 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 $\text{seq}(R) \in \mathbb{C}_{n_0,n_1}$ be the creation sequence such that $G(\text{seq}(R)) = G(\vec{w})$. By Proposition 4, above, this sequence is uniquely defined and independent of the choice of $\vec{w} \in R$.

Thus, we can define a mapping

$$f : \mathcal{R} \to S_{n_0} \times S_{n_1} \times \mathbb{C}_{n_0,n_1} \text{ by } f : R \mapsto (\sigma_R, \tau_R, \text{seq}(R)).$$

Proposition 5. *The mapping f is a bijection.*

Proof. To see that f is injective, take $R, R' \in \mathcal{R}$ such that $R \neq R'$. Then there must exist some hyperplane ϕ, where ϕ is of the form $\alpha_{i,j}$, $\beta_{i,j}$, or $\gamma_{i,j}$, such that R and R' are on opposite sides of ϕ.

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 vice-versa. Thus, σ_R and $\sigma_{R'}$ are different permutations, and $f(R)$ and $f(R')$ differ in the first coordinate.

Similarly, if ϕ 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 ϕ is of the form $\gamma_{i,j}$ and there exist no separating planes of the form α or β, 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.

To show that f is surjective, we take an arbitrary $(\sigma, \tau, C_{n_0,n_1}) \in S_{n_0} \times S_{n_1} \times \mathbb{C}_{n_0,n_1}$. Since $G(C_{n_0,n_1})$ is a difference graph, there must exist some weight vector $\vec{w}$ such that $G(\vec{w}) = G(C_{n_0,n_1})$ by Proposition 3. (The methods in that proof construct a weight vector that does not intersect any of the hyperplanes, and is therefore in $\mathbb{P}_{n_0,n_1}$.)

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

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

Using $\vec{w}$, as well as all four permutations, we define a final weight vector $\vec{w}_2$ by permuting the coordinates of $\vec{w}$. Specifically,

$$\vec{w}_2 = \left\langle x_{\sigma^{-1}(\sigma_w(1))}, \ldots, x_{\sigma^{-1}(\sigma_w(n_0))}, y_{\tau^{-1}(\tau_w(1))}, \ldots, y_{\tau^{-1}(\tau_w(n_1))} \right\rangle \in \mathbb{P}_{n_0,n_1}$$

Let R be the element of $\mathcal{R}$ that contains $\vec{w}_2$. Then, since the permutations σ and τ sort the coordinates of $\vec{w}_2$, they must be the first and second coordinates of $f(R)$. And since the resulting sorted vector produces a difference graph whose creation sequence is C_{n_0,n_1}, we have $f(R) = (\sigma, \tau, C_{n_0,n_1})$. $\square$

Corollary 6. *There are $(n_0 + n_1)!$ connected regions of $\mathbb{P}_{n_0,n_1}$.*

Proof. From the above,

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

$\square$

3.3   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 7. *For any region $R \in \mathcal{R}$, any reflection of R across a hyperplane of type α, β, or γ produces another element of $\mathcal{R}$.*

Proof. Since every such reflection ϕ is a continuous mapping, the image of any connected set is connected. Furthermore, since ϕ 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

$$\vec{w} = \langle x_1, \ldots, x_{n_0}, y_1, \ldots, y_{n_1} \rangle$$

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

As reflections of the α and β 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 $\gamma_{i,j}$:

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

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, as such would also imply that $\vec{w} \notin \mathbb{P}_{n_0,n_1}$.

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

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 8. *Let $\mathbf{1}_{n_0}$ and $\mathbf{1}_{n_1}$ 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:*

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

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

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 α-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 β-reflections to (increasingly) sort the last n_1 coordinates.

The resulting vector has all x_i and y_j listed in increasing order, and thus its region corresponds to the identity permutations $\mathbf{1}_{n_0}$ and $\mathbf{1}_{n_1}$. As for the corresponding difference graph, due to the sorting and Step 2, for any $i \in [n_0], j \in [n_1]$,

$$x_i + y_j < x_{n_0} + y_{n_1} < 1,$$

and therefore there are no edges in the graph.

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

Since every region can be mapped onto this "base region" by some composition of invertible rigid transformations, all of the regions are congruent.

3.4 Conclusion

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

Now let us fix an arbitrary difference graph G with part sizes n_0 and 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 $G(C_{n_0,n_1})$, each of the $\binom{n_0 + n_1}{n_0}$ sequences is equally likely. So the probability of generating $\text{seq}(G)$ is:

$$P(G(C_{n_0,n_1}) = G) = \binom{n_0 + n_1}{n_0}^{-1} = \frac{n_0! n_1!}{(n_0 + n_1)!}$$

On the other hand, when using the random weighting model $G(\vec{w})$, a random vector $\vec{w}$ generates G if and only if it lies in a region R such that the third coordinate of $f(R)$ is the creation sequence of G. Thus,

$$P(G(\vec{w}) = G) = n_0! n_1! P(\vec{w} \in f^{-1}(\mathbf{1}_{n_0}, \mathbf{1}_{n_1}, \text{seq}(G))) = \frac{n_0! n_1!}{(n_0 + n_1)!}$$

Therefore, the two models of generating difference graphs have the exact same distribution.

4 Applications

The power of the previous section is that we can now discard the continuous variables of the traditional difference graph model. Instead, we need only consider the creation sequence.

For each graph property, 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 said property.

4.1 Notation

Let G_{n_0,n_1} be a difference graph with part sizes $|Z| = n_0$ and $|U| = n_1$. Given a binary sequence S, we let $G(S)$ denote the difference graph whose creation sequence is S. Conversely, given a difference graph G, let $\text{seq}(G)$ denote the creation sequence of G.

Please note that although we are 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 G_{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 left to right. As a result 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 z_i has lower index than z_{i+1}, so any one-vertex adjacent to z_{i+1} is also adjacent to z_i.

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 position. 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 $s_i s_{i+1} \ldots s_n$. 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 $1 \leqslant i \leqslant n_0 + n_1 + 1$, $t_i^u(C_{n_0,n_1}) + t_i^z(C_{n_0,n_1}) = n_0 + n_1 - i + 1$.

4.2 Connectivity and Degrees

Proposition 9. *Let G be a difference graph with at least two vertices. Then G is connected if and only if no vertex of G is isolated.*

Proof. By Proposition 2, 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 G, 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 G, there exists a path between them through u or z, and therefore G is connected. $\square$

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

Proposition 10. *Let G_{n_0,n_1} be a random difference graph such that $n_0 + n_1 \geqslant 2$. Then*

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

Proof. By Proposition 9, G 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 have the highest index. The remaining vertices, arranged between these two extremes, can have any configuration without impacting the connectivity. Therefore,

$$P(G \text{ is connected}) = \binom{n_0 + n_1 - 2}{n_0 - 1}\binom{n_0 + n_1}{n_0}^{-1} = \frac{n_0 n_1}{(n_0 + n_1)(n_0 + n_1 - 1)}$$

$\square$

We can generalize the above results to broader statements about the connectivity of G_{n_0,n_1}, specifically vertex-connectivity. A graph is *k-vertex-connected* if it remains connected after the removal of up to $k - 1$ vertices. (As we will not be concerning ourselves with edge-connectivity, such a graph is simply said to be *k-connected*).

Proposition 11. *Let G_{n_0,n_1} be a difference graph. For $k \leqslant n_0 + n_1 - 1$, G_{n_0,n_1} is k-connected if and only if $seq(G_{n_0,n_1})$ begins with at least k consecutive zeroes and ends with at least k consecutive ones.*

Proof. First, suppose that G_{n_0,n_1} is k-connected. Then every vertex v in G_{n_0,n_1} must have degree at least k. 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 $seq(G_{n_0,n_1})$ begins with at least k consecutive zeroes and ends with at least k consecutive ones. Let $V = V(G_{n_0,n_1})$ denote the vertex set of G_{n_0,n_1}, and

take any $A = \{a_1, \ldots, a_{k-1}\} \subset V$; we shall show that the induced subgraph on $V \setminus A$ remains connected.

Since $\mathrm{seq}(G_{n_0,n_1})$ 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. As 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, so G_{n_0,n_1} is k-connected. $\qquad\square$

Corollary 12. *Let G_{n_0,n_1} be a random difference graph. Then for $k \leqslant n_0 + n_1 - 1$,*

$$P(G_{n_0,n_1} \text{ is } k\text{-vertex-connected}) = \binom{n_0 + n_1 - 2k}{n_0 - k}\binom{n_0 + n_1}{n_0}^{-1} = \frac{(n_0)_k (n_1)_k}{(n_0 + n_1)_{2k}},$$

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

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

Corollary 13. *Let G_{n_0,n_1} be a random difference graph. Then*

$$P(\kappa(G_{n_0,n_1}) = k) = \binom{n_0 + n_1}{n_0}^{-1}\left(\binom{n_0 + n_1 - 2k}{n_0 - k} - \binom{n_0 + n_1 - 2k - 2}{n_0 - k - 1}\right)$$

Proof. For the connectivity of G_{n_0,n_1} to equal k, G_{n_0,n_1} must be k-connected but not $(k+1)$-connected. Since the event of G_{n_0,n_1} being a $(k+1)$-connected graph necessitates that it is also k-connected,

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

$\qquad\square$

Having looked at the connectedness of the vertices, we now turn to their degrees, starting with the minimum degree, $\delta(G)$.

Proposition 14. *Let G be a difference graph. Then $\delta(G) = \kappa(G)$.*

Proof. By the same argument from the first part of Proposition 11, we see that $\delta(G) \geqslant k$ if and only if G both begins with at least k consecutive zeroes and ends with at least k consecutive ones. As such, $\delta(G) \geqslant k$ exactly when G is k-connected, so $\delta(G) = k$ if and only if $\kappa(G) = k$. $\qquad\square$

Corollary 15. *Let G_{n_0,n_1} be a random difference graph. Then*

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

Having found the distribution of the minimum degree $\delta(G_{n_0,n_1})$, we continue by finding the distribution of the maximum degree, denoted $\Delta(G_{n_0,n_1})$. This requires us to introduce a construction which will be used both here and in subsequent sections.

Recall that a difference graph G is connected if and only if $\mathrm{seq}(G)$ begins with a 0 and ends with a 1. Consequently, we define the *interior* of creation sequence S, denoted $int(S)$, 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 $int(S) = c_{i+1}c_{i+2}\ldots c_{j-2}c_{j-1}$.

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

Since $int()$ removes the right-most one and the left-most zero, $int(C_{n_0,n_1})$ 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 $0 \leqslant i \leqslant n_0 - 1$ and $0 \leqslant j \leqslant n_1 - 1$, there exists a C_{n_0,n_1} such that $int(C_{n_0,n_1}) = S$. To see this, we define C_{n_0,n_1} as follows:

$$C_{n_0,n_1} = \underbrace{1\cdots1}_{n_1-j-1}0S1\underbrace{0\cdots0}_{n_0-i-1}$$

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

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

through repeated use of the identity $\displaystyle\sum_{a=0}^{b}\binom{c+a}{c} = \binom{c+b+1}{b}$.

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

Proposition 16. *Let G be a nonempty difference graph. Then for $k \geqslant 1$, $\Delta(G) = k$ if and only if the maximum of the number of zeroes and the number of ones in $int(\mathrm{seq}(G))$ is $k - 1$.*

Proof. First, suppose that $\Delta(G) = k \geqslant 1$, and that, without loss of generality, there exists a zero-vertex z of degree k. Since all non-isolated vertices lie in a single connected component, said component must include z, exactly k one-vertices, and some number of other zero-vertices. Because there exists a one-vertex that dominates all zero-vertices in the component, there can be at most k zero-vertices therein. As the non-trivial component of G contains k one-vertices and at most k zero-vertices, $int(\mathrm{seq}(G))$ contains $k - 1$ ones and at most $k - 1$ zeroes.

Similarly, if $int(\mathrm{seq}(G))$ should contain exactly $k - 1$ zeroes and at most $k - 1$ ones, then the non-trivial component of G consists of k zero-vertices and at most k one-vertices. So there exists a one-vertex of degree k, and every other vertex has degree at most k. □

Theorem 17. *For a difference graph G_{n_0,n_1} with $n_0 + n_1 \geqslant 1$, the probability that $\Delta(G_{n_0,n_1}) = k$ equals, for $0 \leqslant k \leqslant \max(n_0, n_1)$,*

$$
\begin{cases}
\dbinom{n_0 + n_1}{n_0}^{-1} & k = 0 \\[2ex]
\dbinom{n_0 + n_1}{n_0}^{-1} \left[\dbinom{2k-2}{k-2} + \dbinom{2k-1}{k-1} \right] & 1 \leqslant k \leqslant \min(n_0, n_1) \\[2ex]
\dbinom{n_0 + n_1}{n_0}^{-1} \dbinom{\min(n_0, n_1) + k - 1}{k} & \min(n_0, n_1) < k \leqslant \max(n_0, n_1)
\end{cases}
$$

Proof. As G_{n_0,n_1} is bipartite, the maximum degree can never exceed the larger of n_0 and n_1. On the other extreme, the only way for the maximum degree to be zero is for G_{n_0,n_1} to be the empty graph, for which there is exactly one configuration, having probability $\dbinom{n_0 + n_1}{n_0}^{-1}$.

Decomposing the event given in Proposition 16, we see that the corresponding probability is given by:

$$
\begin{aligned}
P(\Delta(G_{n_0,n_1}) = k) = \ & P(int(\mathrm{seq}(G_{n_0,n_1})) \text{ has } k - 1 \text{ zeroes}, \leqslant k - 1 \text{ ones}) \\
& + P(int(\mathrm{seq}(G_{n_0,n_1})) \text{ has } k - 1 \text{ ones}, \leqslant k - 1 \text{ zeroes}) \\
& - P(int(\mathrm{seq}(G_{n_0,n_1})) \text{ has } k - 1 \text{ zeroes and } k - 1 \text{ ones})
\end{aligned}
$$

Depending upon the relationships between k, n_0, and n_1, one or more of those probabilities may be zero. (E.g., for $n_0 < k < n_1$, it is impossible to have $k - 1$ zeroes in the interior of $\mathrm{seq}(G_{n_0,n_1})$.) But in the event that $k \leqslant \min(n_0, n_1)$, the total probability can be written as:

$$
\begin{aligned}
P(\Delta(G_{n_0,n_1}) = k) &= \binom{n_0 + n_1}{n_0}^{-1} \left[\sum_{j=0}^{k-1} \binom{j+k-1}{j} + \sum_{i=0}^{k-1} \binom{i+k-1}{i} - \binom{2k-2}{k-1} \right] \\
&= \binom{n_0 + n_1}{n_0}^{-1} \left[2\binom{2k-1}{k-1} - \binom{2k-2}{k-1} \right] \\
&= \binom{n_0 + n_1}{n_0}^{-1} \left[\binom{2k-2}{k-2} + \binom{2k-1}{k-1} \right]
\end{aligned}
$$

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

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

$\square$

4.3   Matching Number

Here we find the distribution of the size of the maximum matching of a difference graph, denoted $\nu(G_{n_0,n_1})$.

We define a function h on the set of all finite binary sequences by

$$h(S) = \max_{1 \leqslant i \leqslant |S|+1} \{t_i^z(S) - t_i^u(S)\}$$

That is, $h(S)$ is the maximum number of excess zeroes across all tails of sequence S. Note that h is always non-negative, as the tail corresponding to $i = |S|+1$ is the empty sequence, with equal numbers of zeroes and ones. Similarly, by taking the tail corresponding to $i = 1$, we see that

$$n_0 - n_1 \leqslant h(\text{seq}(G_{n_0,n_1})) \leqslant n_0$$

Proposition 18. *For a difference graph G_{n_0,n_1} with matching number $\nu(G_{n_0,n_1})$,*

$$n_0 - \nu(G_{n_0,n_1}) = h(seq(G_{n_0,n_1}))$$

Proof. First, we 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 $h(\text{seq}(G_{n_0,n_1})) = t_j^z(\text{seq}(G_{n_0,n_1})) - t_j^u(\text{seq}(G_{n_0,n_1}))$. Then the tail beginning at j has $h(\text{seq}(G_{n_0,n_1}))$ 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 $h(\text{seq}(G_{n_0,n_1}))$ of them must remain unmatched in any matching, and therefore $n_0 - \nu(G_{n_0,n_1}) \geqslant h(\text{seq}(G_{n_0,n_1}))$.

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

Thus there are $n_0 - h(\text{seq}(G_{n_0,n_1}))$ 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 $\text{seq}(G_{n_0,n_1})$, $G(S)$ is an induced subgraph of G_{n_0,n_1}. Thus the original sequence must also have at least that many pairs, and $\nu(G_{n_0,n_1}) \geqslant n_0 - h(\text{seq}(G_{n_0,n_1}))$. □

Having found the properties of the creation sequence that correspond to the matching number $\nu(G_{n_0,n_1})$, we can calculate the distribution. In order to do so, we must first find the distribution of $h(\text{seq}(G_{n_0,n_1}))$:

Lemma 19. *For a random difference graph G_{n_0,n_1},*

$$P\left(h(seq(G_{n_0,n_1})) = k\right) = \binom{n_0 + n_1}{n_0 - k} - \binom{n_0 + n_1}{n_0 - k - 1}$$

Proof. For $h(\text{seq}(G_{n_0,n_1}))$ to equal k, some tail of $\text{seq}(G_{n_0,n_1})$ must have exactly k more zeroes than ones, but in no tail is there ever $k+1$ more zeroes than ones.

To count the number of such sequences, we 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; such sequences of moves are known as "monotonic paths" or "staircase walks". By reading the creation sequence from right to left, the number of G_{n_0,n_1} such that $h(\text{seq}(G_{n_0,n_1})) = k$ is equal to the number of staircase walks from $(0,0)$ to (n_1, n_0) that touch, but do not cross, the line $y = x + k$.

By a reflection argument, we can count the number of walks that touch the line $y = x + k$ as $\binom{n_0 + n_1}{n_0 - k}$. We then subtract all of the walks which cross the line, which are in turn all walks that touch the line $y = x + k + 1$. $\qquad\square$

Theorem 20. *For a random difference graph G_{n_0,n_1}, the distribution of the matching number $\nu(G_{n_0,n_1})$ is given by*

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

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

Proof. By Proposition 18, $\nu(G_{n_0,n_1})) = k$ if and only if $h(\text{seq}(G_{n_0,n_1})) = n_0 - k$. Therefore, the probability that $\nu(G_{n_0,n_1}) = k$ 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. From Lemma 19, this number is exactly

$$\binom{n_1 + n_0}{k} - \binom{n_1 + n_0}{k - 1}$$

Therefore,

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

$\qquad\square$

4.4 Cycles

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

Proposition 21. *For a difference graph G, there exists a cycle in G if and only if the creation sequence $\text{seq}(G)$ contains 0011 as a subsequence.*

Proof. Suppose that $\text{seq}(G)$ 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 G contains $K_{2,2}$ as a subgraph, and a cycle of length four.

Next, suppose that G contains a cycle; since G is bipartite, said cycle must contain at least two vertices from each part. So there exist two zero-vertices, the higher-index of which has degree at least two. Thus, the digits of $\mathrm{seq}(G)$ corresponding to those four vertices are a subsequence of the form 0011. $\qquad\square$

With this result,

Theorem 22. *For a random difference graph G_{n_0,n_1},*

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

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

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

$\qquad\square$

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

Proposition 23. *If G_{n_0,n_1} contains a cycle of length $2k$, then it contains a $2k$-cycle of the form*

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

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

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:

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

Notice that within X, each z_i (with the exception of z_1) is adjacent to both $u_{n_1-k+i-1}$ and u_{n_1-k+i}.

By the existence of Y, there must exist k one-vertices of positive degree, so vertex u_{n_1-k+1} must be non-isolated. 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 $i \geqslant 2$ such that z_i is not adjacent to $u_{n_1-k+i-1}$. (If z_i is not adjacent to u_{n_1-k+i}, then it also cannot be adjacent to $u_{n_1-k+i-1}$, as the latter has lower index.)

Since z_i is not adjacent to $u_{n_1-k+i-1}$, every possible neighbor of z_i must lie in the set $\{u_{n_1-k+i}, \ldots, u_{n_1}\}$, so z_i has at most $(k-i+1)$ neighbors. Furthermore, for all $j > i$, z_j is also not adjacent to $u_{n_1-k+i-1}$; letting $A = \{z_i, \ldots, z_k\}$, we see that $|N(A)| \leqslant k - i + 1 = $

$|A|$. By Hall's Marriage Theorem, since the restriction of Y to $A \cup N(A)$ contains a matching, $|N(A)| \geqslant |A|$.

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. $\qquad\square$

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

Proposition 24. *If G_{n_0,n_1} is a difference graph, then for all $k \geqslant 2$, $\psi(G_{n_0,n_1}) = 2k$ if and only if $\nu(G(int(seq(G_{n_0,n_1})))) = k - 1$.*

Proof. First, suppose that the longest cycle is of length $2k$. Then by Proposition 23, the following cycle must exist:

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

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

Second, to show the converse, suppose that $\nu(G(int(seq(G_{n_0,n_1})))) = k - 1$. Then there must exist edges in G_{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 σ and τ are injective, and for all i, $\sigma(i) \geqslant 2$ and $\tau(i) \leqslant n_1 - 1$. 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 π obeys the same constraints as τ.) And therefore, the following cycle exists in G_{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, $\psi(G_{n_0,n_1}) \geqslant 2k$. $\qquad\square$

Having shown that $\psi(G_{n_0,n_1})$ depends solely on $int(seq(G_{n_0,n_1}))$, we take a deeper look at the latter.

Theorem 25. *If G_{n_0,n_1} is a difference graph, then for $k \geqslant 2$, $P(\psi(G_{n_0,n_1}) = 2k)$ is given by*

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

Proof. Note that

$$P(\psi(G_{n_0,n_1}) = 2k) = \binom{n_0 + n_1}{n_0}^{-1} \sum_{G_{n_0,n_1} \in \mathbb{G}_{n_0,n_1}} \mathbf{1}_{\{\psi(G_{n_0,n_1})=2k\}}$$

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

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

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},

$$P(\psi(G_{n_0,n_1}) = 2k) = \binom{n_0 + n_1}{n_0}^{-1} \sum_{i=0}^{n_0-1} \sum_{j=0}^{n_1-1} \sum_{C_{i,j} \in \mathbb{C}_{i,j}} \mathbf{1}_{\{\nu(G(C_{i,j}))=k-1\}}$$

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

Using Theorem 20, this can be further simplified to

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

Through repeated use of the identity

$$\sum_{i=a}^{b} \binom{c + i}{d} = \binom{c + b + 1}{d + 1} - \binom{c + a}{d + 1},$$

we arrive at the desired expansion. $\qquad\square$

Corollary 26. *If $G_{n,n}$ is a random difference graph with parts of equal size $n \geqslant 2$, then*

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

4.5 Degeneracy

The 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 over H. That is,

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

So by measuring the degeneracy, we can analyze the amount of clustering in difference graphs.

Proposition 27. *For a difference graph G_{n_0,n_1}, $degen(G_{n_0,n_1}) \geqslant d$ if and only if $K_{d,d} \subseteq G_{n_0,n_1}$.*

Proof. First, suppose that $K_{d,d}$ is an induced subgraph of G_{n_0,n_1}. Then there exists a subgraph of G_{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 G_{n_0,n_1} is at least d. Then there exists some subgraph $H \subseteq G_{n_0,n_1}$ such that the degree of every vertex in H is at least d. As G_{n_0,n_1} is bipartite, H must contain at least d vertices from each part. Therefore G_{n_0,n_1} must contain d zero-vertices and d one-vertices, each with degree at least d.

As 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 every vertex in each set is adjacent to all vertices in the other, $K_{d,d} \subseteq G_{n_0,n_1}$. $\qquad\square$

Theorem 28. *For a random G_{n_0,n_1},*

$$P(degen(G_{n_0,n_1}) = d) = \binom{n_0+n_1}{n_0}^{-1} \binom{n_0}{d}\binom{n_1}{d}$$

Proof. From the above, we see that

$$degen(G_{n_0,n_1}) = \max\{d : K_{d,d} \subseteq G_{n_0,n_1}\}$$

Furthermore, G_{n_0,n_1} contains $K_{d,d}$ if and only if the creation sequence $seq(G_{n_0,n_1})$ contains $0^d 1^d$ as a subsequence. So in order for $degen(G_{n_0,n_1})$ to equal d, $seq(G_{n_0,n_1})$ must contain a substring of the form $0^d 1^d$, but *cannot* contain $0^{d+1}1^{d+1}$.

The requirement that $seq(G_{n_0,n_1})$ contain $0^d 1^d$ means that there must be at least d ones after the dth zero; equivalently, there must be at most $n_1 - d$ ones before the dth zero. The requirement that $seq(G_{n_0,n_1})$ 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 dth zero, we have the range $0 \leqslant i \leqslant n_1 - d$. Letting j denote the number of ones between z_d and z_{d+1}, the requirement $i + j \geqslant n_1 - d$ becomes $n_1 - d - i \leqslant j \leqslant n_1 - i$. Thus,

$$P(degen(G_{n_0,n_1}) = d) =$$

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

which simplifies to

$$P(degen(G_{n_0,n_1}) = d) = \binom{n_0+n_1}{n_0}^{-1} \sum_{i=0}^{n_1-d} \binom{i+d-1}{i}\binom{n_0}{d}$$

$$= \binom{n_0+n_1}{n_0}^{-1} \binom{n_0}{d}\binom{n_1}{d}$$

$$\square$$

References

[1] V. Chvátal and P.L. Hammer. Set-packing problems and threshold graphs. CORR 73-21, University of Waterloo, 1973.

[2] V. Chvátal and P.L. Hammer. Aggregation of inequalities in integer programming. In Hammer *et al.*'s *Studies in Integer Programming*, North-Holland, 1977.

[3] J.-P. Doignon, A. Ducamp, and J.-C. Falmagne. On realizable biorders and the border dimension of a relation. *Journal of Mathematical Psychology*, 28:73-109, 1984.

[4] P. Diaconis, S. Holmes, and S. Janson. Threshold Graph Limits and Random Threshold Graphs. *J. Internet Math.*, **5** no. 3, 267-320, 2010.

[5] P.L. Hammer, U.N. Peled, and X. Sun. Difference Graphs. *Discrete Applied Mathematics* 28:35-44, 1990.

[6] N.V.R. Mahadev and U.N. Peled. *Threshold Graphs and Related Topics*. North-Holland, 1995.

[7] E.P. Reilly and E.R. Scheinerman. Random Threshold Graphs. *The Electronic Journal of Combinatorics*, 16:R130, 2009