Random Threshold Graphs

Elizabeth Perez Reilly
Edward R. Scheinerman
Department of Applied Mathematics and Statistics
Johns Hopkins University
Baltimore, Maryland 21218 USA.

Submitted: Feb 3, 2009; Accepted: Oct 13, 2009; Published: Oct 31, 2009
Mathematics Subject Classifications: 05C62, 05C80

Abstract

We introduce a pair of natural, equivalent models for random threshold graphs and use these models to deduce a variety of properties of random threshold graphs. Specifically, a random threshold graph G is generated by choosing n IID values $x_1, \ldots, x_n$ uniformly in $[0, 1]$; distinct vertices i, j of G are adjacent exactly when $x_i + x_j \geqslant 1$. We examine various properties of random threshold graphs such as chromatic number, algebraic connectivity, and the existence of Hamiltonian cycles and perfect matchings.

1 Introduction and Overview of Results

Threshold graphs were introduced by Chvátal and Hammer in [4, 5]; see also [6, 13]. There are several, logically equivalent ways to define this family of graphs, but the one we choose works well for developing a model of random graphs. A simple graph G is a *threshold graph* if we can assign weights to the vertices such that a pair of distinct vertices is adjacent exactly when the sum of their assigned weights is or exceeds a specified threshold. Without loss of generality, the threshold can be taken to be 1 and the weights can be restricted to lie in the interval $[0, 1]$; see Definition 2.1. References [2, 9, 16] provide an extensive introduction to this class of graphs.

If we choose the weights for the vertices at random, we induce a probability measure on the set of threshold graphs and thereby create a notion of a random threshold graph. Given that we may assume the weights lie in $[0, 1]$ it is natural to take the weights independently and uniformly in that interval; a careful definition is given in §3.1. The idea of choosing a random representation has been explored in other contexts such as random geometric graphs [18] (choose points in a metric space at random to represent vertices that are adjacent if their points are within a specified distance) and random interval graphs [19] (choose real intervals at random to represent vertices that are adjacent if their intervals intersect).

A different approach to random threshold graphs that is based on a recursive description of their structure (see Theorem 2.7) was presented in [11] whose goal was to use threshold graphs

to approximate real-world networks (such as social networks). We use the core idea of [11] to develop a second, alternative model of random threshold graphs (see §3.2).

Our principal result is that these two rather different definitions of random threshold graphs result in precisely the same probability distribution on graphs; this is presented in §3.4 and proved in §4. We then exploit this alternative description of random threshold graphs to deduce various properties of these graphs in §5. In nearly all cases, our results are exact; this stands in stark contrast to the theory of Erdős-Rényi random graphs in which most results are asymptotic. In particular we consider the following properties of random threshold graphs:

- degree and connectivity properties, including the algebraic connectivity;
- the clique and chromatic number;
- Hamiltonicity;
- perfect matchings; and
- statistics on small induced subgraphs and vertices of extreme degree.

For example, we prove that the probability a random threshold graph on n vertices has a Hamiltonian cycle is exactly

$$\frac{1}{2^{n-1}}\binom{n-2}{\lfloor (n-2)/2 \rfloor}$$

which is asymptotic to $1/\sqrt{2\pi n}$; see Theorem 5.21.

2 Threshold Graphs

Most of the definitions and results presented in this section are previously known; see [4] but especially [2, 9, 16] for a broad overview.

2.1 Definitions

The graphs we consider are simple graphs (undirected and without loops or multiple edges). Often the vertex set of G, denoted $V(G)$, is $[n] := \{1, 2, \ldots, n\}$. The edge set of G is denoted $E(G)$.

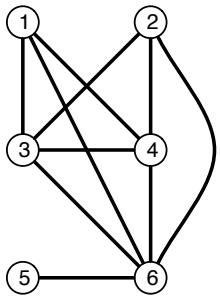
There are a variety of equivalent ways to define threshold graphs; we choose this one as particularly convenient for our purposes.

Definition 2.1 (Threshold graph, representation)**.** Let G be a graph. We say that G is a *threshold graph* provided there is a mapping $f : V(G) \to \mathbb{R}$ such that for all pairs of distinct vertices u, v we have

$$uv \in E(G) \iff f(u) + f(v) \geqslant 1.$$

The mapping f is called a *threshold representation* of f. The number $f(v)$ is called the *weight* assigned to vertex v.

Figure 1 A threshold graph. A representation for this graph is $\mathbf{x} = \left(\frac{1}{2}, \frac{1}{4}, \frac{7}{8}, \frac{15}{16}, \frac{1}{32}, \frac{63}{64}\right)$.



Definition 2.2 (Proper representation)**.** Let G be a threshold graph and let $f : V(G) \to \mathbb{R}_+$ be a threshold representation of G. We say that f is a *proper* representation provided:

1. for all vertices v of G, $0 < f(v) < 1$,
2. for all pairs of distinct vertices u, v of G, $f(u) \neq f(v)$, and
3. for all pairs of distinct vertices u, v of G, $f(u) + f(v) \neq 1$.

The following is well known; see [16].

Proposition 2.3. *Let G be a threshold graph. Then G has a proper threshold representation.* □

Because the graphs we consider have $V(G) = [n]$, a threshold representation $f : V(G) \to \mathbb{R}_+$ can be identified with a vector $\mathbf{x} \in \mathbb{R}^n$ in which the i^{th} coordinate of $\mathbf{x}$, x_i, is $f(i)$. A threshold graph and representation for this graph are shown in Figure 1.

By Proposition 2.3 we may restrict our attention to representing vectors in the following set.

Definition 2.4 (Space of proper representations)**.** Let n be a positive integer. The *space of proper representations* is the set $\mathbb{P}_n$ defined as those vectors $\mathbf{x} \in \mathbb{R}^n$ such that

1. for all i, $0 < x_i < 1$,
2. for all $i \neq j$, $x_i \neq x_j$, and
3. for all $i \neq j$, $x_i + x_j \neq 1$.

Given $\mathbf{x} \in \mathbb{P}_n$ define $\Gamma(\mathbf{x})$ to be the threshold graph G with $V(G) = [n]$ so that $i \mapsto x_i$ is a threshold representation. That is, $ij \in E(G)$ if and only if $x_i + x_j > 1$. Thus Γ is a mapping from $\mathbb{P}_n$ onto the set of threshold graphs on vertex set $[n]$.

We denote the set of threshold graphs with vertex set n as $\mathbb{T}_n$. Therefore $\Gamma : \mathbb{P}_n \to \mathbb{T}_n$.

Note that for a threshold graph G with $V(G) = [n]$, $\Gamma^{-1}(G)$ is the subset of $\mathbb{P}_n$ of all proper representations of G.

2.2 Characterization theorems

See [16] for details on these well-known results.

It is easy to check that the property of being a threshold graph is a hereditary property of graphs. By this we mean

- if G is a threshold graph and H is isomorphic to G, then H is a threshold graph, and
- if G is a threshold graph and H is an induced subgraph of G, then H is a threshold graph.

Therefore, threshold graphs admit a forbidden subgraph characterization; in addition to [16], see also [2].

Theorem 2.5. *[4] Let G be a graph. Then G is a threshold graph if and only if G does not contain an induced subgraph isomorphic to C_4, P_4, or $2K_2$.* □

Of greater utility to us is a structural characterization of threshold graphs based on extremal vertices which we define here.

Definition 2.6. Let G be a graph and let $v \in V(G)$. We say that v is *extremal* provided it is either *isolated* (adjacent to no other vertices of G) or *dominating* (adjacent to all other vertices of G).

Theorem 2.7. *Let G be a graph. Then G is a threshold graph if and only if G has an extremal vertex u and $G - u$ is a threshold graph.*

We include a proof of this well-known result because it is central to the notion of creation sequence developed in section 2.3.

Proof. Suppose first that G is a threshold graph and let $\mathbf{x}$ be a proper threshold representation. Select vertices a and b such that

$$x_a = \min\{x_v : v \in V(G)\} \quad \text{and} \quad x_b = \max\{x_v : v \in V(G)\}.$$

Note that if $x_a + x_b < 1$, then $x_a + x_v < 1$ for all vertices v and so a is an isolated vertex. However, if $x_a + x_b > 1$ then $x_v + x_b > 1$ for all vertices and so b is a dominating vertex. Hence G has an extremal vertex u (either a or b). Furthermore, any induced subgraph of a threshold graph is again a threshold graph, so $G - u$ is threshold.

Conversely, suppose u is an extremal vertex of G and that $G - u$ is a threshold graph. Let $\mathbf{x}$ be a threshold representation of $G - u$. Without loss of generality, we can choose $\mathbf{x}$ so that all weights are strictly between 0 and 1.

Define x_u to be 0 if u is an isolated vertex or to be 1 is u is a dominating vertex. One checks that so augmented, $\mathbf{x}$ is a threshold representation of G, and therefore G is a threshold graph. □

Corollary 2.8. *A graph G is a threshold graph if and only if its complement $\overline{G}$ is a threshold graph.* □

As usual,for a vertex v of a graph G we write $N(v) = \{w \in V(G) : vw \in E(G)\}$ for the set of neighbors of v and $d(v) = |N(v)|$ for the degree of v.

Proposition 2.9. *Let v, w be vertices of a threshold graph G. The following are equivalent:*

1. $d(v) < d(w)$.
2. *In every threshold representation f of G we have $f(v) < f(w)$.* □

Proof. (1) $\Rightarrow$ (2): Suppose $d(v) < d(w)$ and let f be any representation of G. For contradiction, suppose $f(v) \geqslant f(w)$. Choose any vertex $u \neq v, w$. If $u \sim w$ then $f(u) + f(w) \geqslant 1$ which implies $f(v) + f(w) \geqslant 1$ and so $u \sim v$. This implies $d(v) \geqslant d(w)$, a contradiction.

(2) $\Rightarrow$ (1): Suppose in every representation of f of G we have $f(v) < f(w)$. Then, arguing as above, for all $u \neq v, w$, $u \sim v \Rightarrow u \sim w$. This implies $d(v) \leqslant d(w)$. If (for contradiction) we had $d(v) = d(w)$, then for all $u \neq v, w$, $u \sim v \iff u \sim w$. Fix a representation f and define a new function f' by

$$f'(u) = \begin{cases} f(w) & \text{if } x = v, \\ f(v) & \text{if } x = w, \text{ and} \\ f(u) & \text{otherwise.} \end{cases}$$

One checks that f' is also a representation of G but $f'(v) > f'(w)$, a contradiction. □

Proposition 2.10. *Let G be a threshold graph and let $v, w \in V(G)$. The following are equivalent:*

1. $d(v) = d(w)$.
2. $N(v) - w = N(w) - v$.
3. *There is an automorphism of G that fixes all vertices other than v and w and that transposes v and w.*
4. *There is a threshold representation f of G such that $f(v) = f(w)$.* □

Proof. The implications (4) $\Rightarrow$ (3) $\Rightarrow$ (2) $\Rightarrow$ (1) are straightforward, so we are left to argue that (1) $\Rightarrow$ (4). By Proposition 2.9, there are representations f and g of G with $f(v) \leqslant f(w)$ and $g(v) \geqslant g(w)$. Define h by $h(u) = \frac{1}{2}[f(u) + g(u)]$. One checks that h is a representation of G in which $h(v) = h(w)$. □

Vertices v, w that satisfy any (and hence all) of the conditions of Proposition 2.10 are called *twins*.

2.3 Creation sequences

The concept of a creation sequence was developed in [11]. Our definition is a modest modification of their original formulation.

Let G be a threshold graph. Theorem 2.7 implies that G can be constructed as follows. Begin with a single vertex. Iteratively add either an isolated vertex (adjacent to none of the previous vertices) or a dominating vertex (adjacent to all of the previous vertices). We can encode this construction as a sequence of 0s and 1s where 0 represents the addition of an isolated vertex and 1 represents the addition of a dominating vertex.

Definition 2.11 (Creation sequence)**.** Let G be a threshold graph with n vertices. Its *creation sequence* seq(G) is an $n-1$-long sequence of 0s and 1s recursively defined as follows. Let v be an extremal vertex of G. Then seq(G) = seq($G-v$) $\|$ x (here $\|$ represents concatenation) where $x = 0$ if v is isolated and $x = 1$ if v is dominating.

For example, consider the threshold graph G in Figure 1. It has a dominating vertex (6) so the final entry in seq(G) is a 1, i.e., seq(G) = xxxx1. Deleting vertex 6 from G gives a graph with an isolated vertex (5), so seq(G) = xxx01. Deleting that vertex leaves vertex 4 as a dominating vertex. Continuing this way we see seq(G) = 01101.

Note that there is a mild ambiguity in Definition 2.11 in that a threshold graph may have more than one extremal vertex v. One checks, however, that the same creation sequence is generated regardless of which extremal vertex is used to determine the last term of seq(G). The creation sequence of K_1 is the empty sequence.

It is easy to check that for every $n-1$-long sequence s of 0s and 1s, there is a threshold graph G with seq(G) = s. We also have the following.

Proposition 2.12. *Let G and H be threshold graphs. Then $G \cong H$ if and only if* seq(G) = seq(H). □

2.4 Unlabeled graphs

In the sequel we consider both labeled and unlabeled graphs. To deal with these concepts carefully, we include the following discussion.

For us, there is no distinction between the terms *graph* and *labeled graph*.

An *unlabeled graph* is an isomorphism class of graphs, but we define it in a strict way.

Definition 2.13 (Unlabeled graph)**.** Let G be a graph on n vertices. Let $[G]$ denote the set of all graphs on vertex set $[n]$ that are isomorphic to G. We call $[G]$ an *unlabeled* graph.

Since there are only finitely many graphs with vertex set $[n]$, unlabeled graphs are finite sets of (labeled) graphs. Indeed, if the automorphism group of G has cardinality a, then $[G]$ is a set of $n!/a$ graphs.

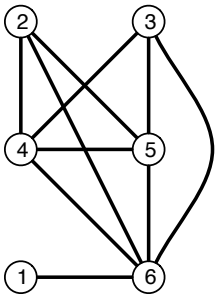
We typically denote labeled graphs with upper case italic letters, G, and unlabeled graphs with upper case bold letters, $\mathbf{G}$.

Let $\mathbf{G}$ be an unlabeled threshold graph. By Proposition 2.12, for all $G, G' \in \mathbf{G}$, we have seq(G) = seq(G'). Therefore, we write seq($\mathbf{G}$) to denote this common sequence.

Proposition 2.14. *[17] Let n be a positive integer. There are 2^{n-1} unlabeled threshold graphs on n vertices.*

Proof. Unlabeled threshold graphs on n vertices are in one-to-one correspondence with $n-1$-long sequences of 0s and 1s. □

Figure 2 The graph from Figure 1 canonically labeled.



2.5 Canonical labeling of threshold graphs

Let $\mathbf{G}$ be an unlabeled threshold graph. It is useful to have a method to select a canonical representative $G \in \mathbf{G}$. We denote the canonical representative of $\mathbf{G}$ by $\ell(\mathbf{G})$ which we define as follows.

Definition 2.15 (Canonical labeling)**.** Let $\mathbf{G}$ be an unlabeled graph. Let $G = \ell(\mathbf{G})$ be the unique graph in $\mathbf{G}$ with the property that

$$\forall v, w \in V(G),\ d_G(v) < d_G(w) \Longrightarrow v < w.$$

In other words, we number sequentially starting with the vertices of lowest degrees working up to the vertices of largest degree.

The uniqueness of $\ell(\mathbf{G})$ follows from Propositions 2.9 and 2.10.

Here is an equivalent description of $\ell(\mathbf{G})$. For a vector $\mathbf{x}$, let $\mathrm{sort}(\mathbf{x})$ be the vector formed from $\mathbf{x}$ by arranging $\mathbf{x}$'s elements in ascending order. Let $\mathbf{x}$ be a proper representation for any graph in $\mathbf{G}$. Then $\ell(\mathbf{G}) = \Gamma(\mathrm{sort}(\mathbf{x}))$. This observation leads to the following result.

Proposition 2.16. *Let* $\mathbf{x}, \mathbf{x}' \in \mathbb{P}_n$ *and suppose* $\Gamma(\mathbf{x}) \cong \Gamma(\mathbf{x}')$. *Let* $\mathbf{y} = \mathrm{sort}(\mathbf{x})$ *and let* $\mathbf{y}' = \mathrm{sort}(\mathbf{x}')$. *Then* $\Gamma(\mathbf{y}) = \Gamma(\mathbf{y}')$. □

For example, let G be the graph in Figure 1. One checks that $\mathbf{x} = \left(\frac{1}{2}, \frac{1}{4}, \frac{7}{8}, \frac{15}{16}, \frac{1}{32}, \frac{63}{64}\right)$ is a proper representation for G. Let $\mathbf{y} = \mathrm{sort}(\mathbf{x}) = \left(\frac{1}{32}, \frac{1}{4}, \frac{1}{2}, \frac{7}{8}, \frac{15}{16}, \frac{63}{64}\right)$ to produce the graph $H = \Gamma(\mathbf{y})$ in Figure 2.

3 Random Models

We now present two models of random threshold graphs. In both cases, a *random threshold graph* on n vertices is a pair $(\mathbb{T}_n, P)$ where P is a probability measure on $\mathbb{T}_n$.

3.1 Random vector model

Let n be a positive integer. A natural way to define a random threshold graph on n vertices is to pick n random numbers independently and uniformly from $[0, 1]$ and use these as the weights.

Equivalently, we pick $\mathbf{x}$ uniformly at random in $[0,1]^n$. Note that with probability 1, $\mathbf{x} \in \mathbb{P}_n$. Let G be the threshold graph $\Gamma(\mathbf{x})$. This leads us to the following formal definition.

Definition 3.1 (Random vector threshold graph)**.** Let n be a positive integer. Define the probability space $(\mathbb{T}_n, P')$ by setting

$$P'(G) = \mu\left(\Gamma^{-1}(G)\right)$$

where $G \in \mathbb{T}_n$ and μ is Lebesgue measure in $\mathbb{R}^n$.

Note: By definition $\Gamma : \mathbb{P}_n \to \mathbb{T}_n$, and so $\Gamma^{-1}(G)$ is a subset of $\mathbb{P}_n$. Observe that $\mu(\mathbb{P}_n) = 1$. Definition 3.1 can be rewritten like this:

$$P'(G) = \mu\{\mathbf{x} \in \mathbb{P}_n : \Gamma(\mathbf{x}) = G\}.$$

Example 3.2. We calculate $P'(G)$ where G is the path on three vertices $1 \sim 2 \sim 3$. To do this we need to find

$$\mu\left\{(x,y,z) \in \mathbb{P}_3 : \Gamma([x,y,z]) = G\right\} = \mu\left\{(x,y,z) \in [0,1]^3 : x+y \geqslant 1, y+z \geqslant 1, x+z < 1\right\}.$$

We break up this calculation into two cases: $x \leqslant z$ and $x \geqslant z$ to get

$$P'(G) = 2\int_{x=0}^{\frac{1}{2}} \int_{z=x}^{1-x} \int_{y=1-x}^{1} dy\, dz\, dx = \frac{1}{12}.$$

(The triple integral is based on the case $x \leqslant z$.)

We define $\mathcal{T}_1$ to be the set $\{(\mathbb{T}_n, P') : n \geqslant 1\}$. We call $\mathcal{T}_1$ the *random vector model* for threshold graphs.

3.2 Random creation sequence model

Our second model of random threshold graphs is based on creation sequences. Let n be a positive integer and let s be an $n-1$-long sequence of 0s and 1s. Define $\gamma(s)$ to be the unlabeled threshold graph $\mathbf{G}$ with $\mathrm{seq}(\mathbf{G}) = s$. In other words,

$$\gamma(s) = \{G \in \mathbb{T}_n : \mathrm{seq}(G) = s\}.$$

Our second model of random threshold graph can be described informally as follows. Let n be a positive integer. Choose a random $n-1$-long sequence of 0s and 1s s; each element of s is an independent fair coin flip; that is, all 2^{n-1} sequences are equally likely. Then randomly apply labels to the unlabeled threshold graph $\gamma(s)$; that is, select a graph uniformly at random from $\gamma(s)$. Here is a formal description.

Definition 3.3 (Random creation sequence threshold graph)**.** Let n be a positive integer. Define the probability space $(\mathbb{T}_n, P'')$ by setting,

$$P''(G) = \frac{1}{2^{n-1} \cdot |[G]|}$$

where $G \in \mathbb{T}_n$.

One checks that

$$\sum_{G \in \mathbb{T}_n} P''(G) = 1.$$

Example 3.4. We calculate $P''(G)$ where G is the path on three vertices $1 \sim 2 \sim 3$. Note that $|[G]| = 3!/|\operatorname{Aut}(G)| = 3!/2 = 3$ and so

$$P''(G) = \frac{1}{2^2|[G]|} = \frac{1}{12}.$$

Note that the calculation of P'' (Example 3.4) is much easier than the calculation of P' (Example 3.2) and gives the same result—a phenomenon that holds in general (Theorem 3.7).

Example 3.5. We calculate P'' for the graph G in Figure 1. Note that $\operatorname{Aut}(G)$ contains exactly four automorphisms as we can independently exchange vertices $1 \leftrightarrow 2$ and $3 \leftrightarrow 4$. Therefore

$$P''(G) = \frac{1}{2^5|[G]|} = \frac{4}{2^5 \cdot 6!} = \frac{1}{5760}.$$

Let $\mathcal{T}_2 = \{(\mathbb{T}_n, P'') : n \geqslant 1\}$. We call $\mathcal{T}_2$ the *random creation sequence model* for threshold graphs.

Note that in this model, the probability that a random threshold graph has a particular creation sequence is $1/2^{n-1}$. Furthermore, all graphs with creation sequence s are equally likely in this model.

3.3 Computing $P''(G)$

As suggested by Examples 3.4 and 3.5, the computation of $P''(G)$ for a threshold graph G is easy.

By Definition 3.3, if G is a threshold graph with vertex set $[n]$, then

$$P''(G) = \frac{1}{2^{n-1}|[G]|}.$$

Of course $|[G]| = n!/|\operatorname{Aut}(G)|$, so this can be rewritten

$$P''(G) = \frac{|\operatorname{Aut}(G)|}{n!2^{n-1}}.$$

For a general graph, the computation of $|\operatorname{Aut}(G)|$ is nontrivial. However, for a threshold graph, it is easy.

Proposition 3.6. *Let G be a threshold graph with n vertices. For $0 \leqslant i \leqslant n-1$, let n_i be the number of vertices of degree i in G. Then*

$$|\operatorname{Aut}(G)| = n_0!n_1!n_2! \cdots n_{n-1}!.$$

Proof. By Proposition 2.10 it follows that every degree-preserving permutation of the vertex set of a threshold graph G is an automorphism of G. Hence $\operatorname{Aut}(G)$ is isomorphic to $S_{n_0} \times S_{n_1} \times \cdots \times S_{n_{n-1}}$, and the result follows. □

3.4 Equivalence of models

Model $\mathcal{T}_1$ is an especially natural way to define threshold graphs—it flows comfortably from the definition of these graphs. Model $\mathcal{T}_2$, however, is more tractable. Fortunately, these two models are equivalent.

Theorem 3.7. *$\mathcal{T}_1 = \mathcal{T}_2$. That is, if G is a threshold graph, then $P'(G) = P''(G)$.*

The proof of this result rests on a geometric analysis (see §4) of the space of proper representations, $\mathbb{P}_n$. Before we present the proof, two comments are in order.

Remark 3.8. The choice of the uniform distribution on $[0, 1]$ for the weights in model $\mathcal{T}_1$ is natural, but other distributions might be considered as well. A close reading of the proof of Theorem 3.7 reveals that replacing the uniform $[0, 1]$ distribution with any continuous distribution that is symmetric about $\frac{1}{2}$ (such as the normal distribution $\mathcal{N}(\frac{1}{2}, 1)$ with mean $\frac{1}{2}$ and variance 1) results in the same model of random threshold graphs.

Remark 3.9. We can maintain the uniform $[0, 1]$ distribution for the vertex weights, but change the threshold for adjacency. Let t be a real number with $0 < t < 2$ and let $\mathbf{x} \in [0, 1]^n$. Define $\Gamma_t(\mathbf{x})$ to be the graph G with vertex set $[n]$ in which ij is an edge exactly when $x_i + x_j \geqslant t$. This gives rise to a model of random threshold graphs $\mathcal{T}_1^t$ generated by choosing the weights uniformly at random in $[0, 1]$. In this model, one can work out that the probability of an edge is

$$P\{ij \in E(G)\} = p = \begin{cases} 1 - \frac{1}{2}t^2 & \text{for } 0 < t \leqslant 1 \text{ and} \\ \frac{1}{2}(2 - t)^2 & \text{for } 1 \leqslant t < 2. \end{cases} \tag{1}$$

In case $t = 1$, this model reduces to $\mathcal{T}_1$.

It is natural to ask if there is an analogue to Theorem 3.7 for the model $\mathcal{T}_1^t$ when $t \neq 1$. Let $\mathcal{T}_2^p$ be the random creation sequence model in which the 0s and 1s of the creation sequence are independent coin tosses, but in which the probability of a 1 is p as given in equation (1).

For $0 < t < 1$, note that the probability of $\overline{K_3}$ in $\mathcal{T}_1^t$ is $\frac{1}{4}t^3$ but in $\mathcal{T}_2^p$ this graph has probability $(1 - p)^2 = \frac{1}{4}t^4$; these are different for all $0 < t < 1$. A similar argument, based on the graph K_3, shows that $\mathcal{T}_1^t \neq \mathcal{T}_2^p$ for $1 < t < 2$.

4 Decomposing $\mathbb{P}_n$ and the Proof of Theorem 3.7

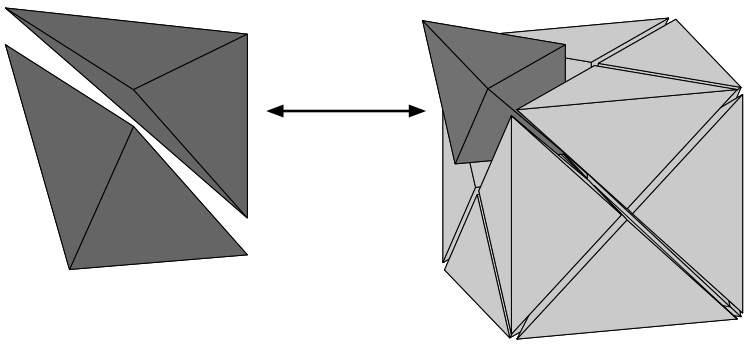
4.1 The regions of $\mathbb{P}_n$

The space of proper representations, $\mathbb{P}_n$, is an open subset of the open cube $(0, 1)^n$. Note that $\mathbb{P}_n$ is dissected into connected regions by slicing the open cube with the following $2\binom{n}{2}$ hyperplanes:

- $\forall i, j \in [n]$ with $i \neq j$, $\Pi_{ij} = \{\mathbf{x} \in (0, 1)^n : x_i = x_j\}$ and
- $\forall i, j \in [n]$ with $i \neq j$, $\Pi'_{ij} = \{\mathbf{x} \in (0, 1)^n : x_i + x_j = 1\}$.

Figure 3 illustrates how $\mathbb{P}_3$ is dissected by these hyperplanes.

Figure 3 The regions of $\mathbb{P}_3$. The left portion of the figure shows two of the 24 connected regions of $\mathbb{P}_3$. The right portion shows how these pieces fit together.



Proposition 4.1. *Let* $\mathbf{x}, \mathbf{x}'$ *be points in the same connected region of* $\mathbb{P}_n$. *Then* $\Gamma(\mathbf{x}) = \Gamma(\mathbf{x}')$.

Proof. Note that for all vertices $i \neq j$, we have $x_i + x_j \neq 1$ and $x'_i + x'_j \neq 1$. Therefore, to establish that $\Gamma(\mathbf{x}) = \Gamma(\mathbf{x}')$, is enough to show

$$\forall i \neq j, \quad x_i + x_j < 1 \iff x'_i + x'_j < 1.$$

But if this were false, then $\mathbf{x}$ and $\mathbf{x}'$ would lie on opposite sides of a hyperplane of the form Π'_{ij}. □

Thus the set of $\mathbf{x} \in \mathbb{P}_n$ that represent a given graph G is a disjoint union of connected regions of $\mathbb{P}_n$.

4.2 Counting the regions

Theorem 4.2. *There are* $2^{n-1}n!$ *connected regions of* $\mathbb{P}_n$. *Moreover, there is a bijection between the set of regions of* $\mathbb{P}_n$ *and the set of ordered pairs* $(\mathbf{G}, \pi)$ *where* $\mathbf{G}$ *is an unlabeled threshold graph on n vertices and* $\pi \in S_n$, *i.e., a permutation of* $[n]$.

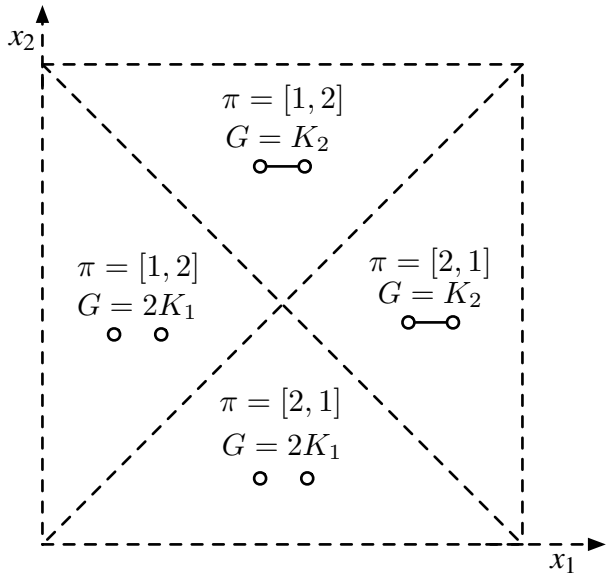
For $n = 1, 2, 3, 4, \ldots$, the number of regions is $1, 4, 24, 192, \ldots$; this is sequence A002866 in [21].

Proof. We establish a bijection between connected regions of $\mathbb{P}_n$ and the set of ordered pairs $(\mathbf{G}, \pi)$ where $\mathbf{G}$ is an unlabeled threshold graph on n vertices and $\pi \in S_n$. The result then follows from Proposition 2.14.

Let R be a region of $\mathbb{P}_n$ and let $\mathbf{x} \in R$. First, to $\mathbf{x}$ we associate a permutation π so that

$$(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) = \mathrm{sort}(\mathbf{x}).$$

Figure 4 The four regions of $\mathbb{P}_2$ corresponding to all ordered pairs $(\pi, \mathbf{G})$ where $\pi \in S_2$ and $\mathbf{G}$ is an unlabeled threshold graph on two vertices.



This unambiguously defines π because no two components of $\mathbf{x}$ are equal. Furthermore, if $\mathbf{x}, \mathbf{x}'$ are distinct points of R, they determine the same permutation. [Otherwise, we have, say $x_i < x_j$ and $x'_i > x'_j$ placing the points on opposite sides of the hyperplane Π_{ij}, a contradiction.] Thus we may associate this permutation with the entire region and refer to it as π_R.

Next, to a point $\mathbf{x} \in R$ we associate the unlabeled graph $[\Gamma(\mathbf{x})]$. Furthermore, given two points $\mathbf{x}$ and $\mathbf{x}'$ of R, note that $\Gamma(\mathbf{x}) = \Gamma(\mathbf{x}')$. [Otherwise, we have, say, $ij \in E[\Gamma(\mathbf{x})]$ but $ij \notin E[\Gamma(\mathbf{x}')]$. This gives $x_i + x_j > 1$ and $x'_i + x'_j < 1$, placing the points on opposite sides of the hyperplane Π'_{ij}.$\Rightarrow\Leftarrow$] Thus, all points $\mathbf{x}$ in R yield the same graph G and *a fortiori*, the same unlabeled graph $[\Gamma(\mathbf{x})]$. We call this graph $\mathbf{G}_R$.

Hence the mapping $R \mapsto (\mathbf{G}_R, \pi_R)$ is well defined. We claim that this mapping is a bijection. For example, see Figure 4 for the simple case $n = 2$.

We first show that $R \mapsto (\mathbf{G}_R, \pi_R)$ is surjective. Let $\mathbf{G}$ be any unlabeled threshold graph on n vertices and let π be any permutation in S_n.

Choose any $G \in \mathbf{G}$ and let $\mathbf{y}$ be a proper representation of G. Rearrange the coordinates of $\mathbf{y}$ to give $\mathbf{x}$ subject to the condition that $x_{\pi(1)} < x_{\pi(2)} < \cdots < x_{\pi(n)}$. Let R be the region that contains $\mathbf{x}$. Note that $\Gamma(\mathbf{x}) \cong \Gamma(\mathbf{y})$ and so $\mathbf{G}_R = [\Gamma(\mathbf{x})] = [\Gamma(\mathbf{y})] = \mathbf{G}$. In addition, $\mathbf{x}$ was constructed so that

$$(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) = \text{sort}(\mathbf{x})$$

and so $\pi_R = \pi$.

Finally, we need to show that $R \mapsto (\mathbf{G}_R, \pi_R)$ is injective. Let R, R' be distinct regions of $\mathbb{P}_n$, and choose $\mathbf{x} \in R$ and $\mathbf{x}' \in R'$. If $\pi_R \neq \pi_{R'}$ then we are done, so suppose $\pi_R = \pi_{R'}$. Since $\mathbf{x}$ and $\mathbf{x}'$ are from different regions, there exist $i \neq j$ so that (without loss of generality) $x_i + x_j < 1$ but

$x'_i + x'_j > 1$. Therefore $\Gamma(\mathbf{x}) \neq \Gamma(\mathbf{x}')$. Since $\mathbf{G}_R = [\Gamma(\mathbf{x})]$ and $\mathbf{G}_{R'} = [\Gamma(\mathbf{x}')]$ it suffices to show that $\Gamma(\mathbf{x}) \ncong \Gamma(\mathbf{x}')$.

Suppose, for contradiction, that $\Gamma(\mathbf{x}) \cong \Gamma(\mathbf{x}')$. Then, $\mathbf{G} = [\Gamma(\mathbf{x})] = [\Gamma(\mathbf{x}')]$. Let $\ell(\mathbf{G})$ be the canonical labeling of $\mathbf{G}$. By definition, we have $\ell(\mathbf{G}) = \Gamma(\mathrm{sort}(\mathbf{x})) = \Gamma(\mathrm{sort}(\mathbf{x}'))$. Define $\mathbf{y} = \mathrm{sort}(\mathbf{x})$ and $\mathbf{y}' = \mathrm{sort}(\mathbf{x}')$. Because $\Gamma(\mathbf{y}) = \Gamma(\mathbf{y}')$, we see that $y_i + y_j > 1$ if and only if $y'_i + y'_j > 1$. By our earlier assumption that $\pi_R = \pi_{R'}$, we know that $y_i = x_{\pi_R(i)}$ and $y'_i = x'_{\pi_R(i)}$. Thus we have $x_{\pi_R(i)} + x_{\pi_R(j)} > 1$ if and only if $x'_{\pi_R(i)} + x'_{\pi_R(j)} > 1$ implying that $\mathbf{x}$ and $\mathbf{x}'$ admit the same threshold graph. This is a contradiction. Therefore, we conclude that $\Gamma(\mathbf{x}) \ncong \Gamma(\mathbf{x}')$ and $R \mapsto (\mathbf{G}_R, \pi_R)$ is injective. □

Definition 4.3. Let n be a positive integer. Let $\mathbf{G}$ be an unlabeled threshold graph and let $\pi \in S_n$. Define $R(\mathbf{G}, \pi)$ to denote the connected region of $\mathbb{P}_n$ corresponding to the ordered pair $(\mathbf{G}, \pi)$ given by the bijection in the proof of Theorem 4.2.

4.3 Congruence of the regions

We have established that $\mathbb{P}_n$ decomposes into $2^{n-1}n!$ regions, and each region R is uniquely associated with an ordered pair $(\mathbf{G}_R, \pi_R)$. Our next goal is to establish that these regions all have the same shape, and hence the same n-dimensional volume: $1/(2^{n-1}n!)$.

Theorem 4.4. *All regions of $\mathbb{P}_n$ are congruent and therefore have the same n-dimensional volume.*

Proof. To show that the $n!2^{n-1}$ regions of $\mathbb{P}_n$ are congruent we perform the following transformation:

$$\mathbf{x} \mapsto \hat{\mathbf{x}} := \mathbf{x} - \tfrac{1}{2}\mathbf{1}$$

where $\mathbf{1}$ is a vector of all ones. This translates the cube whose corners are $\{0,1\}^n$ to the cube whose corners are $\{-\frac{1}{2}, \frac{1}{2}\}^n$.

The hyperplanes $x_i = x_j$ and $x_i + x_j = 1$ are transformed as follows:

$$x_i = x_j \quad \mapsto \quad \hat{x}_i + \frac{1}{2} = \hat{x}_j + \frac{1}{2} \quad \Longrightarrow \quad \hat{x}_i = \hat{x}_j$$

and

$$x_i + x_j = 1 \quad \mapsto \quad \left(\hat{x}_i + \frac{1}{2}\right) + \left(\hat{x}_j + \frac{1}{2}\right) = 1 \quad \Longrightarrow \quad \hat{x}_i = -\hat{x}_j$$

Thus the translated $\mathbb{P}_n$ now centered at the origin is dissected by the $2\binom{n}{2}$ hyperplanes $x_i = \pm x_j$. By symmetry, all the regions have the same shape, and therefore the same n-dimensional volume. □

Corollary 4.5. *Let R be a connected region of $\mathbb{P}_n$. Then*

$$\mu(R) = \frac{1}{2^{n-1}n!}.$$

Proof. From Theorem 4.4 we deduce that all regions R have the same n-dimensional volume. Since by Theorem 4.2 there are $2^{n-1}n!$ regions and $\mu(\mathbb{P}_n) = 1$, the result follows. □

4.4 Proof of $\mathcal{T}_1 = \mathcal{T}_2$

Proof of Theorem 3.7. Let $G \in \mathbb{T}_n$ be a threshold graph. We must show that $P'(G) = P''(G)$.

Recall (Definition 3.1) that $P'(G)$ is the measure of the set $\{\mathbf{x} \in \mathbb{P}_n : \Gamma(x) = G\}$. This set is the disjoint union of regions whose points represent G (see Proposition 4.1).

Let $\mathcal{R}_G$ denote the set of regions $R \subset \mathbb{P}_n$ such that $\mathbf{x} \in R \Longrightarrow \Gamma(\mathbf{x}) = G$. Then

$$P'(G) = \frac{|\mathcal{R}_G|}{n!2^{n-1}}$$

because every region in $\mathcal{R}_G$ has the same volume (Corollary 4.5).

Recall (Section 3.3) that

$$P''(G) = \frac{|\operatorname{Aut}(G)|}{n!2^{n-1}}$$

the result follows once we establish $|\mathcal{R}_G| = |\operatorname{Aut}(G)|$.

Let $\mathbf{G} = [G]$ be the unlabeled version of G.

Claim 1. *Let $R(\mathbf{G}, \pi) \in \mathcal{R}_G$ and let $\sigma \in \operatorname{Aut}(G)$. Then $R(\mathbf{G}, \pi \circ \sigma) \in \mathcal{R}_G$.*

> *Proof.* By Theorem 4.2, there is a bijection between regions, R, and unlabeled graph and permutation pairs, $(\mathbf{G}, \pi)$. Thus, it follows that $R(\mathbf{G}, \pi \circ \sigma) \in \mathbb{P}_n$.
>
> It is clear that $R(\mathbf{G}, \pi)$ and $R(\mathbf{G}, \pi \circ \sigma)$ correspond to isomorphic graphs. By Proposition 2.16, they have the same canonical labeling $\ell(\mathbf{G})$. To obtain the graph $G = \Gamma(R(\mathbf{G}, \pi))$, we apply the isomorphism π^{-1} to $\ell(\mathbf{G})$. Similarly, to obtain the graph $G' = \Gamma(R(\mathbf{G}, \pi \circ \sigma))$, we apply the isomorphism $(\pi \circ \sigma)^{-1}$ to $\ell(\mathbf{G})$.
>
> Because σ is an automorphism of G (and therefore so is σ^{-1}), we obtain the same graph, G, after applying σ^{-1} to G. In other words, by first applying π^{-1} to $\ell(\mathbf{G})$ and then applying σ^{-1} to the result, we obtain the same graph G as we would by simply applying π^{-1} to $\ell(\mathbf{G})$. However, applying π^{-1} and then σ^{-1} is equivalent to applying $(\pi \circ \sigma)^{-1}$ to $\ell(\mathbf{G})$ which results in G' as defined above. Thus, $G' = G$ and $R(\mathbf{G}, \pi \circ \sigma) \in \mathcal{R}_G$. □

Claim 2. *Let $R(\mathbf{G}, \pi), R(\mathbf{G}, \sigma) \in \mathcal{R}_G$. Then $\pi^{-1} \circ \sigma \in \operatorname{Aut}(G)$.*

> *Proof.* Let $\ell(\mathbf{G})$ be the canonical labeling of G. Notice that σ is the isomorphism that takes us from $\Gamma(R(\mathbf{G}, \sigma))$ to $\ell(\mathbf{G})$ and π^{-1} is the isomorphism that takes us from $\ell(\mathbf{G})$ to $\Gamma(R(\mathbf{G}, \pi))$. Thus, $\pi^{-1} \circ \sigma$ is an isomorphism from $\Gamma(R(\mathbf{G}, \sigma))$ to $\Gamma(R(\mathbf{G}, \pi))$. Since $R(\mathbf{G}, \pi), R(\mathbf{G}, \sigma) \in \mathcal{R}_G$, we have $G = \Gamma(R(\mathbf{G}, \sigma)) = \Gamma(R(\mathbf{G}, \pi))$. Therefore, $\pi^{-1} \circ \sigma$ is, in fact, an automorphism of G. □

Let $R(\mathbf{G}, \pi) \in \mathcal{R}_G$. The claims show that every region of $\mathcal{R}_G$ is precisely of the form $R(\mathbf{G}, \pi \circ \sigma)$ for some $\sigma \in \operatorname{Aut}(G)$. Therefore $|\mathcal{R}_G| = |\operatorname{Aut}(G)|$, completing the proof of Theorem 3.7. □

5 Properties of Random Threshold Graphs

Having established the equivalence of models $\mathcal{T}_1$ and $\mathcal{T}_2$, we drop the subscripts and simply call these *random threshold graphs*. Furthermore, we now write $\Pr(G)$ to denote the probability of a graph G in this common model.

The bits of a creation sequence s are denoted $s_1 s_2 \dots s_{n-1}$. If $s = s_1 s_2 \dots s_{n-1}$, we define $\overline{s} = \overline{s}_1 \overline{s}_2 \dots \overline{s}_{n-1}$ to be the *complement* of s. That is, $\overline{s}_i = 1 - s_i$. The following is easy to establish.

Proposition 5.1. *Let G be a threshold graph. If $s = \mathrm{seq}(G)$, then $\overline{s} = \mathrm{seq}(\overline{G})$ where $\overline{G}$ is the complement of G.* □

Corollary 5.2. *Let G be a threshold graph. Then,* $\Pr\{G\} = \Pr\{\overline{G}\}$.

Proof. Notice that $\mathrm{seq}(G)$ and $\overline{\mathrm{seq}(G)}$ are equally likely to occur. The result follows by Proposition 5.1. □

5.1 Degree and connectivity properties

Proposition 5.3. *Let G be an instance of a random threshold graph. Then,*

$$\Pr\{G \text{ is connected}\} = \frac{1}{2}.$$

Proof. G is connected if and only if the last bit of $\mathrm{seq}(G)$ is 1, and that occurs with probability $\frac{1}{2}$. □

Proposition 5.4. *Let G be an instance of a random threshold graph on n vertices. Then, the maximum degree of G has the following distribution:*

$$\Pr\{\Delta(G) = i\} = \begin{cases} 1/2^{n-1} & \text{for } i = 0, \\ 1/2^{n-i} & \text{for } 1 \leqslant i \leqslant n-1, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. First, notice that $\Delta(G) = 0$ if and only if $s_i = 0$ for all $1 \leqslant i \leqslant n-1$. So, $\Pr\{\Delta(G) = 0\} = 1/2^{n-1}$. For $1 \leqslant i \leqslant n-1$, $\Delta(G) = i$ if and only if $s_i = 1$ and $s_j = 0$ for all $j > i$. Thus, $\Pr\{\Delta(G) = i\} = \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right)^{n-1-i} = \left(\frac{1}{2}\right)^{n-i}$. □

Proposition 5.5. *Let G be an instance of a random threshold graph on n vertices. Then, the expected maximum degree of G is $E[\Delta(G)] = n - 2 + \frac{1}{2^{n-1}}$.*

Proof. Using Proposition 5.4,

$$\begin{aligned} E[\Delta(G)] &= 0 \cdot \frac{1}{2^{n-1}} + 1 \cdot \frac{1}{2^{n-1}} + 2 \cdot \frac{1}{2^{n-2}} + \cdots + (n-1) \cdot \frac{1}{2} \\ &= \sum_{i=1}^{n-1} \frac{i}{2^{n-i}} = n - 2 + \frac{1}{2^{n-1}}. \end{aligned}$$

□

Corollary 5.6. *Let G be an instance of a random threshold graph on n vertices. Then,*

$$\Pr\{\delta(G) = i\} = \begin{cases} 1/2^{n-1} & \text{for } i = n-1, \\ 1/2^{i+1} & \text{for } 0 \leqslant i \leqslant n-2, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Recall that $\delta(G) = n - 1 - \Delta(\overline{G})$. Thus,

$$\Pr\{\delta(G) = i\} = \Pr\{n - 1 - \Delta(\overline{G}) = i\} = \Pr\{\Delta(\overline{G}) = n - 1 - i\}.$$

The result then follows from Proposition 5.4 and Corollary 5.2. □

Corollary 5.7. *Let G be an instance of a random threshold graph. Then, $E[\delta(G)] = 1 - \frac{1}{2^{n-1}}$.*

Proof. The result follows from the fact that $\delta(G) = n - 1 - \Delta(\overline{G})$ and Proposition 5.5. □

Let G be a graph with n vertices. The *Laplacian* of G, denoted $L(G)$, is an $n \times n$-matrix defined by $L(G) = D(G) - A(G)$ where $D(G)$ is the diagonal matrix of G's degrees and $A(G)$ is G's adjacency matrix. In other words, taking $V(G) = [n]$ we have

$$D(G)_{ij} = \begin{cases} d(i) & \text{when } i = j, \\ -1 & \text{when } ij \in E(G), \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

The matrix $L(G)$ is positive semidefinite and with spectrum

$$0 = \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n.$$

The second smallest eigenvalue, λ_2, is known as the graph's *algebraic connectivity*.

Note that $\lambda_2 > 0$ if and only if the graph is connected.

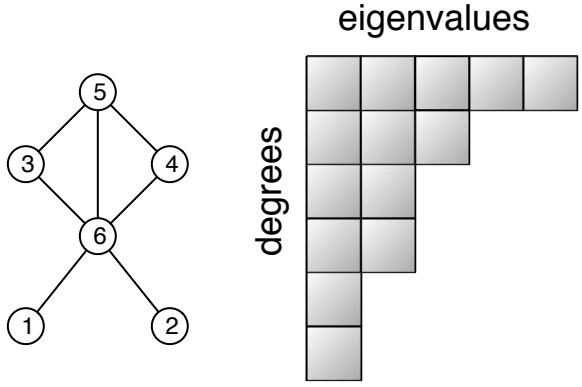
There is a beautiful relation between the eigenvalues of $L(G)$ and the degree sequence of G for threshold graphs due to Merris [10]. Merris observed that the eigenvalues of a threshold graph's Laplacian are all integers. Furthermore, considering the trace of $L(G)$ gives

$$\sum_{i=2}^{n} \lambda_i = \operatorname{tr}[L(G)] = \sum_{j=0}^{n} d(j) = 2|E(G)|.$$

Thus, the eigenvalues of $L(G)$ and the degrees of G are partitions of the same integer. Moreover, Merris proved the following relationship between these partitions.

Theorem 5.8. *Let G be a connected threshold graph, let $0 < d_1 \leqslant d_2 \leqslant \cdots \leqslant d_n$ be the degrees of its vertices and let $0 < \lambda_2 \leqslant \lambda_3 \leqslant \cdots \leqslant \lambda_n$ be the nonzero eigenvalues of G's Laplacian matrix. Then the sequences $(d_n, \ldots, d_1)$ and $(\lambda_n, \ldots, \lambda_2)$ are Ferrer's conjugates of each other.* □

Figure 5 The degree sequence and the nonzero Laplacian eigenvalues of a threshold graph G are conjugate partitions of $2|E(G)|$. In this example, the degrees of the vertices are $(5, 3, 2, 2, 1, 1)$ and the nonzero eigenvalues of $L(G)$ are $(6, 4, 2, 1, 1)$.



For example, see the graph in Figure 5. The degrees of the vertices are $(5, 3, 2, 2, 1, 1)$ which is conjugate to the nonzero eigenvalues of the graph's Laplacian: $(6, 4, 2, 1, 1)$.

Corollary 5.9. *Let G be a threshold graph that is not a complete graph. Then its algebraic connectivity equals its minimum degree, i.e., $\lambda_2(G) = \delta(G)$.*

Note that $\lambda_2(K_n) = n$ but $\delta(K_n) = n - 1$.

Proof. Let $G \neq K_n$ be a threshold graph on n vertices and let $0 = \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ be the eigenvalues of its Laplacian.

If G is not connected, then $\delta(G) = \lambda_2(G) = 0$.

Otherwise, G is connected and let $s = \text{seq}(G)$. Because G is not complete, s contains at least one zero. The vertex of smallest degree corresponds to the last zero in s. Its degree is the number of 1s to its right, which is the number of vertices of maximum degree. Since there are δ vertices of maximum degree, the last column in the Ferrer's conjugate has δ boxes, and so $\lambda_2 = \delta$. □

Corollary 5.10. *Let G be an instance of a random threshold graph on n vertices. Then*

$$\Pr\{\lambda_2(G) = i\} = \begin{cases} 1/2^{n-1} & \text{for } i = n, \\ 1/2^{i+1} & \text{for } 0 \leqslant i \leqslant n-2, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

In particular $E[\lambda_2] = 1$.

Proof. Immediate from Corollaries 5.6 and 5.9 and the fact that $\lambda_2(K_n) = n$. □

We can also deduce from Theorem 5.8 that the largest eigenvalue of a threshold graph G equals $|V(G)| - i(G)$ where $i(G)$ is the number of isolated vertices in G whose distribution is given in Proposition 5.26.

Another degree property that can be readily deduced from the creation sequence is the number of distinct degrees in a threshold graph.

Proposition 5.11. *Let G be a threshold graph and let $s = \mathrm{seq}(G)$ be its creation sequence. The number of contiguous blocks of* 1*s and* 0*s equals the number of different degrees in G.*

Proof. If $\mathrm{seq}(G)$ is entirely 0s or 1s, then the graph is either edgeless or complete, respectively. In either case, all vertices have the same degree.

Otherwise s consists of alternating blocks of 0s and 1s. Note that all vertices within a contiguous run have the same degree. Furthermore, the one vertex that does not correspond to an entry in s has the same degree as the vertices in the first block. □

Proposition 5.12. *Let G be an instance of a random threshold graph on n vertices and let g denote the number of distinct degrees in G. Then, for $1 \leqslant i \leqslant n-1$ we have*

$$\Pr\{g = i\} = \frac{1}{2^{n-2}}\binom{n-2}{i-1}.$$

Proof. We count the number of creation sequences the i runs. The first bit can be either zero or one (2 choices). After that, we select $i-1$ locations from the $n-2$ "spaces" between the bits to show where a block of 1s changes to 0s and vice versa. Hence there are $2\binom{n-2}{i-1}$ creation sequences with i runs, and the result follows. □

It follows that the expected number of distinct degrees in a random threshold graph on n vertices is

$$E[g] = \frac{1}{2^{n-2}} \sum_{i=1}^{n-1} i\binom{n-2}{i-1} = \frac{n}{2}.$$

5.2 Chromatic number

Because threshold graphs are perfect (see, for example, [9]) we can deduce information about the chromatic number from the clique number which is, in turn, directly available from the creation sequence.

Proposition 5.13. *Let G be an instance of a random threshold graph on $n \geqslant 1$ vertices. Then, the chromatic number and the clique number of G have the following distribution with support $[n]$:*

$$\Pr\{\chi(G) = k\} = \Pr\{\omega(G) = k\} = \binom{n-1}{k-1}/2^{n-1} \quad \text{for } 1 \leqslant k \leqslant n.$$

Proof. Threshold graphs are perfect. Therefore, the chromatic number is the size of the maximum clique of the graph. However, the size of the maximum clique is one more than the number of 1s in the creation sequence. This implies that for $1 \leqslant k \leqslant n$, $\Pr\{\chi(G) = k\} = \Pr\{\omega(G) = k\} = \binom{n-1}{k-1}/2^{n-1}$. □

Corollary 5.14. *Let G be an instance of a random threshold graph on $n \geqslant 1$ vertices. Then, the independence number of G has the following distribution with support $[n]$:*

$$\Pr\{\alpha(G) = k\} = \binom{n-1}{k-1}/2^{n-1} \quad \text{for } 1 \leqslant k \leqslant n.$$

Proof. This follows from the fact that $\alpha(G) = \omega(\overline{G})$. □

Proposition 5.15. *Let G be an instance of a random threshold graph. Then, the expected chromatic number of G, and thus the expected clique number of G, is $\frac{n+1}{2}$.*

Proof. By Proposition 5.13,

$$E[\chi(G)] = \frac{1}{2^{n-1}} \sum_{k=1}^{n} k\binom{n-1}{k-1} = \frac{n+1}{2}.$$ □

Corollary 5.16. *Let G be an instance of a random threshold graph. Then, the expected independence number of G is $\frac{n+1}{2}$.*

Proof. Apply Proposition 5.15 and the fact that $\alpha(G) = \omega(\overline{G})$. □

5.3 Cycles

Proposition 5.17. *Let G be an instance of a random threshold graph on n vertices. Then,*

$$\Pr\{G \text{ is acyclic}\} = \frac{n}{2^{n-1}}.$$

Proof. Let $s = \text{seq}(G)$. Because G is a threshold graph, then by Theorem 2.5, it cannot contain C_4 as an induced subgraph. Thus, G contains a cycle if and only if it contains K_3 as an induced subgraph. However, this occurs if and only if there are at least two 1s in s.

Thus,

$$\Pr\{G \text{ is acyclic}\} = \Pr\{s \text{ has at most one } 1\} = \frac{\binom{n-1}{0} + \binom{n-1}{1}}{2^{n-1}} = \frac{n}{2^{n-1}}.$$ □

Corollary 5.18. *Let G be an instance of a random threshold graph on n vertices. Then, the probability G has a cycle is $1 - n/2^{n-1}$.* □

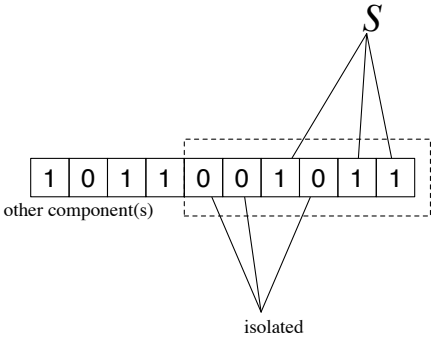
Notice that, as n goes to infinity, the probability that G has a cycle goes to 1.

Next, we consider the probability that a random threshold graph is Hamiltonian. There is a nice connection between Hamiltonicity and a threshold graph's creation sequence. For more background on Hamiltonian threshold graphs, see [12].

For a sequence s of 1s and 0s, let $u_k(s)$ be the number of 1s in the last k bits and $z_k(s)$ be the number of 0s in the last k bits.

Definition 5.19. Let G be a graph and $c(G)$ denote the number of connected components of G. We say that G is *tough* if for every nonempty subset $S \subseteq V(G)$ we have $c(G - S) \leqslant |S|$.

Figure 6 The (b)⇒(c) case for Theorem 5.20: toughness implies the strict partial Dyck property. At some point k, we have $u_k(s) \leqslant z_k(s)$ (illustrated by the dotted box). If S is the set of vertices corresponding to the 1s in the box, then $c(G - S) > |S|$.



Note that a tough graph with three or more vertices must be connected.

Theorem 5.20. *Let G be a threshold graph with $n \geqslant 3$ vertices. The following are equivalent:*

(a) G is Hamiltonian.

(b) G is tough.

(c) If $s = \mathrm{seq}(G)$, then $u_k(s) > z_k(s)$ for all $1 \leqslant k \leqslant n-1$.

The condition $u_k(s) > z_k(s)$ for all k means that the reversal of s (i.e., $s_{n-1}s_{n-2}\ldots s_1$) satisfies the strict partial Dyck property; see Appendix A.

Proof. (a) ⇒ (b): This is well-known.

(b) ⇒ (c): Suppose G is tough. Label the vertices of G by the integers 0 through $n-1$ so that vertex i (with $i > 0$) corresponds to the i^{th} bit of $s = \mathrm{seq}(G)$.

Suppose, for contradiction, there is an index k so that $u_k(s) \leqslant z_k(s)$. Let S be the set of those vertices corresponding to 1s in the last k bits of s. Note that if we delete S from G, the resulting graph has at least $z_k(s) + 1$ components: the component of $G - S$ containing vertex 0 and the $z_k(s)$ isolated vertices. This is illustrated in Figure 6. It follows that

$$c(G - S) \geqslant z_k(s) + 1 > z_k(s) \geqslant |S|$$

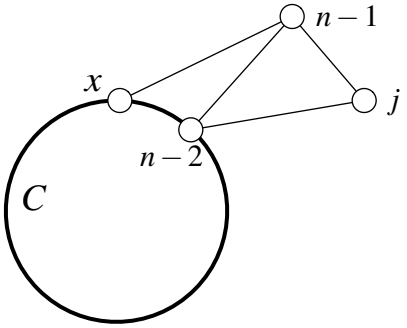
contradicting the fact that G is tough.

(c) ⇒ (a): Suppose that $s = \mathrm{seq}(G)$ satisfies $u_k(s) > z_k(s)$ for all k with $1 \leqslant k \leqslant n-1$. This implies that the last two bits of s are both 1s.

We prove that G is Hamiltonian by induction on the number of vertices, n.

In case $n = 3$, then $\mathrm{seq}(G) = 11$ and so $G = K_3$ which is Hamiltonian. In case $n = 4$, then $\mathrm{seq}(G) = 111$ or 011. In the first case $G = K_4$ and in the second case $G = K_4 - e$, both of which are Hamiltonian.

Figure 7 Induction step in (c)⇒(a).



We now assume the theorem has been shown for all graphs with fewer than n vertices (where we may assume $n \geqslant 5$), and let G be a threshold graph with n vertices that satisfies condition (c).

Without loss of generality, we assume the vertices of G are numbered from 0 to $n-1$ corresponding to their position in the creation sequence $s = \text{seq}(G)$. If s does not contain any zeros, then $G = K_n$ which is Hamiltonian. Otherwise, let j be the index of the last 0 in s; note that $j < n-2$.

Let H be the graph formed by deleting vertices j and $n-1$ from G. Observe that H is a threshold graph whose creation sequence is formed from s by deleting bits j and $n-1$. One checks that H's creation sequence satisfies property (c) and so, by induction, H is Hamiltonian.

Fix a Hamiltonian cycle C of H and let x be a vertex of H that is adjacent to vertex $n-2$ on the cycle C. (See Figure 7.) Note that because the last two bits of s are 1s, vertices $n-2$ and $n-1$ are adjacent to both j and x. Thus, if we delete the edge $\{x, n-2\}$ from C and insert the path $x \sim n-1 \sim j \sim n-2$ in its stead, we create a Hamiltonian cycle in G. □

Theorem 5.21. *Let G be an instance of a random threshold graph with $n \geqslant 3$ vertices. Then*

$$\Pr\{G \text{ is Hamiltonian}\} = \frac{1}{2^{n-1}}\binom{n-2}{\lfloor (n-2)/2 \rfloor} \sim \frac{1}{\sqrt{2\pi n}}.$$

Proof. The number of sequences of length $n-1$ that satisfy condition (c) of Theorem 5.20 is $\binom{n-2}{\lfloor (n-2)/2 \rfloor}$. This is shown in Proposition A.2. The asymptotic value follows from a routine application of Stirling's formula. □

5.4 Perfect matchings

The existence of a perfect matching in a threshold graph is equivalent to a condition that is similar to that for a Hamiltonian cycle. Recall that for a sequence s of 1s and 0s that $u_k(s)$ and $z_k(s)$ denote the number of 1s and 0s, respectively, in the last k bits of s. We have the following result that is analogous to Theorem 5.20.

Theorem 5.22. *Let G be a threshold graph on n vertices with n even and let $s = \text{seq}(G)$. Then G has a perfect matching if and only if $u_k(s) \geqslant z_k(s)$ for all k.*

Proof. First, suppose that for some k, $u_k(s) < z_k(s)$. Let S be the set of vertices corresponding to the 1s in the last k bits of s, so $|S| = u_k(s)$. Note that $G - S$ contains $z_k(s)$ isolated vertices plus (perhaps) other odd components. Therefore, by Tutte's theorem G does not have a perfect matching.

Conversely, suppose that for all k, $u_k(s) \geqslant z_k(s)$. We assume that the vertices $V(G)$ are numbered from 0 to $n-1$ with vertex $i > 0$ corresponding to the i^{th} bit in s. Let

$$U = \{v : s_v = 1\} \cup \{0\} \quad \text{and} \quad Z = \{v : s_v = 0\}.$$

Note that U is a clique and Z is an independent set.

Claim. *G contains a matching M of edges between U and Z that saturates Z.*

> *Proof of claim.* Consider the bipartite graph G' consisting of all vertices of G and all edges of G with one end in Z and the other in U. Let $Y \subseteq Z$. The set of neighbors of Y, $N(Y) = \{u \in U : u \sim y \; \exists y \in Y\}$, corresponds exactly to the set of all 1s in s to the right of position y for $y \in Y$. Since $u_k(s) \geqslant z_k(s)$ for all k, we have $|N(Y)| \geqslant |Y|$ for all $Y \subseteq Z$. Therefore, by Hall's theorem, G' has a matching M that saturates Z. □

Finally, we can extend M to a perfect matching since all vertices unsaturated by M (which are necessarily even in number) lie in the clique U. □

Theorem 5.23. *Let n be an even integer and let G be an instance of a random threshold graph on n vertices. Then*

$$\Pr\{G \text{ has a perfect matching}\} = \frac{1}{2^{n-1}}\binom{n-1}{\lfloor (n-1)/2 \rfloor} \sim \sqrt{\frac{2}{\pi n}}.$$

Proof. From Theorem 5.22, G has a perfect matching if and only if $s = \text{seq}(G)$ is the reverse of a partial Dyck sequence of length $n-1$ (see Appendix A). By Proposition A.1 there are $\binom{n-1}{\lfloor (n-1)/2 \rfloor}$ such sequences. The asymptotic expression follows from Stirling's formula. □

5.5 Edges and extremal vertices

Proposition 5.24. *Let G be an instance of a random threshold graph on n vertices, m denote the number of edges of G, and $Q(k, \ell)$ denote the number of partitions of k into distinct parts whose largest part is less than or equal to ℓ. Then, for $0 \leqslant i \leqslant \binom{n}{2}$, we have that $\Pr\{m = i\} = Q(i, n-1)/2^{n-1}$.*

Proof. Let $s = \text{seq}(G)$. Then,

$$m = \sum_{j=1}^{n-1} s_j \cdot j. \tag{2}$$

Thus, a creation sequence results in a graph with i edges whenever i can be written as the sum of distinct integers between 1 and $n-1$. There are $Q(i, n-1)$ ways to do this and 2^{n-1} creation sequences total. The result follows. □

Proposition 5.25. *Let G be an instance of a random threshold graph. Then, the expected number of edges of G is $E[m] = \frac{1}{2}\binom{n}{2}$. Also, the variance of the number of edges is* $\mathrm{Var}(m) = \frac{n(n-1)(2n-1)}{24}$.

Proof. Let $s = \mathrm{seq}(G)$. Since for all i we have $E[s_i] = \frac{1}{2}$, equation (2) gives

$$\mu = E[m] = E\left[\sum_{i=1}^{n-1} s_i \cdot i\right] = \sum_{i=1}^{n-1} \frac{i}{2} = \frac{1}{2}\binom{n}{2}.$$

Using the independence of the s_i and taking the variance of equation (2), we obtain

$$\sigma^2 = \mathrm{Var}[m] = \mathrm{Var}\left[\sum_{i=1}^{n-1} s_i \cdot i\right] = \sum_{i=1}^{n-1} \mathrm{Var}[s_i] \cdot i^2 = \sum_{i=1}^{n-1} \frac{i^2}{4} = \frac{n(n-1)(2n-1)}{24}. \qquad \square$$

It is interesting to note that an Erdős-Rényi random graph with $p = \frac{1}{2}$ has the same expected number of edges, but the variance of the number of edges is on the order of n^2 while the variance for a random threshold graph is on the order of n^3.

Later (§5.6) we show that $(m - \mu)/\sigma$ converges to a normal distribution.

Next, we consider the number of isolated and universal vertices of a random threshold graph. For a graph G, we let $i(G)$ denote the number of isolated vertices of G and let $u(G)$ denote the number of universal vertices of G.

Proposition 5.26. *Let G be an instance of a random threshold graph on n vertices. Then, the number of isolated vertices of G has the following distribution:*

$$\Pr\{i(G) = j\} = \begin{cases} 1/2^{n-1} & \text{for } j = n, \\ 1/2^{j+1} & \text{for } 0 \leqslant j \leqslant n-2, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. First, notice that it is impossible to have $n-1$ isolated vertices. So, $\Pr\{i(G) = n-1\} = 0$. Now, let $s = \mathrm{seq}(G)$. Then, $i(G) = n$ if and only if $s_i = 0$ for all i. Therefore, $\Pr\{i(G) = n\} = \frac{1}{2^{n-1}}$. For $0 \leqslant j \leqslant n-2$, we notice that $i(G) = j$ if and only if $s_{n-1-j} = 1$ and the last j bits equal 0. Thus, we have $\Pr\{i(G) = j\} = \frac{1}{2^{j+1}}$. $\square$

Corollary 5.27. *Let G be an instance of a random threshold graph on n vertices. Then the number of universal vertices of G has the following distribution:*

$$\Pr\{u(G) = j\} = \begin{cases} 1/2^{n-1} & \text{for } j = n, \\ 1/2^{j+1} & \text{for } 0 \leqslant j \leqslant n-2, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Notice that $i(G) = u(\overline{G})$. The result follows from Proposition 5.26. $\square$

Proposition 5.28. *Let G be an instance of a random threshold graph. Then $E[i(G)] = E[u(G)] = 1$.*

Proof. Note that $i(G)$ and $u(G)$ have the same distribution, so it is enough to find the expected value of just one of them.

$$E[i(G)] = \frac{n}{2^{n-1}} + \sum_{j=1}^{n-2} j \cdot \frac{1}{2^{j+1}} = \frac{n}{2^{n-1}} + \left(1 - \frac{n}{2^{n-1}}\right) = 1. \qquad \square$$

We note that the existence of a common neighbor between two vertices increases the likelihood that those vertices are adjacent. This clustering phenomena may be a reason that some have considered random threshold graphs as a model for social networks [11]. Here is a formal statement.

Proposition 5.29. *Let a, b, c be distinct vertices of a random threshold graph. Then*

$$\Pr\{a \sim b \mid a \sim c \sim b\} > \Pr\{a \sim b\}.$$

Proof. Using Example 3.2, we have

$$\Pr\{a \sim c \sim b\} = \Pr\{a \sim c \sim b \text{ and } a \sim b\} + \Pr\{a \sim c \sim b \text{ and } a \nsim b\} = \frac{1}{4} + \frac{1}{12} = \frac{1}{3}.$$

Therefore,

$$\Pr\{a \sim b \mid a \sim c \sim b\} = \frac{\Pr\{a \sim b \text{ and } a \sim c \sim b\}}{\Pr\{a \sim c \sim b\}} = \frac{1/4}{1/3} = \frac{3}{4} > \frac{1}{2} = \Pr\{a \sim b\}. \qquad \square$$

5.6 Small induced subgraphs

Let H be a threshold graph. We are interested in determining the number of copies of H appearing in a random threshold graph G. Specifically, we wish to understand the behavior of the random variable $N_H(G)$ which we define to be the number of induced copies of H. This is an extension of Proposition 5.25 in which $H = K_2$. Of course, if H is not a threshold graph, then $N_H = 0$.

With a modest abuse of notation, we also write $N_H(\mathbf{x})$ to mean $N_H[\Gamma(\mathbf{x})]$, i.e., the number of copies of H in the threshold graph represented by $\mathbf{x}$.

Theorem 5.30. *Let H be a threshold graph on h vertices and let N_H be the number of induced copies of H in an n-vertex random threshold graph. Then the expected value of N_H is $E[N_H] = \binom{n}{h}/2^{h-1}$ and its variance is $\mathrm{Var}(N_H) \sim cn^{2h-1}$ for some constant $c > 0$.*

For example, for $H = K_2$, Proposition 5.25 gives

$$E(N_H) = \frac{1}{2}\binom{n}{2} \qquad \text{and} \qquad \mathrm{Var}(N_H) = \frac{n(n-1)(2n-1)}{24} \sim \frac{1}{12}n^3.$$

Proof. Let A be an h-element subset of $[n]$ and define X_A to be the indicator random variable that $G[A]$ (the induced subgraph of the random threshold graph G on vertex set A) is isomorphic to H, i.e., $X_A = \mathbf{1}\{G[A] \cong H\}$. Hence $N_H = \sum X_A$.

As there are $h!/|\operatorname{Aut}(H)|$ different ways in which H might be realized on a set of h vertices, we have

$$E[X_A] = \Pr\{G[A] \cong H\} = \frac{h!}{|\operatorname{Aut}(H)|}\Pr\{H\} = \frac{h!}{|\operatorname{Aut}(H)|} \cdot \frac{|\operatorname{Aut}(H)|}{2^{h-1}h!} = \frac{1}{2^{h-1}}.$$

It follows that $E[N_H] = \binom{n}{h}/2^{h-1}$ by linearity of expectation.

It is useful to present a second derivation for $E[N_H]$ based on creation sequences. In this approach, we prepend a "wild card" symbol ($*$) to all creation sequences to stand for the first vertex in the graph. This wild card can be considered either a 1 or a 0; it does not matter as it is the first vertex in the creation list.

Let s_H be the creation sequence for the graph H (including the initial wild card) and let S be a random creation sequence (an initial $*$ followed by a random sequence of $n-1$ 1s and 0s). Then the number of induced copies of H in the random threshold graph generated by S equals the number of h-long subsequences of S that match s_H where the $*$ in s_H can match any symbol in S.

Therefore, given a fixed subset A of h entries in S, the probability that those entries match s_H is $1/2^{h-1}$, and so $\Pr\{X_A = 1\} = 1/2^{h-1}$. As there are $\binom{n}{h}$ such subsets we have $E[N_H] = \binom{n}{h}/2^{h-1}$.

For the second claim, note that

$$\operatorname{Var}(N_H) = \operatorname{Var}\left[\sum_A X_A\right] = \sum_A \sum_B \operatorname{Cov}(X_A, X_B) = \sum_{i=0}^{h} \sum_{\substack{A,B:\\ |A\cap B|=i}} \operatorname{Cov}(X_A, X_B) \tag{3}$$

where the double sums are over all $A, B \subset [n]$ with $|A| = |B| = h$, but the second is organized by the size of the intersection of A and B. Note that the number of summands in which $|A \cap B| = i$ is

$$\binom{n}{h}\binom{h}{i}\binom{n-h}{h-i} = \Theta(n^{2h-i}).$$

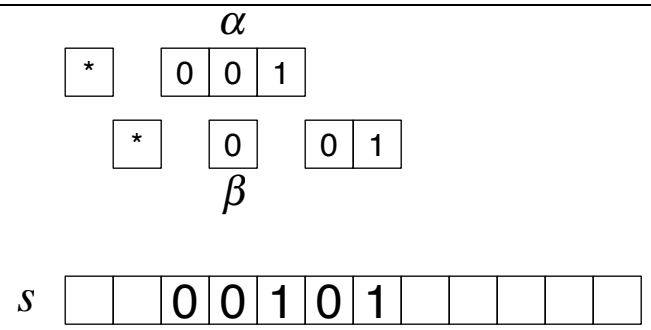
Note that when A and B are disjoint, then $\operatorname{Cov}(X_A, X_B) = 0$. When $i > 1$, this expression is $o(n^{2h-1})$. We therefore concentrate solely on the terms $\operatorname{Cov}(X_A, X_B)$ in (3) for which $|A \cap B| = 1$.

There are $\binom{n}{2h-1}$ ways to choose the elements of $A \cup B$ for which $|A \cap B| = 1$. For each such union, there are $\binom{2h-1}{h}h$ choices for the ordered pair (A, B) and the restricted sum of $\operatorname{Cov}(X_A, X_B)$ is the same for all possible choices of $A \cup B$ of size $2h-1$. That is, we have

$$\sum_{|A\cap B|=1} \operatorname{Cov}(X_A, X_B) = \binom{n}{2h-1} \sum_{\substack{A\cup B=[2h-1]\\ |A\cap B|=1}} \operatorname{Cov}(X_A, X_B) \tag{4}$$

We consider a particular term in the second sum in (4). Let $A = \{a_1 < a_2 < \cdots < a_h\}$ and $B = \{b_1 < b_2 < \cdots < b_h\}$ where $A \cup B = [2h-1]$. Let α, β be the indices of A, B [respectively] of their unique common element; that is, $a_\alpha = b_\beta$.

Figure 8 An illustration of the constraint (5). In this example $h = 4$, $s_H = *001$, $A = \{1, 3, 4, 5\}$, $B = \{2, 4, 6, 7\}$, $A \cap B = \{4\}$, $\alpha = 3$, and $\beta = 2$. The five elements of s must be as shown in order for $X_A = X_B = 1$. The probability this happens is $2^{-(2h-3)} = 2^{-5}$.



Recall that s is the (random) creation sequence of the graph G and s_H is the (fixed) creation sequence for H. We calculate $\mathrm{Cov}(X_A, X_B)$ in the three cases: (a) $\alpha = 1$ or $\beta = 1$, (b) $\alpha, \beta > 1$ and $s_H(\alpha) = s_H(\beta)$, and (3) $\alpha, \beta > 1$ and $s_H(\alpha) \neq s_H(\beta)$. In all cases $E(X_A)E(X_B) = \left(2^{-(h-1)}\right)^2 = 2^{2-2h}$. We also have the following constraint:

$$X_A = X_B = 1 \qquad \Longrightarrow \qquad \forall i > 1,\ s_H(i) = s(a_i) \quad \text{and} \quad \forall j > 1,\ s_H(j) = s(b_j). \tag{5}$$

This condition is illustrated in Figure 8.

(a) **Claim**: *If* $\alpha = 1$ *or* $\beta = 1$ *then* $\mathrm{Cov}(X_A, X_B) = 0$.

Consider the case $\alpha = 1$ (as the case $\beta = 1$ is the same). Condition (5) imposes $2h - 2$ constraints on s, and so $E[X_A X_B] = \Pr\{X_A = X_B = 1\} = 2^{-(2h-2)}$ and so $\mathrm{Cov}(X_A, X_B) = 0$.

(b) **Claim**: *If* $\alpha, \beta > 1$ *and* $s_H(\alpha) = s_H(\beta)$ *then* $\mathrm{Cov}(X_A, X_B) = 2^{-(2h-2)}$.

In this case, condition (5) imposes $2h - 3$ constraints on s (the constraints $s_H(a_i) = s(a_i)$ and $s_H(b_j) = s(b_j)$ are redundant). Thus $\Pr\{X_A = X_B = 1\} = 2^{-(2h-3)}$ and so

$$\mathrm{Cov}(X_A, X_B) = E[X_A X_B] - E[X_A]E[X_B] = 2^{-2h+3} - 2^{-2h+2} = 2^{-2h+2}.$$

(c) **Claim**: *If* $\alpha, \beta > 1$ *and* $s_H(\alpha) \neq s_H(\beta)$*, then* $\mathrm{Cov}(X_A, X_B) = -2^{-(2h-2)}$.

If $s_H(\alpha) \neq s_H(\beta)$ we have the contradictory requirements that $s(a_\alpha) \neq s(b_\beta)$ even though $a_\alpha = b_\beta$. Thus $X_A = X_B = 1$ is not possible. Therefore

$$\mathrm{Cov}(X_A, X_B) = E[X_A X_B] - E[X_A]E[X_B] = 0 - 2^{-2h+2}.$$

Thus, the only possible values of $\mathrm{Cov}(X_A, X_B)$ are 0 and $\pm 2^{-(2h-2)}$ and so next we determine how often each of these cases [(a), (b), or (c)] occur. In particular, we must show that case (b) occurs more often than case (c) to complete the proof.

We continue our assumption that $A \cup B = [2h-1]$ and $A \cap B = \{a_\alpha\} = \{b_\beta\}$. For a fixed choice of α, β, the number of pairs (A, B) is

$$\binom{\alpha+\beta-2}{\alpha-1}\binom{2h-\alpha-\beta}{h-\alpha}$$

as there are $\alpha+\beta-2$ elements of $A \cup B$ smaller than $a_\alpha = b_\beta$ (and we put $\alpha-1$ in A and the others in B) and there are $2h-\alpha-\beta$ elements of $A \cup B$ greater than $a_\alpha = b_\beta$.

To calculate the right-hand sum in (4) we use the following handy notation. For $1 \leqslant i \leqslant h$ put

$$x_i = \begin{cases} 0 & \text{for } i = 1, \\ +1 & \text{for } i > 1 \text{ and } s_H(i) = 0, \text{ and} \\ -1 & \text{for } i > 1 \text{ and } s_H(i) = 1. \end{cases}$$

Then, the right-hand sum in (4) equals

$$\frac{1}{2^{2h-2}} \sum_{\alpha,\beta} \binom{\alpha+\beta-2}{\alpha-1}\binom{2h-\alpha-\beta}{h-\alpha} x_\alpha x_\beta \tag{6}$$

and we want to show that this term is positive.

Note that this double sum in (6) is a quadratic form $\mathbf{x}^T M \mathbf{x}$ where M is the $h \times h$ matrix whose α, β-entry is

$$M_{\alpha,\beta} = \binom{\alpha+\beta-2}{\alpha-1}\binom{2h-\alpha-\beta}{h-\alpha}. \tag{7}$$

Note that M is symmetric. Thus, to show that $\mathbf{x}^T M \mathbf{x} > 0$, we show that M is positive definite.

Thanks to the representation (7), we note that M is the Hadamard product $M = P \circ \hat{P}$ where P is the Pascal-triangle matrix whose i, j-entry is $\binom{i+j-2}{i-1}$ and $\hat{P}$ is formed from P by reversing the rows and columns of P. For example, with $h = 5$, the matrix $M = P \circ \hat{P}$ is

$$\begin{bmatrix} 70 & 35 & 15 & 5 & 1 \\ 35 & 40 & 30 & 16 & 5 \\ 15 & 30 & 36 & 30 & 15 \\ 5 & 16 & 30 & 40 & 35 \\ 1 & 5 & 15 & 35 & 70 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 6 & 10 & 15 \\ 1 & 4 & 10 & 20 & 35 \\ 1 & 5 & 15 & 35 & 70 \end{bmatrix} \circ \begin{bmatrix} 70 & 35 & 15 & 5 & 1 \\ 35 & 20 & 10 & 4 & 1 \\ 15 & 10 & 6 & 3 & 1 \\ 5 & 4 & 3 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

It is known [1, 3] that the Pascal matrix P is positive definite and since $\hat{P}$ is formed from P simply by reversing the rows and columns, it is also positive definite. Finally, since M is the Hadamard product of positive definite matrices, Schur's product theorem (see, for example, [15] Theorem 7.5.3) implies that M is positive definite. □

As an example, we find the (asymptotic) variance in the number of copies of a complete subgraph on h vertices.

Proposition 5.31. *Let $H = K_h$, the complete graph on $h \geqslant 2$ vertices. Then*

$$\operatorname{Var} N_H \sim \frac{2(h-1)\binom{2h-2}{h-2}}{2^{2h-2}}\binom{n}{2h-1}.$$

Proof. Recall that by equations (3) and (4) we have that

$$\operatorname{Var} N_H \sim \binom{n}{2h-1} \sum_{\substack{A\cup B=[2h-1]\\ |A\cap B|=1}} \operatorname{Cov}(X_A, X_B)$$

We count the number of terms in equation (4) that fall under each of the cases in the proof of Theorem 5.30. Because $\operatorname{seq}(H) = *111\cdots1$, the only terms we have are those covered in claims (a) and (b). Because terms in case (a) contribute zero covariance, we count the number of terms in case (b); that is,

$$A \cap B = \{a_1 < a_2 < \cdots < a_h\} \cap \{b_1 < b_2 < \cdots < b_h\} = \{a_\alpha\} = \{b_\beta\}$$

where $\alpha, \beta > 1$.

First, notice that the element 1 belongs to exactly one of A or B; there are 2 choices for which. Without loss of generality, suppose $1 \in A$. Then, there are $\binom{2h-2}{h-2}$ ways to choose the rest of the set A from $[2h-1] - \{1\}$, not including element a_α where the sets A and B intersect. This leaves h elements from $[2h-1]$ all of which are all to be assigned to B and one of which is the intersection of A and B. The smallest element of B cannot be the common element of A and B (because we are not in case (a) of the proof of Theorem 5.30). So there are $h-1$ choices for the intersection $a_\alpha = b_\beta$. Once the intersection of A and B has been chosen, the sets A and B are fully determined. Thus, there are

$$2(h-1)\binom{2h-2}{h-2}$$

positive terms in equation (4) where $A \cup B = [2h-1]$, each contributing $2^{-(2h-2)}$ to the variance. The right hand side of equation (4) becomes

$$\frac{2(h-1)}{2^{2h-2}}\binom{2h-2}{h-2}\binom{n}{2h-1}$$

as desired. □

Note that among all threshold graphs H on h vertices, the graphs K_h and $\overline{K_h}$ give the largest value for $\operatorname{Var} N_H$ because there are no negative covariance terms in (4).

Next we show that when suitably centered and rescaled, N_H has an asymptotically normal distribution. This is a direct consequence of a theorem of Hoeffding [14] (see also [20], Theorem 5.5.1.A) on U-statistics.

We can write N_H as

$$N_H = \sum_{1\leqslant i_1<i_2<\cdots<i_h\leqslant n} k(x_{i_1}, x_{i_2}, \ldots, x_{i_h}) \tag{8}$$

where

$$k(y_1, \ldots, y_h) = \begin{cases} 1 & \text{if } \Gamma(\mathbf{y}) \cong H \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Writing N_H as in (8) shows that N_H (or more precisely, $N_H/\binom{n}{h}$ is a U-statistic and Hoeffding's result enables us to establish asymptotic normality.

One technical condition needs to be checked. Let

$$k_1(x) = E\left\{k(x, X_2, X_3, \ldots, X_h)\right\}$$

where x is a fixed value in $[0, 1]$ and the expected value is computed by integrating over the remaining $h - 1$ variables (indicated by capital letters). Note that if H has a universal vertex, then $k_1(0) = 0$ because one vertex would necessarily be isolated. On the other hand, $k_1(1) = 2^{1-(h-1)} = 2^{2-h}$. The point is that k_1 is not constant. Put

$$\zeta_1 = \operatorname{Var} k_1$$

and therefore we have $\zeta_1 > 0$. With this notation in place, Hoeffding's theorem immediately gives the following result.

Theorem 5.32. *Let H be a threshold graph on h vertices and let N_H denote the number of induced copies of H in a random threshold graph on n vertices. Then*

$$\sqrt{n}\left[\frac{N_H}{\binom{n}{h}} - 2^{1-h}\right]$$

converges in distribution to a normal distribution with mean 0 and variance $h^2\zeta_1$. □

A Partial Dyck Sequences

The results in this section are known; proofs are included here for convenience. References for this work include [7, 8].

A *Dyck path* is a lattice path from $(0, 0)$ to (n, n) (for some nonnegative integer n) that never goes below the diagonal. Equivalently, a *Dyck sequence* is a sequence of $2n$ symbols (say 1s and 0s) so that (a) there are an equal number of 1s and 0s and (b) in each initial portion of the sequence the number of 1s is equal to or greater than the number of 0s. It is well known that the number of Dyck sequences of length $2n$ is the Catalan number $c_n = \frac{1}{n+1}\binom{2n}{n}$ with generating function

$$C(x) = \sum_{n=0}^{\infty} c_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x}. \tag{9}$$

By a *partial Dyck sequence* we mean an initial portion of a Dyck sequence. That is, a sequence of 1s and 0s so that each initial portion of the sequence has at least as many 1s as 0s. In other words, the sequence satisfies condition (b) but not necessarily (a) above.

Let p_n denote the number of partial Dyck paths of length n. The list $p_0, p_1, p_2, \ldots$ is sequence A001405 in [21] in which the following result is implicitly asserted.

Proposition A.1. *Let p_n be the number of partial Dyck sequences of length n. Then for $n \geqslant 0$ we have*

$$p_n = \binom{n}{\lfloor n/2 \rfloor}.$$

Proof. We begin with the case then n is even. We claim that

$$p_{2m} = 2^{2m} - \sum_{k=0}^{m-1} c_k 2^{2m-2k-1}. \tag{10}$$

Proof of equation (10). Consider those sequences of length $2m$ that fail to be partial Dyck paths. We index these sequences beginning with index 0.

There are 2^{2m} length-$2m$ sequences of 1s and 0s. If such a sequence fails to be a partial Dyck sequence, the first initial subsequence that violates condition (b) does so at an odd index $2k+1$. [The subsequence from index 0 to index $2k$ is a Dyck sequence and then the next symbol is a 0.] The sum in the right hand side of (10) counts these failures exactly. □

With equation (10) established, we define the generating functions

$$A(x) = \sum_{m=0}^{\infty} p_{2m} x^m \qquad \text{and}$$

$$T(x) = 0 + 2x + 8x^2 + 32x^3 + \cdots = \sum_{k=1}^{\infty} 2^{2k-1} x^k = \frac{2x}{1-4x}.$$

Using $C(x)$ from (9), we note that the coefficient of x^m in $C(x)T(x)$ is

$$c_0 2^{2m-1} + c_1 2^{2m-3} + \cdots + c_{m-1} 2^1 + c_m \cdot 0 = \sum_{k=0}^{m-1} c_k 2^{2m-2k-1}$$

which is precisely the sum in (10). Therefore

$$\begin{aligned} A(x) &= \sum_{m=0}^{\infty} p_{2m} x^m \\ &= \sum_{m=0}^{\infty} 2^{2m} x^m - \sum_{m=0}^{\infty} \left(\sum_{k=0}^{m-1} c_k 2^{2m-2k-1} \right) x^m \\ &= \frac{1}{1-4x} - C(x)T(x) = \frac{1}{1-4x} - \left(\frac{1-\sqrt{1-4x}}{2x} \right) \left(\frac{2x}{1-4x} \right) \\ &= \frac{1}{\sqrt{1-4x}} = \sum_{m=0}^{\infty} \binom{2m}{m} x^m \end{aligned}$$

and therefore $p_{2m} = \binom{2m}{m}$ or, equivalently, when n is even, $p_n = \binom{n}{\lfloor n/2 \rfloor}$.

We now turn to the case when n is odd, say, $n = 2m - 1$ where $m > 0$. Such a sequence has more 1s than 0s, so it can be extended to a partial Dyck sequence of length $2m$ by the addition of a 0 or a 1. Conversely, any partial Dyck sequence of length $2m$ can be truncated to a partial Dyck sequence of length $2m - 1$ by deleting the last symbol. Hence, there is a 2-to-1 correspondence between partial Dyck sequences of length $2m$ with those of length $2m - 1$. Therefore $p_{2m} = 2p_{2m-1}$.

Thus for n odd

$$\begin{aligned} p_n = p_{2m-1} &= \frac{1}{2}p_{2m} = \frac{1}{2}\binom{2m}{m} \\ &= \frac{1}{2}\left[\binom{2m-1}{m-1} + \binom{2m-1}{m}\right] \qquad \text{(Pascal's identity)} \\ &= \binom{2m-1}{m-1} = \binom{n}{\lfloor n/2 \rfloor} \end{aligned}$$

as claimed. □

Let us call a partial Dyck sequence *strict* if in every initial subsequence the number of 1s is *greater* than the number of 0s.

Proposition A.2. *Let s_n denote number of strict partial Dyck sequences of length n. For $n > 0$ we have*

$$s_n = \binom{n-1}{\lfloor (n-1)/2 \rfloor}.$$

Proof. Every strict partial Dyck sequence of length n begins with a 1 followed by a (not necessarily strict) partial Dyck sequence of length $n - 1$. Therefore $s_n = p_{n-1}$ and the result now follows from Proposition A.1. □

References

[1] Rajendra Bhatia. *Positive Definite Matrices*. Princeton University Press, 2007.

[2] Andreas Brandstädt, Van Ban Le, and Jeremy Spinrad. *Graph Classes: A Survey*. SIAM, 1999.

[3] Robert Brawer and Magnus Pirovino. The linear algebra of the Pascal matrix. *Linear Algebra and Its Applications*, 174:13–23, 1992.

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

[5] V. Chvatal and P.L. Hammer. Aggregation of inequalities in integer programming. In P.L. Hammer, E.L. Johnson, B.H. Korte, and G.L. Nemhauser, editors, *Studies in Integer Programming*, volume 1 of *Annals of Discrete Mathematics*, pages 145–162. North-Holland, 1977.

[6] K. Ecker and S. Zaks. On a graph labelling problem. Technical report, Gesellschraft für Mathematik und Datenverarbeitung MBH, Bonn, 1977.

[7] Ömer Eğecioğlu and Alastair King. Random walks and Catalan factorization. *Congressus Numerantium*, 138:129–140, 1999.

[8] William Feller. *An Introduction to Probability Theory*. Wiley, 3 edition, 1957.

[9] Martin C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, 1980.

[10] Robert Grone and Russell Merris. The Laplacian spectrum of a graph II. *SIAM Journal on Discrete Mathematics*, 7:221–229, 1994.

[11] A. Hagberg, P.J. Swart, and D.A. Schult. Designing threshold networks with given structural and dynamical properties. *Physical Review E*, 74, 2006.

[12] Frank Harary and Uri Peled. Hamiltonian threshold graphs. *Discrete Applied Mathematics*, 16:11–15, 1987.

[13] P.B. Henderson and Y. Zalcstein. A graph-theoretic characterization of the PV_{chunk} class of synchronizing primitives. *SIAM Journal of Computing*, 6:88–108, 1977.

[14] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(293–325), 1948.

[15] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.

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

[17] Uri N. Peled. Threshold graph enumeration and series-product identities. *Congressus Numerantium*, 29:735–738, 1980.

[18] Matthew Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[19] Edward R. Scheinerman. Random interval graphs. *Combinatorica*, 8:357–371, 1988.

[20] Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, 1980.

[21] Neil J. Sloane. The on-line encyclopedia of integer sequences. `http://www.research.att.com/~njas/sequences/index.html`.