RANDOM GRAPHS WITH A MARKOV FLAVOUR

MOHABAT TARKESHIAN
THE UNIVERSITY OF WESTERN ONTARIO
DEPARTMENT OF MATHEMATICS

1. Introduction

The theory that follows has the hopeful goal of leading to a description of a rich class of random graphs whose probability distributions are especially nice. This *niceness* notion is the strongly Rayleigh property. It was first described in [BBL09] with respect to having useful closure properties as well as satisfying the strongest form of negative association for probability measures.

The story of random graphs begins with the understanding that networks exist ubiquitously in the world: social networks, electric networks, and networks within the human body to name a few. These networks grow and form relationships in a somewhat unpredictable way. This gives rise to the theory of *random graphs*. That is, assigning a *probability* to some observed network representing the actual network. The theory started in 1984 with [ER84] who defined the foundational model of random graphs. This random graph model relies on *independent* structures within the graph. Although rich in theory in its own right, we take this theory in a different direction, in that we assume there exist *dependencies* between different relationships. For example, in the context of social networks, this means that if Bob and Sally are both friends with Amy, then we assign a high probability to a friendship forming between Bob and Sally; either now or in the near future (see [RPKL07, New03] for more on these types of networks). This is sometimes called the transitivity effect. This phenomenon also exists the physical world, as in the classical *Ising model*.

The Ising model is a type of *Gibbs distribution*[1] and it describes the changing spins of particles on a piece of ferromagnetic material as the external temperature is varied. It is traditionally used to understand *phase transitions* but we will use it as a motivating example to understand why dependence structures within graphs are useful. The model starts with a graph and assigns a positive or negative spin to each vertex. As the temperature increases, the spins begin to behave as their neighbours do: that is, adjacent vertices will most likely have the same spin. This is because if neighbouring particles have the same spin, this configuration has lower energy and higher entropy, so it forms an *optimal* configuration. At a certain point, all of the spins are the same and the material behaves as a magnet. The crucial property of this model that we will elaborate on is this *local affinitive* behaviour. That is, the states of neighbouring (or *local*) vertices depend on each other and do not depend on vertices sufficiently far away. This is exactly the *Markov property* on random graphs, leading to the definition of the *Markov random graph*, which is a type of Gibbs distribution of graphs[2]. Specifically, we look at the collection of all possible graphs on a fixed number of vertices, and assign a probability to each graph. The graphs "close to each other" are defined by a neighbourhood structure on graphs.

[1] the name comes from its use in statistical mechanics.

[2] this is sometimes called an *exponential random graph* model (as in [New03, RPKL07]).

A generalization of the Ising model is the so-called *Potts model*, where the only difference is that instead of each vertex being in one of two states, a vertex can be in any one of q different states for $q > 0$. The Potts model has connections to the multivariate Tutte polynomial, which is discussed in section 5 and in more detail in [Sok05].

Some probability distributions have the *strongly Rayleigh* property. This property implies that the model has strong negative dependencies (see [BBL09, Pem00]). Negatively dependent probability models have applications to determinantal point processes and machine learning (see [ALGV20, KT12]). Then, a natural question is to ask whether the Markov random graph model has this property if we restrict our attention to a certain subset of graphs or parameters. This characterization is our goal. This paper describes the setup and a characterization of the smallest interesting complete graph (the one on 3 vertices).

To analyze the strongly Rayleigh property, we make use of a *dictionary* that allows us to go back and forth between probability models and polynomials. This dictionary was described in [ALGV19, BBL09, Sok05] to name a few. In particular, the strongly Rayleigh property on probability spaces corresponds to whether or not a polynomial is *stable* via the dictionary. Stable polynomials (see [Brä07, COW02, Wag09]) generalize the notion of real-rootedness to the multivariate setting and are used widely in both graph and matroid theory. More recently, the slightly weaker notion of *Lorentzian polynomials* (see [BH20]) was explored and found to have useful properties.

We will also describe a strongly Rayleigh characterization of another random graph model called the *random cluster model*, which connects the Potts model with the Tutte polynomial.

1.1. **Organization of the paper.** Section 2 begins with an overview of the probability theory that we will use throughout. In section 3, we define the Markov random graph model by first describing Markov random fields and showing that those of interest are the same as Gibbs distributions. We then move towards the polynomial side of the dictionary in section 4 by defining stability and its consequences for probability distributions: namely, the strongly Rayleigh property. We conclude the discussion in section 5 with a description of the random cluster model on graphs and prove a strongly Rayleigh characterization under this model.

2. Review of probability theory

We begin with a general review of the probability theory that will be used henceforth.

To properly study random phenomena, we assign probabilities to certain *outcomes* of some observations. The collection of all possible outcomes of a certain phenomenon is called the **sample space**, usually denoted by Ω. For example, the sample space of the experiment of tossing a die once is $\Omega = \{1, 2, \ldots, 6\}$. Any subset $A \subseteq \Omega$ can be thought of as an **event**. Probability theory assigns a number $P(A) \in [0, 1]$ to an event A. This is called the **probability** of observing event A. In the general setting, not all events in the powerset $\{0, 1\}^\Omega$ are assigned a probability. For our purposes, however, we will assign a probability to every subset of Ω.

Definition 2.1. A **random variable** is a function on the sample space $X : \Omega \to \mathbb{R} \cup \{\infty\}$. If D is a countable set and $X : \Omega \to D$ is a function, then X is called a **discrete random variable**.

Definition 2.2. A **probability measure** on $(\Omega, \{0, 1\}^\Omega)$ is a map $P : \{0, 1\}^\Omega \to [0, 1]$ such that the following hold:

(1) $P(\Omega) = 1$
(2) If $\{A_k\}_{k=1}^{\infty}$ are pairwise disjoint and each $A_k \subseteq \Omega$, then $P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$

The triple $(\Omega, \{0,1\}^{\Omega}, P)$ is called a **probability space** or a **probability distribution**.

We will now define what it means for events and random variables to be *independent* because the probability models of interest to us require some *dependencies*.

Definition 2.3. Two *events* $A, B \subseteq \Omega$ are said to be **independent** if $P(A \cap B) = P(A)P(B)$. Two *random variables* X, Y are said to be independent if for all $a, b \in \mathbb{R}$, $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$.

We finish with a statement of *Bayes' rule* that we will use in one of the forthcoming proofs.

Definition 2.4. The **conditional probability of** A **given** B when $P(B) > 0$ is

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}.$$

Theorem 1. *(Bayes' rule). If* $A, B \subseteq \Omega$ *and* $P(B) > 0$, *then*

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Proof. This result follows by some applications of the definition of conditional probability (definition 2.4). We have that $P(A \cap B) = P(A \mid B)P(B)$ and $P(B \cap A) = P(B \mid A)P(A)$. Since $A \cap B = B \cap A$, we equate the two expressions to obtain

$$P(A \cap B) = P(B \mid A)P(A). \tag{1}$$

That is, we can write $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ as $P(B) > 0$ and (1) gives us $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$ as required. □

3. Markov random fields and Gibbs distributions: one and the same

The random graphs described herein will involve network structures that have inherent *dependencies*. The study of random graph models began with the seminal paper [ER84] where Erdős and Rényi showed the so-called *threshold phenomenon* for random graphs. In one of their classical random graph models, there are two parameters $n \in \mathbb{N}$ and $0 < p < 1$. The model starts with n vertices and edges are drawn *independently* between each pair of vertices with a fixed probability p. That is, pairs of edges are independent of one another. Although this laid the groundwork for the study of random graphs[3], in many situations it is beneficial to encode some dependency between pairs of edges. For example, when modelling real-world networks[4]. Throughout, we assume every graph is finite and undirected.

The model starts with a complete graph on m vertices, $G = K_m$. A **complete graph** is a graph where every pair of vertices are connected by an edge[5]. As an example, figure 1 displays K_8. The probability measure is defined on **spanning subgraphs** of G. By *spanning*, this means that every vertex is included in the subgraph; in other words, we are selecting some subset of edges of G. The model describes all possible relationships between the n vertices. The model encodes the probability of a certain subgraph occurring under some hypotheses. We make this precise below.

[3] for a comprehensive survey of the theory of random graphs, see [FK15].

[4] see [RPKL07, New03].

[5] vertices that are connected by an edge are sometimes said to be *adjacent*.

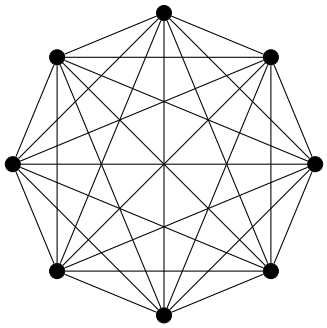

FIGURE 1. K_8, the complete graph on 8 vertices

3.1. **Markov random fields.**

Let $G = K_m = (V, E)$ for $m \in \mathbb{N}_{>0}$ where V is its set of vertices and E its set of edges. Suppose G has n edges, in symbols: $|E| = n$.

Definition 3.1. The **configuration space** of G is the set $\Lambda := \{0, 1\}^E$. We call some element $x = (x_e)_{e \in E} \in \Lambda$ a **configuration** of G[6]. Edges labelled with a 1 are sometimes called **open edges** and those labelled with a 0 are said to be **closed edges**.

The set of open edges of a configuration x defines a **spanning subgraph** of G. We denote this subgraph by G_x so that $V(G_x) := V(G)$ and $E(G_x) := \{e : x_e = 1\}$. The purpose of this graph is to encode the configuration x as a spanning subgraph of G. If we run over all possible configurations of G, the value for an edge e can vary between being open (if $x_e = 1$) or closed (if $x_e = 0$).

Definition 3.2. For a fixed $e \in E$, let X_e denote the **Bernoulli random variable** with values in $\{0, 1\}$. Denote the collection of these edge random variables by $X := \{X_e\}_{e \in E}$. The collection X is called a **random field** on E with **phases** in $\{0, 1\}$.

The random variables $\{X_e\}_{e \in E}$ are *locally* dependent in general, meaning that the value of some X_e depends on the values of edge random variables for some edges *close to* the edge e. This locality or *closeness* is encoded using a so-called *neighbourhood system*. The neighbourhood structure is inherited from the graph structure, since graphs, by nature, encode information about relationships between objects. We start by defining a general neighbourhood system and then proceed to define the one of interest: the Markovian neighbourhoods.

Definition 3.3. A **neighbourhood system** on E is a family of subsets of edges $\mathcal{N} := \{N_e \subseteq E\}_{e \in E}$ such that the following conditions are satisfied:

(1) $e \notin N_e$ for all $e \in E$, and
(2) $e_1 \in N_{e_2}$ if and only if $e_2 \in N_{e_1}$ for all $e_1, e_2 \in E$.

Further, given a neighbourhood system $\mathcal{N}$ on E, we can define a graph that completely describes the dependencies determined by the neighbourhoods. This is called the **dependency graph** of G with respect to

[6]we can also think of a configuration as a map $x : e \in E \mapsto x_e \in \{0, 1\}$.

$\mathcal{N}$, denoted $\Gamma_{\mathcal{N}}$. That is,

$$V(\Gamma_{\mathcal{N}}) := E$$
$$E(\Gamma_{\mathcal{N}}) := \{(e_1, e_2) : e_1 \in E, e_2 \in N_{e_1}\}.$$

For $e \in E$, denote $\bar{e} := E \setminus \{e\}$. For $A \subseteq E$ and a configuration $x \in \Lambda$, denote $X_A := \{X_e : e \in A\}$ and $x_A := (x_e)_{e \in A}$. We are now ready to define a *Markov random field* - a random field on G where the dependencies are determined by neighbourhoods.

Definition 3.4. ([Bré20, Gri18]) The random field X is said to be a **Markov random field (MRF)** (with respect to $\mathcal{N}$) if for all $e \in E$ and $x \in \Lambda$,

$$P(X_e = x_e \mid X_{\bar{e}} = x_{\bar{e}}) = P(X_e = x_e \mid X_{N_e} = x_{N_e}).$$

Further, if $P(X = x) > 0$ for all $x \in \Lambda$, then X is said to be a **positive Markov random field**.

Positive Markov random fields are special in that their distributions depend solely on local interactions. This is the essence of the following result.

Theorem 2. *If X is a positive Markov random field, then $P(X = x)$ is completely determined by the "local specification" map π_e where*

$$\pi_e : e \mapsto \{P(X_e = x_e \mid X_{N_e} = x_{N_e})\}.$$

Proof. The proof requires the following lemma.

Lemma 3.5. *For all $x, y \in \Lambda$*[7]*:*

$$P(x) = P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{P(x_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{P(y_i \mid x_1, x_2, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)} P(y)$$

For now, assume the lemma is true without proof. Let P and Q be two positive probability distributions of a random field on E. Suppose further that they have the same local specification maps. That is, if we fix $e \in E$, then for all $x \in \Lambda$,

$$P(X_e = x_e \mid X_{N_e} = x_{N_e}) = Q(X_e = x_e \mid X_{N_e} = x_{N_e}).$$

Now, let $y \in \Lambda$ be arbitrary. By Lemma 3.5, then for all $x \in \Lambda$,

$$\frac{P(x)}{P(y)} = \prod_{i=1}^{n} \frac{P(x_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{P(y_i \mid x_1, x_2, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}.$$

We now note that *any* random field is a MRF with respect to the topology where we take the neighbourhoods $N_e := E \setminus \{e\}$ for all $e \in E$. Taking this neighbourhood system, then the above can be rewritten as:

$$\frac{P(x)}{P(y)} = \prod_{i=1}^{n} \frac{P(x_i \mid N_{x_i})}{P(y_i \mid N_{y_i})}.$$

Since P and Q have the same local specification maps, then we have

$$\frac{P(x)}{P(y)} = \frac{Q(x)}{Q(y)} \text{ and } \frac{Q(x)}{P(x)} = \frac{Q(y)}{P(y)}.$$

[7] notice the notation $P(x_i) := P(X = x_i)$.

Since $y \in \Lambda$ was fixed, this must hold for all $x \in \Lambda$, and hence $c = {}^{Q(x)}/_{P(x)}$ is a constant. That is,

$$cP(x) = Q(x).$$

As $1 = \sum_{x\in\Lambda} Q(x) = c \sum_{x\in\Lambda} P(x) = c$, then we must have $c = 1$. Therefore, $P(x) = Q(x)$ for all x and so a positive P on a MRF is completely determined by its local specification. It remains to prove the lemma.

Proof of Lemma 3.5. We write

$$P(x) = \prod_{i=1}^{n} \frac{P(x_1, \ldots, x_{i-1}, x_i, y_{i+1}, \ldots, y_n)}{P(x_1, x_2, \ldots, x_{i-1}, y_i, \ldots, y_n)} P(y). \tag{2}$$

The left-hand side (after multiplication) gives us $\prod_{i=1}^{n} P(x_1, x_2, \ldots, x_{i-1}, y_i, \ldots, y_n)P(x)$. This is the probability of y and x occurring together with all of the intermediate (mixed) steps. Unravelling the product, everything is cancelled except for the $i = n^{\text{th}}$ term, which gives us exactly $P(x_1, x_2, \ldots, x_n) = P(x)$, so (2) holds.

Now, we recall *Bayes' rule* as stated in Theorem 1. We note that we can apply Bayes' rule since P is assumed to be positive. That is, for events A, B,

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Let $B := \{x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n\}$ and apply Bayes' rule twice to the numerator and denominator of (2) where for the numerator we take $A := x_i$ and for the denominator we take $A := y_i$. We obtain the following:

$$\begin{aligned} P(x_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n) &= \frac{P(x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n \mid x_i)P(x_i)}{P(x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)} \\ &= \frac{P(x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, x_i)}{P(B)}. \end{aligned}$$

The second line holds because $P(B \mid A)P(A) = P(A \cap B)$. Similarly, the denominator simplifies to:

$$\begin{aligned} P(y_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n) &= \frac{P(x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n \mid y_i)P(y_i)}{P(x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)} \\ &= \frac{P(x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, y_i)}{P(B)}. \end{aligned}$$

Therefore, for each i, expression (2) is equivalent to

$$\frac{P(x_1, \ldots, x_{i-1}, x_i, y_{i+1}, \ldots, y_n)}{P(x_1, \ldots, x_{i-1}, y_i, \ldots, y_n)} = \frac{P(x_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{P(y_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}.$$

Hence, the lemma holds. □

3.2. **Gibbs distributions and exponential random graphs.**

We will now describe the probability distribution of interest to us, called the *exponential random graph model (ERGM)* in some contexts ([RPKL07, New03]) and the *Gibbs distribution on graphs* in others ([Bré20, Gri18]). The name *Gibbs distribution* stems from its origins in statistical mechanics.

The Gibbs probability distribution encodes information about *local* dependencies. To accomplish this, there are functions defined on so-called *cliques* of the graph. As in the previous sections, let $G = (V, E) = K_m$, $\Lambda = \{0, 1\}^E$, and $\mathcal{N}$ a neighbourhood system on E (as in definition 3.3).

Definition 3.6. A subset $C \subseteq E$ is called an **$\mathcal{N}$-clique** if the induced dependency subgraph $\Gamma_C \subseteq \Gamma_{\mathcal{N}}$ is complete. The set of all $\mathcal{N}$-cliques of E is denoted by $\mathcal{C}$.

The Gibbs probability distribution (or *Gibbs field*) relies on defining functions on these cliques.

Definition 3.7. A **Gibbs potential** on Λ relative to $\mathcal{N}$ is a family of functions $\mathcal{P} := \{\mathcal{P}_C : \Lambda_C \to \mathbb{R} \cup \{\infty\}\}_{C \in \mathcal{C}}$. A **Gibbs energy function** derived from the Gibbs potential $\mathcal{P}$ is the sum of all of the potentials on each clique. In other words, it is the function

$$\mathcal{E}_{\mathcal{P}} : x \mapsto \sum_{C \in \mathcal{C}} \mathcal{P}_C(x_C).$$

Let $T > 0$ be a parameter sometimes referred to as the **temperature** (of the external field). A random field X on E with phases in $\{0, 1\}$ is called a **Gibbs field** or **Gibbs distribution** with potential $\mathcal{P}$ if there exists some $Z > 0$ such that for all $x \in \Lambda$,

$$P(X = x) = \frac{1}{Z} \exp\left(-\frac{1}{T}\mathcal{E}_{\mathcal{P}}(x)\right).$$

The Gibbs field X is said to be **finite** if $\mathcal{E}_{\mathcal{P}}(x) < \infty$ for all $x \in \Lambda$.

We note that it suffices to take $Z = \sum_{x \in \Lambda} \exp\left(-\frac{1}{T}\mathcal{E}_{\mathcal{P}}(x)\right)$, for Z to act as the *normalizing constant* of the distribution.

There is an interesting connection between a Gibbs distribution and the MRFs discussed in 3.1. In fact, they are one and the same. A proof of the following theorem can be found in [Bré20], Theorem 10.1.9.

Theorem 3. *(Hammersley-Clifford Theorem). X is a finite Gibbs field if and only if X is a positive Markov random field.*

Remark 3.8. As previously mentioned, the motivation for defining Gibbs distributions stems from statistical mechanics. Specifically, it defines a probability distribution that maximizes the entropy subject to certain constraints. If such a solution exists, it takes the form of a Gibbs distribution.

Example 3.9. *Bernoulli/Erdős-Rényi graphs.* Let $p \in (0, 1)$ and for all $e \in E$, draw each edge independently with probability p by assigning $P(X_e = 1) = p$. Define the (discrete) neighbourhood system by $N_e = \varnothing$ for each $e \in E$. Then, the set of all $\mathcal{N}$-cliques is $\mathcal{C} = \{\{e\} : e \in E\}$. Define the Gibbs potential (and corresponding energy function) on Λ relative to $\mathcal{N}$ by the following functions for each $e \in E, x \in \Lambda$.

$$\mathcal{P}_{\{e\}}(x_e) := -(x_e \ln(p) + (1 - x_e)\ln(1 - p))$$
$$\mathcal{E}_{\mathcal{P}}(x) = -\sum_{e \in E}(x_e \ln(p) + (1 - x_e)\ln(1 - p))$$

Define $n_i(x) := |\{e : x_e = i\}|$. The Gibbs field with the above potential is then the following distribution.

$$\begin{aligned} P(X = x) &= \frac{1}{Z}\exp\left(\sum_{e \in E} x_e \ln(p) + (1 - x_e)\ln(1 - p)\right) \\ &= \frac{1}{Z}p^{n_1(x)}(1 - p)^{n_0(x)} \end{aligned}$$

We now define the type of random graph that we are most interested in: the *Markov random graph*. This model relies on neighbourhoods that are the most intuitive: two edges have overlapping neighbourhoods if and only if they share an endpoint.

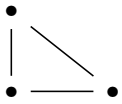

FIGURE 2. C_3, the 3-cycle graph

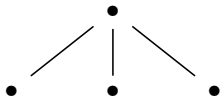

FIGURE 3. a k-star for $k = 3$

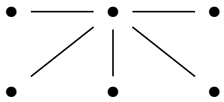

FIGURE 4. a k-star for $k = 5$

Definition 3.10. Define a neighbourhood system $\mathcal{N}$ on E by $N_e := \{e' \in E \setminus e : e, e' \text{ share a vertex}\}$ for each $e \in E$. An $\mathcal{N}$-clique C then takes two forms: it is either a **3-cycle** or a k**-star** in G (see figures 2, 3, and 4 for examples of both). The corresponding Gibbs distribution is called a **Markov random graph**. Let t, s_k be parameters, $\tau(G_x)$ the number of 3-cycles in G_x, and $\sigma_k(G_x)$ the number of k-stars in G_x. The Markov Gibbs energy function takes the following form.

$$\mathcal{E}_{\mathcal{P}}(x) = -(t\tau(G_x) + \sum_k s_k \sigma_k(G_x))$$

We can rewrite this in terms of subsets of $S \subseteq \{1, 2, \ldots, n\} =: [n] = E$ as follows. Let $S \subseteq [n]$ be a subset of edges that defines a spanning subgraph of G, denoted G_S. Let $\tau(G_S)$ be the number of 3-cycles in G_S and $\sigma_k(G_S)$ the number of k-stars. The energy function is $\mathcal{E}_{\mathcal{P}}(S) = \sum_C \mathcal{P}_C(S_C) = -(t\tau(G_S) + \sum_k s_k \sigma_k(G_S))$ and the probability distribution takes the form:

$$P(S) = \frac{1}{Z} \exp(-\mathcal{E}_{\mathcal{P}}(S)).$$

Now that we have defined the probability distribution of interest, we discuss the relationship between polynomials and probability distributions. This is the relationship we want to leverage in order to make conclusions about whether such distributions have a "nice" property.

4. Stable polynomials and strongly Rayleigh distributions

We are now ready to describe the so-called dictionary[8] between probability distributions and polynomials. Throughout, for $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, 2, \ldots, n\}$ and $\mathbb{1}_n \in \mathbb{R}^n$ is the all ones vector.

A real or complex multivariate polynomial is said to be **multi-affine** exactly when it has degree at most one in each variable. The dictionary alluded to before is between the following two objects:

(i) $\mathcal{P}_n$: multi-affine polynomials f where $f(\mathbb{1}_n) = 1$; and
(ii) $\mathcal{B}_n$: probability measures μ on the Boolean algebra $([n], \{0, 1\}^{[n]})$.

To describe the dictionary, let $\mu \in \mathcal{B}_n$ be a probability measure. We can associate a (multi-affine) polynomial to μ in a canonical way.

[8] this correspondence was described in the seminal paper [BBL09] and expanded further by [ALGV19] for one.

Definition 4.1. The **generating polynomial of** μ, denoted g_μ, is the multi-affine polynomial where the coefficients are the relevant probabilities. That is, if $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ and $\mathbf{z}^S := \prod_{i \in S} z_i$, then

$$g_\mu(\mathbf{z}) := \sum_{S \subseteq [n]} \mu(S)\mathbf{z}^S.$$

The one-to-one correspondence between $\mathcal{B}_n$ and $\mathcal{P}_n$ works as follows: if $f \in \mathcal{P}_n$ is such that $f(\mathbf{z}) = \sum_{S \subseteq [n]} a_S \mathbf{z}^S$, then define a measure $\mu_f \in \mathcal{B}_n$ by $\mu_f(S) := a_S$ for all $S \subseteq [n]$. Notice that this is a valid probability measure since $f(\mathbb{1}_n) = 1$.

Using this dictionary, we associate a multivariate notion of real-rootedness of polynomials with a nice class of probability distributions. That is, there is a connection between *stable* polynomials and *strongly Rayleigh* probability measures.

4.1. **Stable polynomials.**

We start describing the aforementioned dictionary by defining the stability property for polynomials in $\mathcal{P}_n$.

Definition 4.2. A polynomial $f \in \mathbb{C}[z_1, \ldots, z_n]$ is said to be **stable** if either $f \equiv 0$ or it does not have any roots in the open upper half of the complex plane. That is, if $\mathrm{Im}(z_j) > 0$ for all j, then $f(z_1, \ldots, z_n) \neq 0$. Further, if f has real coefficients, then it is said to be **real stable**.

Remark 4.3. We notice that if $f \in \mathbb{R}[z]$ is a univariate polynomial, then f is stable if and only if whenever $\mathrm{Im}(z) > 0$, then $f(z) \neq 0$. This implies that f must be *real-rooted*. This is precisely why real stability is regarded as a generalization of the friendly notion of real-rootedness in the univariate case. We also notice that a *product* of stable polynomials is again stable.

There is an equivalent definition of the stability of a multi-affine polynomial with real coefficients. If checking real stability directly[9], usually the following condition is verified.

Proposition 4.4. [Brä07] *A multi-affine polynomial $f \in \mathbb{R}[z_1, \ldots, z_n]$ is (real) stable if and only if for all $\mathbf{x} \in \mathbb{R}^n$ and $1 \leq i < j \leq n$,*

$$\frac{\partial f}{\partial z_i}(\mathbf{x})\frac{\partial f}{\partial z_j}(\mathbf{x}) \geq \frac{\partial^2 f}{\partial z_i \partial z_j}(\mathbf{x}) f(\mathbf{x}).$$

Remark 4.5. The difference $\frac{\partial f}{\partial z_i}(x)\frac{\partial f}{\partial z_j}(x) - \frac{\partial^2 f}{\partial z_i \partial z_j}(x) f(x)$ is sometimes called the **Rayleigh difference** of f.

A slightly weaker notion of stability is that of being *Rayleigh*.

Definition 4.6. A polynomial $f \in \mathcal{P}_n$ is **Rayleigh** if for all $\mathbf{x} \in \mathbb{R}^n_+$ and $1 \leq i < j \leq n$,

$$\frac{\partial f}{\partial z_i}(\mathbf{x})\frac{\partial f}{\partial z_j}(\mathbf{x}) \geq \frac{\partial^2 f}{\partial z_i \partial z_j}(\mathbf{x}) f(\mathbf{x}).$$

That is, the condition holds in the positive orthant of $\mathbb{R}^n$ but not necessarily for all $\mathbf{x} \in \mathbb{R}^n$.

Another way to verify stability is to make use of its closure properties. These closure properties indicate that the class of stable polynomials is well-behaved, at least in certain respects.

Proposition 4.7. [Wag09] *The following are stability-preserving operations for polynomials $f \in \mathbb{C}[z_1, \ldots, z_n]$.*

[9] by *directly*, this means without making use of the nice closure properties as outlined Proposition 4.7.

(i) *Permutation: For any* $\sigma \in S_n$,

$$f \mapsto f(z_{\sigma(1)}, z_{\sigma(2)}, \ldots, z_{\sigma(n)})$$

(ii) *Scaling: For* $c \in \mathbb{C}, \mathbf{v} \in \mathbb{R}_+^n$,

$$f \mapsto c f(v_1 z_1, v_2 z_2, \ldots, v_n z_n)$$

(iii) *Diagonalization: For* $\{i, j\} \in [n]$,

$$f \mapsto f(z_1, z_2, \ldots, z_n)|_{z_i = z_j}$$

(iv) *Specialization: For* $\mathcal{H} := \{w \in \mathbb{C} : \mathrm{Im}(w) > 0\}$ *and* $a \in \overline{\mathcal{H}}$,

$$f \mapsto f(a, z_2, \ldots, z_n)$$

(v) *Inversion: If* $\deg_1(f) = d$,

$$f \mapsto z_1^d f(-z_1^{-1}, z_2, \ldots, z_n)$$

(vi) *Differentiation: For* $i \in [n]$ *and* $\partial_i f(\mathbf{z}) := \frac{\partial f}{\partial z_i}$,

$$f \mapsto \partial_i f(\mathbf{z})$$

Proof. The proofs of these are generally straightforward. We will prove the *specialization* property (iv) to illustrate the use of *Hurwitz's theorem*.

First, we recall *Hurwitz's theorem: Suppose* $\Omega \subseteq \mathbb{C}^m$ *is an open connected set and* $\{f_m : \Omega \to \mathbb{C}\}_{m \in \mathbb{N}}$ *is a sequence of nonvanishing holomorphic functions that converges uniformly on compact subsets of* Ω. *Then, the limit* f *is either nonvanishing or* $f \equiv 0$ *on* Ω.

Throughout the proof, let $\mathcal{H}^n := \{\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{C}^n : w_i \in \mathcal{H}\}$. Suppose $a \in \overline{\mathcal{H}}$. If $\mathrm{Im}(a) > 0$, then as f is stable, so too is $f(a, z_2, \ldots, z_n)$ by definition 4.2.

It remains to consider the case when $a \in \mathbb{R}$. For each $m \in \mathbb{N}$, define the function $f_m : \mathcal{H}^n \to \mathbb{C}$ by

$$f_m(z_1, \ldots, z_n) := f\left(a + \frac{i}{2^m}, z_2, \ldots, z_n\right).$$

We notice that if we set $a' := a + \frac{i}{2^m}$ so that $\mathrm{Im}(a') > 0$, then we can apply the first case to $f(a', z_2, \ldots, z_n)$ and conclude that each f_m is stable. Each f_m is a polynomial and hence holomorphic on $\mathcal{H}^n$. Also, by stability we have that each f_m is non-vanishing on $\mathcal{H}^n$. Denote the specialization by $g := f(a, z_2, \ldots, z_n)$. Then, on compact subsets of $\mathcal{H}$, $f_m \overset{m \to \infty}{\rightrightarrows} g$ uniformly. By *Hurwitz's theorem*, then either g is nonvanishing or $g \equiv 0$ on $\mathcal{H}^n$. That is, g is stable and so specialization is a stability-preserving operation. □

As expected, these operations correspond to operations on elements of $\mathcal{B}_n$. For example, differentiating with respect to z_i corresponds to conditioning on $i \in [n]$.

4.2. **Strongly Rayleigh probability measures.**

The strongly Rayleigh property for probability measures corresponds to stability via the dictionary.

Definition 4.8. A probability measure $\mu \in \mathcal{B}_n$ is said to be **strongly Rayleigh** when its corresponding generating polynomial $g_\mu \in \mathcal{P}_n$ is stable. As alluded to by the name, $\mu \in \mathcal{B}_n$ is a **Rayleigh** measure if its corresponding generating polynomial is Rayleigh (see definition 4.6).

Remark 4.9. Although the Rayleigh and strongly Rayleigh conditions are similar, it is important to note that Rayleigh measures do not satisfy many of the relevant dependence properties that strongly Rayleigh measures do. In fact, the geometry associated to stability is a key tool in studying so-called *negative dependence* of probability measures[10]. This is because Rayleigh polynomials do not have a geometric description via their vanishing loci.

A probability measure that is strongly Rayleigh satisfies the strongest form of negative association for probability measures. This is described in [BBL09], Theorem 4.9. Let us first define some important operations on probability measures in $\mathcal{B}_n$.

Definition 4.10. Let $S \subseteq [n]$, $\mu \in \mathcal{B}_n$. The **projection** of μ onto the powerset $\{0,1\}^S$ is the measure $\mu' \in \mathcal{B}_{|S|}$ with generating polynomial given by the restriction to all indices in S. That is,

$$g_{\mu'} := g_\mu(z_1, \ldots, z_n)|_{z_i=1, i\in[n]\setminus S}$$

so that $g_{\mu'} \in \mathcal{P}_{|S|}$.

Suppose μ has generating polynomial $g_\mu(z_1, \ldots, z_n) \in \mathcal{P}_n$ and for $1 \leq i \leq n$, $a_i \in \mathbb{R}_{\geq 0}$. The measure obtained from μ by imposing the **external field**[11] $(a_1, \ldots, a_n)$ is the measure $\mu_{(a_1,\ldots,a_n)} \in \mathcal{B}_n$ with the following generating polynomial:

$$g_{\mu_{(a_1,\ldots,a_n)}} := \frac{g_\mu(a_1 z_1, a_2 z_2, \ldots, a_n z_n)}{g_\mu(a_1, a_2, \ldots, a_n)} \in \mathcal{P}_n.$$

Definition 4.11. A measure $\mu \in \mathcal{B}_n$ is said to be **negatively associated (NA)** if for any increasing functions $F, G : \{0,1\}^{[n]} \to \mathbb{R}$ that depend on disjoint sets of coordinates, then

$$\int F d\mu \int G d\mu \geq \int FG d\mu.$$

The measure μ is **conditionally negatively associated (CNA)** if all measures obtained from μ by conditioning on any of the variables is NA. Further, it is said to be **strongly conditionally negatively associated (CNA+)** if all measures obtained from μ by imposing *external fields* or *projections* (see definition 4.10) are CNA.

Proposition 4.12. [BBL09] *Strongly Rayleigh probability measures are strongly conditionally negatively associated (CNA+).*

Strongly Rayleigh measures satisfy some other properties as well, such as *stochastic domination* for truncations. Also, their so-called *rank sequences* are *ultra-log-concave* (see [BBL09]). In general, finding negatively dependent probability measures is a hard task, so it is helpful to have this class of negatively dependent measures that correspond to a rich class of polynomials.

4.3. **Examples of strongly Rayleigh measures.**

Recall that our motivation for studying probability measures is to study such distributions on graphs. A natural question, then, is to ask whether the simplest random graph model satisfies this property. That is, if every edge is independently drawn with some probability p, is this model strongly Rayleigh? As aforementioned, this is sometimes called an **Erdős-Rényi random graph**. In other contexts, it is referred to as the *Bernoulli distribution* on graphs (see example 3.9).

[10] this is described in [BB09].

[11] the term *external field* stems from its connections to the Ising model, where an external magnetic field is applied to a piece of iron.

Let $G = (V, E)$ be a finite undirected graph with $E = [n]$. We describe a probability distribution on the set of (spanning) subgraphs of G.

Definition 4.13. Let $\mathbb{B} : \{0,1\}^{[n]} \to \mathbb{R}$ be the probability distribution on the set of induced subgraphs of G where an edge $i \in [n]$ is chosen independently with probability p_i. This is named the **Bernoulli distribution**[12] on G.

Proposition 4.14. *The Bernoulli distribution on G is strongly Rayleigh.*

Proof. Let $S \subseteq [n]$ be an arbitrary subset. As each edge is chosen independently with some fixed probability, the probability of the subset (or spanning subgraph) S is $\prod_{i\in S} p_i \prod_{i\notin S}(1-p_i)$. Then, the generating polynomial (see definition 4.1) of $\mathbb{B}$ is

$$g_{\mathbb{B}}(\mathbf{z}) = \sum_{S\subseteq[n]} \prod_{i\in S} p_i \prod_{i\notin S}(1-p_i)\mathbf{z}^S.$$

By definition of $\mathbf{z}^S$, then we have the following simplification:

$$\begin{aligned} g_{\mathbb{B}}(\mathbf{z}) &= \sum_{S\subseteq[n]} \prod_{i\in S} p_i z_i \prod_{i\notin S}(1-p_i) \\ &= \prod_{i=1}^{n}(p_i z_i + (1-p_i)). \end{aligned}$$

As $p_i \in \mathbb{R}$, then each polynomial $f_i := p_i z_i + (1-p_i) \in \mathbb{R}[z_i]$ is real-rooted, i.e., real stable. The product of real stable polynomials is real stable and hence $g_{\mathbb{B}}$ is real stable. Therefore, $\mathbb{B}$ is strongly Rayleigh. □

The random graphs that we are interested in as described in section 3 do not behave like the Bernoulli distribution. This is because, by definition, Gibbs distributions on graphs (and specifically Markov graphs) assume some form of *dependence* between the edge random variables. This brings us to our research question.

Question: *Can we characterize strongly Rayleigh Gibbs distributions on graphs?*

There is some basis for this question, as we will see that the Markov random graph model on the complete graph K_3 is indeed strongly Rayleigh under some mild hypotheses. To prove this, we use the following definitions and theorem from [BBL09].

Definition 4.15. The **diagonal specialization** of a polynomial $f \in \mathbb{C}[z_1, \ldots, z_n]$ is the univariate polynomial $t \mapsto \Delta(f)(t) := f(t, t, \ldots, t)$.

Definition 4.16. The **rank sequence** of a measure $\mu \in \mathcal{P}_n$ is the sequence $\{r_k\}_{k=0}^n$ where for each $k = 0, 1, \ldots, n$,

$$r_k := \frac{\Delta(g_\mu)^{(k)}(0)}{k!}.$$

Theorem 4. ([BBL09], *Theorem 3.8) Suppose $\mu \in \mathcal{P}_n$ has a symmetric generating polynomial and rank sequence $\{r_k\}_{k=0}^n$. Then μ is strongly Rayleigh if and only if the univariate polynomial $\sum_{k=0}^n r_k z^k$ is real-rooted.*

Proposition 4.17. *The Markov random graph model on K_3 is strongly Rayleigh when any of the following hold: all of its parameters are positive, $s_1 = -s_2$, or $t = -s_2$.*

[12] this is sometimes called a *product measure* on $\{0,1\}^{[n]}$ in certain contexts (such as [BBL09]). This is also briefly described in [Ami19].

Proof. Let $G = K_3 = (V, E)$ be the complete graph on 3 vertices with $E = [3]$. Let P denote the Markov random graph measure on G. Recall that the Markov random graph model (see definition 3.10) takes the form $P(S) = \frac{1}{Z}\exp(-\mathcal{E}_{\mathcal{P}}(S))$ for $Z > 0$ and subsets of edges $S \subseteq [n]$ that define spanning subgraphs of G. The energy function is given by $\mathcal{E}_{\mathcal{P}}(S) = \sum_C \mathcal{P}_C(S_C) = -(t\tau(G_S) + \sum_k s_k\sigma_k(G_S))$ where $\tau(G_S)$ is the number of 3-cycles in the subgraph S and σ_k is the number of k-stars. The variables t, s_k represent parameters for each of these types of cliques.

The generating function of P is given by

$$g_P(\mathbf{z}) = \sum_{S\subseteq[3]} P(S)\mathbf{z}^S,$$

where $\mathbf{z} = (z_1, z_2, z_3)$. Writing this polynomial out explicitly using all spanning subgraphs of G (see figure 5 for all spanning subgraphs of K_3), we have the following expression.

$$g_P(\mathbf{z}) = \frac{1}{Z}\left(1 + e^{-s_1}(z_1 + z_2 + z_3) + e^{-(2s_1+s_2)}(z_1z_2 + z_1z_3 + z_2z_3) + e^{-(3s_1+3s_2+t)}z_1z_2z_3\right)$$

To determine whether P is strongly Rayleigh, as g_P is symmetric, we continue by verifying the condition stated in Theorem 4. First, we compute the terms of its rank sequence.

$$r_0 = \frac{1}{Z} \qquad r_1 = \frac{1}{Z}\cdot 3e^{-s_1} \qquad r_2 = \frac{1}{Z}\cdot 3e^{-(2s_1+s_2)} \qquad r_3 = \frac{1}{Z}\cdot e^{-(3s_1+3s_2+t)}$$

By Theorem 4, it suffices to show that $\sum_{k=0}^n r_k x^k$ is real-rooted.

$$\sum_{k=0}^{n} r_k x^k = \frac{1}{Z}\left(1 + 3e^{-s_1}x + 3e^{-(2s_1+s_2)}x^2 + e^{-(3s_1+3s_2+t)}x^3\right)$$

As $Z > 0$, it remains to determine the roots of the cubic $h(x) := 1 + 3e^{-s_1}x + 3e^{-(2s_1+s_2)}x^2 + e^{-(3s_1+3s_2+c_3)}x^3$. The cubic h has all real roots if and only if its discriminant is nonnegative, i.e., if

$$\begin{aligned}\operatorname{disc}(h) = &-27\left(e^{-(3s_1+3s_2+t)}\right)^2 + 18\left(e^{-(3s_1+3s_2+t)}\right)\left(3e^{-(2s_1+s_2)}\right)(3e^{-s_1}) - 4\left(e^{-(3s_1+3s_2+t)}\right)^3 - 4\left(3e^{-(2s_1+s_2)}\right)^3 \\ &+ \left(3e^{-(2s_1+s_2)}\right)^2(3e^{-s_1})^2 \geq 0.\end{aligned}$$

That is, we verify the following for $s_1, s_2, t > 0$:

$$\operatorname{disc}(h) = e^{-(6s_1+2s_2)}\left(-27e^{-(4s_2+t)} + 162e^{-(2s_2+t)} - 4e^{-(3s_1+7s_2+3t)} - 108e^{-s_2} + 81\right) \geq 0.$$

We notice that since all parameters are positive, then $\left(81 + 162e^{-(2s_2+t)}\right)$ dominates the negative terms. That is, notice that $\frac{81}{3} = 27 > 27e^{-(4s_2+t)}$, $27 > 4e^{-(3s_1+7s_2+3t)}$, and $27 + 162e^{-(2s_2+t)} > 108e^{-s_2}$. Hence, the difference is a positive value. We also notice that if $s_1 = -s_2$, the same argument applies since the only term affected is $4e^{-(3s_1+7s_2+3t)} = 4e^{-(4s_2+3t)}$. As well, if $t = -s_2$, then again we have a positive difference. Therefore, the discriminant is always positive for these choices of parameters and hence P is strongly Rayleigh by Theorem 4.

□

The Markov random graph model on K_3 is strongly Rayleigh. The question remains whether this will hold for K_m when $m > 3$. Although there is no proof of this as of yet, there is a different probability measure on graphs that does have a strongly Rayleigh characterization. This is the *random-cluster model* on graphs and how we will conclude our discussion.

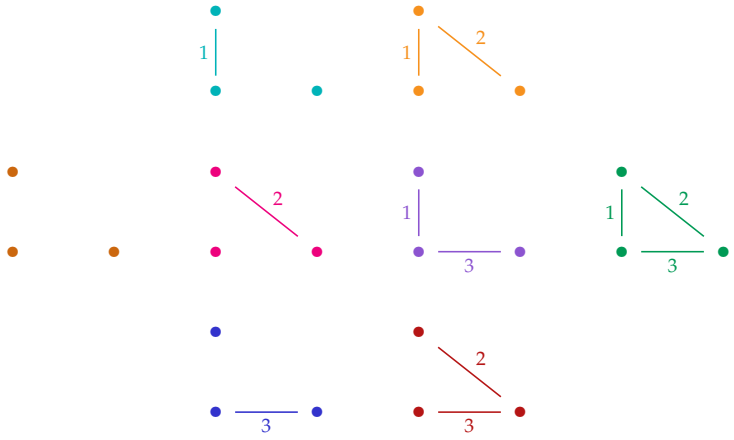

FIGURE 5. all spanning subgraphs of K_3

5. A strongly Rayleigh distribution on graphs: the random-cluster model

There is a characterization of strongly Rayleigh measures on random graphs under a different probability measure than the one discussed in section 3, specifically, the *random-cluster model* as as outlined in [Gri03]. In this section, we will describe this model and prove that the model on graphs is strongly Rayleigh precisely when the underlying graph is acyclic.

Remark 5.1. As an aside, it is relevant to note that there does exist a probability coupling of the random-cluster probability measure with the *Potts model*, under which the *correlations* in the Potts model are related to *connections* in the random-cluster model. We will not get into that discussion here, but it is important to note that there does exist a relationship between the aforementioned model(s) and the random-cluster model.

The random-cluster model was initially formulated in relation to modelling electrical networks and their series and parallel laws (see [Gri03] for more details). Throughout this section, let $G = (V, E)$ be a finite undirected graph.

Definition 5.2. Let $q > 0, 0 \leq p \leq 1$ be parameters. The former is called the **edge-weight parameter** and the latter the **cluster-weight parameter**. The **random-cluster (RC) model** $\mu : \{0,1\}^E \to \mathbb{R}$ is defined by the following. For subsets of edges $A \subseteq E$, let $k(A)$ denote the number of *connected components* in the spanning subgraph (V, A) and $Z_{RC} > 0$ the appropriate normalizing constant. Then,

$$\mu(A) := \frac{1}{Z_{RC}} q^{k(A)} p^{|A|} (1-p)^{|E \setminus A|}.$$

For our purposes (similar to those of [BBL09]), we assume that $q \leq 1$ and $p = \frac{1}{2}$. The latter is chosen because this ensures that the properties are closed under *external fields*. The RC model in this case is equal to the following for each $F \subseteq E$:

$$\mu(F) := \frac{1}{Z_{RC}} q^{k(A)}.$$

The generating polynomial of μ is $g_\mu(\mathbf{z}, q) = \sum_{A \subseteq E} q^{k(A)} \mathbf{z}^A$ where $\mathbf{z} = (z_e)_{e \in E}$. This polynomial is the ubiquitous (multivariate) **Tutte polynomial** of the graph G (see [Sok05]). To proceed with the results on the RC model, we briefly describe this polynomial.

5.1. **The multivariate Tutte polynomial.**

Definition 5.3. Let $\mathbf{z}$ and $k(A)$ be defined as above. The **multivariate Tutte polynomial** of G is the polynomial

$$Z_G(\mathbf{z}, q) := \sum_{A \subseteq E} q^{k(A)} \prod_{e \in A} z_e.$$

If $c(A)$ denotes the **cyclomatic number** of (V, A) (i.e., the minimum number of edges that must be removed to make the graph acyclic), then we can rewrite the definition as follows using the fact that $k(A) = |V| - |A| + c(A)$.

$$Z_G(\mathbf{z}, q) = q^{|V|} \sum_{A \subseteq E} q^{c(A)} \prod_{e \in A} \frac{z_e}{q} \tag{3}$$

Example 5.4. *Cycle graphs.* Let $G = C_n$ be the cycle graph on n vertices[13]. Using the equivalent expression in (3), the Tutte polynomial of C_n is the following.

$$Z_{C_n}(\mathbf{z}, q) = \prod_{e \in E} (q + z_e) + (q - 1) \prod_{e \in E} z_e \tag{4}$$

This is because the cyclomatic number $c(A) = 0$ whenever $A \subsetneq E$ and $c(E) = 1$.

Example 5.5. *Trees.* Let G be a *tree*[14]. Using (3) again, since $c(A) = 0$ for all $A \subseteq E$ as trees have no cycles, then the Tutte polynomial of G is the following.

$$Z_G(\mathbf{z}, q) = q \prod_{e \in E} (q + z_e) \tag{5}$$

The Tutte polynomial is a fascinating polynomial in its own right: it is studied both on graphs and on matroids. As well, many graph polynomials are generalizations or specializations of the Tutte polynomial (see [Tut54, Sok05, BM14] for more).

5.2. **Strongly Rayleigh RC models on graphs.**

Throughout, μ denotes the RC model on the corresponding graph.

Lemma 5.6. *If $G = C_n$ for $n \geq 3$, then the RC model on G is Rayleigh but not strongly Rayleigh.*

Proof. Denote $E(C_n) = [n]$. Using the fact that $g_\mu(\mathbf{z}, q) = Z_{C_n}(\mathbf{z}, q)$, then by example 5.4, we have the following:

$$g_\mu(\mathbf{z}, q) = \prod_{i=1}^{n} (q + z_i) + (q - 1) \prod_{i=1}^{n} z_i.$$

To determine whether μ is strongly Rayleigh, we check the condition stated in Proposition 4.4 as all of its coefficients are real. Since the polynomial is symmetric in all its variables, it suffices to check it for the pair $\{z_1, z_2\}$. That is, consider the Rayleigh difference:

$$\begin{aligned} \frac{\partial g_\mu}{\partial z_1} \frac{\partial g_\mu}{\partial z_2} - \frac{\partial^2 g_\mu}{\partial z_1 \partial z_2} g_\mu &= (1 - q) \prod_{i=1}^{n} z_i (q + z_i)((q + z_1)(q + z_2) + z_1 z_2 - z_1(q + z_2) - z_2(q + z_1)) \\ &= q^2 (1 - q) \prod_{i=3}^{n} z_i (q + z_i). \end{aligned}$$

[13] $V(C_n) := [n]$ and $E(C_n) := \{(i, j) : |i - j| = 1\}$

[14] a finite undirected graph G is a tree if and only if it is acyclic and connected.

As $q < 1$, then $(1 - q) > 0$ and hence the positivity relies on the positivity of the product. Notice that the product is positive on the positive orthant $\mathbb{R}^n_+$ (i.e., when all $z_i \geq 0$), but not necessarily positive otherwise. Hence, μ is Rayleigh but not strongly Rayleigh as per definition 4.6. □

Lemma 5.7. *If G is a tree, then the RC model on G is strongly Rayleigh.*

Proof. Using example 5.5, we have the following expression:

$$g_\mu(\mathbf{z}, q) = q \prod_{e \in E} (q + z_e).$$

The Rayleigh difference of g_μ is identically equal to zero and so by Proposition 4.4, g_μ is stable. That is, the RC model μ is strongly Rayleigh on trees. □

The above ingredients together with some nice properties of the Tutte polynomial provide a full characterization of strongly Rayleigh random cluster models on graphs.

Theorem 5. *The RC model on a graph G is strongly Rayleigh if and only if G is acyclic.*

Proof.

$\Rightarrow$: Suppose G is a finite graph with a cycle subgraph. That is, $C_m \subseteq G$ for some $m \geq 2$.

Recall that by Proposition 4.7 (iv), *specialization* is a stability-preserving operator. We specialize the generating polynomial g_μ to $\{z_e = 0\}_{e \in E(G) \setminus E(C_m)}$, where $E(C_m) \subseteq E(G)$ denotes the subset of edges that make up the cycle subgraph C_m in G. Notice that $z_e \in \overline{\mathcal{H}}$[15] for all $e \in E(G) \setminus E(C_m)$. We also note that the specialization of g_μ to this subset returns the Tutte polynomial of C_m (all other edge variables are set to zero). Therefore, $g_\mu(z_e, z_{e'}, q)$ where $e \in E(G) \setminus E(C_m)$ and $e' \in E(C_m)$ is equal to $Z_{C_m}(\mathbf{z}, q)$. By Lemma 5.6, $Z_{C_m}(\mathbf{z}, q)$ is not stable and hence the specialization of g_μ is not stable. Then by Proposition 4.7 (iv), it follows that g_μ is not stable.

$\Leftarrow$: Suppose G is a finite acyclic graph. If G is connected, then it is a tree and this is Lemma 5.7. If G is not connected, then by a property of the Tutte polynomial (section 4.1 of [Sok05]), the Tutte polynomial of the disjoint union of two graphs is the product of the Tutte polynomials of each component. As each component is connected, then each component is a tree and hence each Tutte polynomial is stable. As the product of stable polynomials is stable, then the generating polynomial of G is stable and hence μ is strongly Rayleigh.

□

[15] $\mathcal{H} = \{w \in \mathbb{C} : \mathrm{Im}(w) > 0\}$ as in Proposition 4.7

References

[ALGV19] N. Anari, K. Liu, S. O. Gharan, and C. Vinzant. Log-concave polynomials ii: High-dimensional walks and an fpras for counting bases of a matroid. arXiv:1811.01816 [cs.DS], January 2019.

[ALGV20] N. Anari, K. Liu, S. O. Gharan, and C. Vinzant. Log-concave polynomials iv: Approximate exchange, tight mixing times, and near-optimal sampling of forests. arXiv:2004.07220 [cs.DS], April 2020.

[Ami19] N. Amini. Stable multivariate generalizations of matching polynomials. arXiv:1905.02264 [math.CO], May 2019.

[BB09] J. Borcea and P. Brändén. Pólya-schur master theorems for circular domains and their boundaries. *Annals of Mathematics*, 170(1):465–492, July 2009.

[BBL09] J. Borcea, P. Brändén, and T.M. Liggett. Negative dependence and the geometry of polynomials. *Journal of the American Mathematical Society*, 22(2):521–567, April 2009.

[BH20] P. Brändén and J. Huh. Lorentzian polynomials. *Annals of Mathematics*, 192(3):821–891, November 2020.

[BM14] P. Brändén and L. Moci. The multivariate arithmetic tutte polynomial. *Transactions of the American Mathematical Society*, 366(10):5523–5540, October 2014.

[Brä07] P. Brändén. Polynomials with the half-plane property and matroid theory. *Advances in Mathematics*, 216:302–320, June 2007.

[Bré20] P. Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer Nature Switzerland AG, second edition, May 2020.

[COW02] Y. Choe, J. Oxley, and D. Wagner. Homogeneous multivariate polynomials with the half-plane property. *Advances in Applied Mathematics*, 32:88–187, March 2002.

[ER84] P. Erdős and A. Rényi. On the evolution of random graphs. *Transactions of the American Mathematical Society*, 286, 1984.

[FK15] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, November 2015.

[Gri03] G. Grimmett. The random-cluster model. arXiv:math/0205237 [math.PR], June 2003.

[Gri18] G. Grimmett. *Probability on graphs: Random processes on graphs and lattices*. Cambridge University Press, University Printing House, Cambridge CB2 8BS, United Kingdom, second edition, 2018.

[KT12] A. Kulesza and B. Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2-3):123–286, December 2012.

[New03] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, March 2003.

[Pem00] R. Pemantle. Towards a theory of negative dependence. *Journal of Mathematical Physics*, 41(1371), March 2000.

[RPKL07] G. Robins, P. Pattison, Y. Kalish, and D. Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks: An International Journal of Structural Analysis*, 29:173–191, May 2007.

[Sok05] A. D. Sokal. The multivariate tutte polynomial (alias potts model) for graphs and matroids. *Surveys in Combinatorics*, pages 173–226, March 2005.

[Tut54] W. T. Tutte. A contribution to the theory of chromatic polynomials. *Canadian Journal of Mathematics*, 6:80–91, 1954.

[Wag09] D. Wagner. Multivariate stable polynomials: theory and applications. *Bulletin of the American Mathematical Society*, 48(1), November 2009.