Péter Csikvári

Counting in Sparse Graphs

Lecture note

Contents

1. Statistical Physical Models

1.1 Combinatorics and statistical physics

The counting problems studied in this book are directly related to statistical physical models. These models in turn provide a large amount of motivation for the investigated phenomenons. Here we survey these models and their relationships with each other.

1.1.1 Monomer-dimer and dimer model

For a graph G let $\mathcal{M}(G)$ be the set of all matchings. Recall that a matching M of G is simply a set of edges such that now two edges in the set intersect each other. When this set has k edges, then we say that it is a k-matching or alternatively, the matching M is of size k. For a $\lambda > 0$ we can associate a probability space on $\mathcal{M}(G)$ by choosing a random matching $\mathbf{M}$ as follows:

$$\mathbb{P}(\mathbf{M} = M) = \frac{\lambda^{|M|}}{M(G, \lambda)},$$

where $M(G, \lambda)$ is the normalizing constant:

$$M(G, \lambda) = \sum_{M \in \mathcal{M}(G)} \lambda^{|M|}.$$

This model is the monomer-dimer model. The name has the following origin. In statistical physics the vertices of the graph represent particles, and edges represent some sort of interaction between certain pair of particles. A dimer is then simply a pair of particles where the interaction is active. Supposing that one particle can be in active with at most one other particles we get that dimers form a matchings. The uncovered vertices are then called monomers. We say that $M(G, \lambda)$ is the partition function of the monomer-dimer model. In mathematics it is called the matching generating function. Let $m_k(G)$ denote the number of k-matchings. Then

$$M(G, \lambda) = \sum_{k} m_k(G) \lambda^k.$$

Note that the sum runs from $k = 0$ as the empty set is a matching by definition. Naturally, once we introduced a probability distribution we can ask various natural question like what $\mathbb{E}|\mathbf{M}|$ is. It is not hard to see that

$$\mathbb{E}|\mathbf{M}| = \sum_{M \in \mathcal{M}(G)} \frac{|M|\lambda^{|M|}}{M(G,\lambda)} = \frac{\lambda M'(G,\lambda)}{M(G,\lambda)}.$$

If G has $2n$ vertices then we call an n-matching a perfect matching as it covers all vertices. The dimer model is the model where we consider a uniform distribution on the perfect matchings. Clearly, a dimer model is a monomer-dimer model without monomers. The number of perfect matchings is $\mathrm{pm}(G)$. With our previous notation $\mathrm{pm}(G) = m_n(G)$.

1.1.2 Hard-core model

For a graph G let $\mathcal{I}(G)$ be the set of all independent sets. Recall that an independent set is a subset I of the vertices such that no two element of I is adjacent in G. Here the vertices of G represent possible places for particles that repulse each other so no two adjacent vertex can be occupied by particles. For a $\lambda > 0$ we can associate a probability space on $\mathcal{I}(G)$ by choosing a random matching $\mathbf{I}$ as follows:

$$\mathbb{P}(\mathbf{I} = I) = \frac{\lambda^{|I|}}{I(G,\lambda)},$$

where $I(G,\lambda)$ is the normalizing constant:

$$I(G,\lambda) = \sum_{I \in \mathcal{I}(G)} \lambda^{|M|}.$$

Then $I(G,\lambda)$ is the partition function of the hard-core model. In mathematics it is called the independence polynomial of the graph G. Let $i_k(G)$ denote the number of independent sets of size k. Then

$$I(G,\lambda) = \sum_{k=0}^{n} i_k(G)\lambda^k.$$

Note that the sum runs from $k = 0$ as the empty set is an independent set by definition. Similar to the case of the matchings we have

$$\mathbb{E}|\mathbf{I}| = \sum_{I \in \mathcal{I}(G)} \frac{|I|\lambda^{|I|}}{I(G,\lambda)} = \frac{\lambda I'(G,\lambda)}{I(G,\lambda)}.$$

1.1.3 Ising-model

In the case of the Ising-model the vertices of the graph G represent particles. These particles have a spin which can be up $(+1)$ or down (-1). Two adjacent particles have an interaction e^{β} if they have the same spin, and $e^{-\beta}$ if they have different spin. Suppose also that there is an external magnetic field that breaks the symmetry between $+1$ and

-1. This defines a probability distribution on the possible configurations as follows: for a random spin configuration $\mathbf{S}$:

$$\mathbb{P}(\mathbf{S}=\sigma)=\frac{1}{Z}\exp\left(\sum_{(u,v)\in E(G)}\beta\sigma(u)\sigma(v)+B\sum_{u\in V(G)}\sigma(u)\right),$$

where Z is the normalizing constant:

$$Z_{\mathrm{Is}}(G,B,\beta)=\sum_{\sigma:V(G)\to\{-1,1\}}\exp\left(\sum_{(u,v)\in E(G)}\beta\sigma(u)\sigma(v)+B\sum_{u\in V(G)}\sigma(u)\right).$$

Similarly to the previous cases, Z is the partition function of the Ising-model.

When $\beta>0$ we say that it is a ferromagnetic Ising-model, and when $\beta<0$ then we say that it is an antiferromagnetic model. It turns out that the model behaves very differently in the two regimes even if we only consider extremal graph theoretic questions.

1.1.4 Potts-model and random-cluster model

Potts-model is a generalization of the Ising-model. Again the vertices of the graph G represent particles, but this time their spin or state can be one of q different states, i. e., $\sigma: V(G)\to[q]$. Two adjacent particles have an interaction e^{β} if they are in the same state, and no interaction otherwise. This defines a probability distribution on the possible configurations as follows: for a random configuration $\mathbf{S}$: let $\mathbf{1}(\text{statement})=1$ if the statement is true and 0 otherwise, then

$$\mathbb{P}(\mathbf{S}=\sigma)=\frac{1}{Z}\exp\left(\sum_{(u,v)\in E(G)}\beta\mathbf{1}(\sigma(u)=\sigma(v))\right),$$

where Z is the normalizing constant:

$$Z_{\mathrm{Po}}(G,q,\beta)=\sum_{\sigma:V(G)\to[q]}\exp\left(\sum_{(u,v)\in E(G)}\beta\mathbf{I}(\sigma(u)=\sigma(v))\right).$$

Similarly to the previous cases, Z is the partition function of the Potts-model. In the case when β is large then the system prefers configurations where the particles are in the same state. When β is a large negative number then the system prefers configuration where adjacent vertices are in different states. In the limiting case $\beta=-\infty$ when $e^{\beta}=0$, the partition function Z counts the proper colorings of G with q colors. Formally, a map $f:V\to\{1,2,\dots,q\}$ is a proper coloring if for all edges $(x,y)\in E$ we have $f(x)\neq f(y)$. For a positive integer q let $\mathrm{ch}(G,q)$ denote the number of proper colorings of G with q colors. Then it turns out that $\mathrm{ch}(G,q)$ is a polynomial in q, and so we can study this quantity even at non-integer or negative q's. This polynomial is called the chromatic polynomial [27]. It is a monic polynomial of degree $v(G)$.

Potts-model is very strongly related to the so-called random cluster model. In this model, the probability distribution is on the subsets of the edge set $E(G)$ and for a random subset $\mathbf{F}$ we have

$$\mathbb{P}(\mathbf{F} = F) = \frac{1}{Z} q^{k(F)} w^{|F|},$$

where $k(F)$ is the number of connected components of the graph $G' = (V(G), F)$. Here q is non-negative and $w \geq -1$, but not necessarily integers, and Z is the normalizing constant

$$Z_{\mathrm{RC}}(G, q, w) = \sum_{F \subseteq E(G)} q^{k(F)} w^{|F|}.$$

Lemma 1.1.1. *Let q be a positive integer, and $e^{\beta} = 1 + w$. Then*

$$Z_{\mathrm{Po}}(G, q, \beta) = Z_{\mathrm{RC}}(G, q, w).$$

Proof. Clearly,

$$\begin{aligned}
Z_{\mathrm{Po}}(G, q, \beta) &= \sum_{\sigma: V(G) \to [q]} \exp\left(\sum_{(u,v) \in E(G)} \beta \mathbf{I}(\sigma(u) = \sigma(v)) \right) \\
&= \sum_{\sigma: V(G) \to [q]} \prod_{(u,v) \in E(G)} (1 + (e^{\beta} - 1)\mathbf{I}(\sigma(u) = \sigma(v))) \\
&= \sum_{\sigma: V(G) \to [q]} \prod_{(u,v) \in E(G)} (1 + w\mathbf{I}(\sigma(u) = \sigma(v))) \\
&= \sum_{\sigma: V(G) \to [q]} \sum_{A \subseteq E(G)} w^{|A|} \prod_{(u,v) \in A} \mathbf{I}(\sigma(u) = \sigma(v)) \\
&= \sum_{A \subseteq E(G)} w^{|A|} \left(\sum_{\sigma: V(G) \to [q]} \prod_{(u,v) \in A} \mathbf{I}(\sigma(u) = \sigma(v)) \right) \\
&= \sum_{A \subseteq E(G)} w^{|A|} q^{k(A)} \\
&= Z_{\mathrm{RC}}(G, q, w).
\end{aligned}$$

□

1.1.5 Homomorphisms and partition functions

It turns out that many things in the previous sections have a common generalization.

Definition 1.1.2. Let $A = (a_{ij})$ be a $q \times q$ symmetric matrix. let us introduce a weight function $\nu : [q] \to \mathbb{R}_+$ (here $[q]$ stands for $\{1, 2, \ldots, q\}$), and let

$$Z(G, A, \nu) = \sum_{\varphi: V(G) \to [q]} \prod_{u \in V(G)} \nu(\varphi(u)) \cdot \prod_{(u,v) \in E(G)} a_{\varphi(u), \varphi(v)}.$$

When $\nu \equiv 1$ we simply write $Z(G, A)$ instead of $Z(G, A, \nu)$. Then

$$Z(G, A) = \sum_{\varphi: V(G) \to [q]} \prod_{(u,v) \in E(G)} a_{\varphi(u), \varphi(v)}.$$

If A is the adjacency matrix of a graph H and $\nu \equiv 1$ then we get the concept of the number of homomorphism from G to H:

Definition 1.1.3. For graphs G and H a map $\varphi : V(G) \to V(H)$ is a homomorphism if $(\varphi(u), \varphi(v)) \in E(H)$ whenever $(u, v) \in E(G)$. Let $\hom(G, H)$ denote the number of homomorphisms from the graph G to the graph H.

Note that when $H = K_q$, then $\hom(G, K_q) = \mathrm{ch}(G, q)$ counts the number proper colorings of G with q colors. As we mentioned, if we identify H with its adjacency matrix A_H we get that $Z(G, A_H) = \hom(G, H)$. Below we list some notable matrices for which we get that the partition function $Z(G, A)$ is indeed the partition function of a previously introduced statistical physical model.

$$A_{\mathrm{Is}(\beta)} = \begin{pmatrix} e^{\beta} & e^{-\beta} \\ e^{-\beta} & e^{\beta} \end{pmatrix}, \ A_{\mathrm{ind}} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \ A_q(w) = \begin{pmatrix} 1+w & 1 & 1 \\ 1 & 1+w & 1 \\ 1 & 1 & 1+w \end{pmatrix}.$$

The $A_q(w)$ is a matrix of size $q \times q$, in the picture $A_3(w)$ is depicted.

The expression $Z(G, A_{\mathrm{Is}(\beta)}, \nu_B)$ with $\nu_B(1) = e^B, \nu_B(1) = e^{-B}$ is the partition function of the Ising-model. The expression $Z(G, A_{\mathrm{ind}}, \nu_\lambda) = I(G, \lambda)$, where $\nu_\lambda(1) = 1$ and $\nu_\lambda(2) = \lambda$ is the partition function of the hard-core model. Finally, for a positive integer q, and the matrix $A_q(w)$ with $\equiv 1$ we have

$$Z_{\mathrm{RC}}(G, q, w) = Z(G, A_q(w)),$$

the partition function of the Potts-model.

1.2 Factor graphs

In the previous chapter we have seen that many counting problems can be phrased as homomorphism counting problem. In this chapter we introduce an even more general concept, the so-called *graphical model.* Graphical models are described by factor graphs, and generalizes both homomorphisms and permanents, in particular, matchings of bipartite graphs. A factor graph $\mathcal{G}$ naturally comes with a partition function $Z(\mathcal{G})$ and the so-called Bethe partition function $Z_B(\mathcal{G})$. We will see that a would-be inequality $Z(\mathcal{G}) \geq Z_B(\mathcal{G})$ is strongly related to both Sidorenko's conjecture and Schrijver's theorem on the number of perfect matchings of a d–regular bipartite graph.

Definition 1.2.1. A factor graph $\mathcal{G} = (F, V, E, \mathcal{X}, (g_a)_{a\in F})$ is a bipartite graph equipped with a set of functions. Its vertex set is $F \cup V$, where F is the set of function nodes, and V is the set of variable nodes. The edge set of $\mathcal{G}$ will be denoted by $E(\mathcal{G})$. The neighbors of a factor node a or variable node v will be denoted by ∂a or ∂v, respectively. For each variable node v we associate a variable x_v taking its values from a set $\mathcal{X}$. For each a there is an associated function $g_a : \mathcal{X}^{\partial a} \to \mathbb{R}_+$. The partition function of the factor graph $\mathcal{G}$ is

$$Z(\mathcal{G}) = \sum_{\underline{x}\in\mathcal{X}^V} \prod_{a\in F} g_a(\underline{x}_{\partial a}),$$

where $\underline{x}_{\partial a}$ is the restriction of $\underline{x}$ to the set ∂a.

When $\mathcal{X} = \{0, 1\}$ we speak about a binary factor graph.

Let us consider a few examples.

Example 1.2.2. Set $\mathcal{X} = [q]$. Suppose that each factor node a is connected to exactly two different vertices, and $g_a(x_1, x_2)$ is a symmetric function independent of the factor node a, and for $s, t \in [q]$ we have $g_a(s, t) = A_{s,t}$, i. e., we put the values of the function into a symmetric matrix A. Then clearly, if G denotes the (ordinary) graph with vertex set V, and for each factor node a we associate an edge $i, j \in E(G)$ if $\partial a = \{i, j\}$. Then

$$Z(\mathcal{G}) = Z(G, A).$$

Example 1.2.3. If there is non-negative matrix A of size $n \times n$ then we can associate the following binary factor graph: $V = [n]^2$ representing each element of the matrix, $F = \{r_1, \dots, r_n\} \cup \{c_1, \dots, c_n\}$ representing the rows and columns. Set

$$g_{r_i}(\sigma_{i,1}, \dots, \sigma_{i,n}) = \begin{cases} A_{i,j}^{1/2} & \text{if } \sum_{k=1}^n \sigma_{i,k} = 1 \text{ and } \sigma_{i,j} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$g_{c_j}(\sigma_{1,j}, \dots, \sigma_{n,j}) = \begin{cases} A_{i,j}^{1/2} & \text{if } \sum_{k=1}^n \sigma_{k,j} = 1 \text{ and } \sigma_{i,j} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$Z(\mathcal{G}) = \mathrm{per}(A).$$

Example 1.2.4. We can modify the definitions of g_{r_i} and g_{c_j} to obtain a version such that we obtain

$$\mathrm{allper}(A) = \sum_{I,J:|I|=|J|} \mathrm{per}(A_{I,J}).$$

So we can express the number of all matchings of a bipartite graph too. Set

$$g_{r_i}(\sigma_{i,1}, \dots, \sigma_{i,n}) = \begin{cases} 1 & \text{if } \sum_{k=1}^n \sigma_{i,k} = 0, \\ A_{i,j}^{1/2} & \text{if } \sum_{k=1}^n \sigma_{i,k} = 1 \text{ and } \sigma_{i,j} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$g_{c_j}(\sigma_{1,j},\dots,\sigma_{n,j}) = \begin{cases} 1 & \text{if } \sum_{k=1}^{n} \sigma_{k,j} = 0, \\ A_{i,j}^{1/2} & \text{if } \sum_{k=1}^{n} \sigma_{k,j} = 1 \text{ and } \sigma_{i,j} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$Z(\mathcal{G}) = \text{allper}(A).$$

1.2.1 Normal factor graphs

Certain functions can be more naturally expressed as partition functions of so-called normal factor graphs.

Definition 1.2.5. A normal factor graph is a graph equipped with a function at each vertex: $\mathcal{H} = (V, E, (f_v)_{v\in V})$. At each edge e there is a variable x_e taking values in an alphabet $\mathcal{X}$. The partition function

$$Z(\mathcal{H}) = \sum_{\sigma\in\mathcal{X}^E} \prod_{v\in V} f_v(\sigma_{\partial v}),$$

where $\sigma_{\partial v}$ is the restriction of σ to the the edges incident to the vertex v.

For instance, if $\mathcal{X} = \{0,1\}$ and

$$f_v(\sigma_1,\dots\sigma_d) = \begin{cases} 1 & \text{if } \sum_{i=1}^{d} \sigma_i = 1, \\ 0 & \text{otherwise,} \end{cases}$$

then $Z(\mathcal{H})$ is exactly the number of perfect matchings of the underlying graph.

It is clear that every normal factor graph can be turned into a factor graph by introducing new variable nodes at each edge. It is also true that each factor graph can be transformed to an equivalent normal factor graph. Given a factor graph $\mathcal{G} = (F, V, E, (g_a)_{a\in F})$ let us consider the normal factor graph $\mathcal{H}$ with vertex set $F \cup V$ and edge set E, but now the variables are at the edges and we introduce a new function f_v at each vertex v:

$$f_v(\sigma_1,\dots\sigma_d) = \begin{cases} 1 & \text{if } \sigma_1 = \dots = \sigma_d, \\ 0 & \text{otherwise,} \end{cases}$$

The point is that instead of having a variable σ_v at each vertex v we have variables $\sigma_{a,v}$ for each edge $(a, v) \in F \times V$, and the function f_v makes sure that $\sigma_{a_1,v} = \dots = \sigma_{a_k,v}$ is equal to some σ_v for each variable vertex v, otherwise $f_v(\sigma_{a_1,v},\dots,\sigma_{a_k,v})$ would be zero. Hence $Z(\mathcal{G}) = Z(\mathcal{H})$.

Since factor graphs and normal factor graphs are equivalent we will use whichever is most convenient. Homomorphisms can be expressed more naturally with factor graphs whereas matchings and perfect matchings are formulated more naturally by normal factor graphs.

1.3 Tutte polynomial

In this section we introduce the so-called Tutte polynomial.

Definition 1.3.1. Let $G = (V, E)$ be an arbitrary graph. Then the Tutte polynomial of G is defined as

$$T_G(x, y) = \sum_{A \subseteq E} (x - 1)^{k(A)-k(E)}(y - 1)^{k(A)+|A|-|V|},$$

where $k(A)$ denotes the number of connected components of the graph (V, A).

In this section we allow loops and multiple edges. The Tutte polynomial satisfies the following recursion:

$$T_G(x, y) = \begin{cases} xT_{G-e}(x, y) & \text{if } e \in E(G) \text{ is a bridge,} \\ yT_{G-e}(x, y) & \text{if } e \in E(G) \text{ is a loop,} \\ T_{G-e}(x, y) + T_{G/e}(x, y) & \text{if } e \in E(G) \text{ is neither a bridge, nor a loop.} \end{cases}$$

Bridge is just another word for a cut edge.

From this recursion it follows that

$$T_G(x, y) = \sum_{i,j} t_{i,j} x^i y^j,$$

where $t_{i,j} \geq 0$. These numbers have a combinatorial meaning in terms of spanning trees.

1.3.1 Special points

Theorem 1.3.2. *Let G be a connected graph.*
(a) $T_G(1, 1)$ counts the number of spanning trees.
(b) $T_G(2, 1)$ counts the number of spanning forests, that is, acyclic edge subsets.
(c) $T_G(1, 2)$ counts the number of connected subgraphs.
(d) $T_G(2, 2) = 2^{e(G)}$.

Proof. We have

$$T_G(x, y) = \sum_{A \subseteq E} (x - 1)^{k(A)-k(E)}(y - 1)^{k(A)+|A|-|V|},$$

where $k(E) = 1$ since G is connected. Note that $0^k = 0$ except if $k = 0$, thus $(1 - 1)^{k(A)-k(E)}(1 - 1)^{k(A)+|A|-|V|} = 1$ if and only if $k(A) = k(E) = 1$, and $|A| = |V| - k(A) = |V| - 1$, thus A is the edge set of a spanning tree. Parts (b) and (c) follow the same way, part (d) is completely trivial. □

Theorem 1.3.3. *Let G be a connected graph.*
(a) $T_G(2,0)$ counts the number of acyclic orientations.
(b) $T_G(1,0)$ counts the number of acyclic orientations with a unique source vertex that is fixed.
(c) $T_G(0,2)$ counts the strongly connected orientations.

Proof. (a) Let $a(G)$ be the number of acyclic orientations. If G contains a loop, then $a(G) = 0$. If G contains a bridge e, then $a(G) = 2a(G-e)$ since we can direct the edge e anyway. If an edge e is neither a loop, nor a bridge, then let us consider an acyclic orientation $\mathcal{O}$ of the graph $G-e$. Such an orientation always comes from a topological order, that is, an ordering of the vertices in such a way that every edge is oriented according to the orientation, left to right. Thus we can always orient the edge e at least one way: simply orient it according to some topological order consistent with the acyclic orientation $\mathcal{O}$. It can happen that we van orient the edge e in both ways. This means that in $G-e$ there is no directed path between the end vertices of e in any direction. Then it means that $\mathcal{O}$ corresponds to an acyclic orientation of G/e. Hence $a(G) = (a(G-e) - e(G/e)) + 2a(G/e) = a(G-e) + a(G/e)$. For the empty graph O_n on n vertices we have $a(O_n) = 0 = T_{O_n}(2,0) = 1$ (empty edge set A), thus the recursion and base cases of $a(G)$ and $T_G(2,0)$ coincide. Hence $T_G(2,0) = a(G)$.

(b) Let u be the unique source vertex, and let $a_u(G)$ denote the number of acyclic orientations with unique source vertex u. The proof follows the same ideas as in part (a), but we have to be a bit careful since the claim is not true for disconnected graphs. In case of a disconnected graph the Tutte polynomial factorizes, but the number of acyclic orientations with a unique source vertex is 0 since every connected component contains at least one source vertex. So the base graphs of the induction are trees. For a tree H with k edges the Tutte polynomial is x^k since all edges are bridges, and there is exactly one orientation with source vertex u, and $T_H(1,0) = 1$. If G contains a loop, then it is still true that $a_u(G) = 0$. Note that when e is a bridge, then $a_u(G-e) = a_u(G/e)$, but this is not exactly the sort of recursion we need. If $e = (u,v)$ is neither a loop, nor a bridge, then e is directed towards v since u is a source. Now if for v only e is directed towards v, then in $G-e$ there would be 2 sources, otherwise only u is the source, but fortunately this corresponds to an acyclic orientation of G/e, where u is the unique source. If aside e there are other edges that directed towards v, then this corresponds to an acyclic orientation of $G-e$, where u is the unique source. Hence $a_u(G) = a_u(G-e) + a_u(G/e)$.

To carry out the induction properly one may use the identity if $B_1, \dots, B_k$ are the 2-connected blocks of a graph G, then $T_G(x,y) = \prod_{i=1}^{k} T_{B_i}(x,y)$. Then if G is not 2-connected we can use this recursion to prove that $a_u(G) = T_G(1,0)$. If G is 2-connected, then it has a spanning tree in which u is a leaf (take a spanning tree in $G-u$ and add u

to it) and there is an edge (u, v) not in the tree for which we can use the recursion. This proves that $a_u(G) = T_G(1, 0)$.

(c) Let $s(G)$ be the number of strongly connected orientations. If G contains a bridge, then $s(G) = 0$. If e is a loop, then $s(G) = 2s(G - e)$. Observe that if $\mathcal{O}$ is a strongly connected orientation of G, then after contracting an edge e we also get a strongly connected orientation of G/e. It is an elementary exercise that if G/e has a strongly connected orientation, then one can orient the edge e at least one way such that it results a strongly connected orientation of G, and it can be directed in both ways if it corresponds to a strongly connected orientation of $G - e$. Hence $s(G) = (s(G/e) - s(G - e)) + 2s(G - e) = s(G/e) + s(G - e)$. Now the claim follows by induction.

□

1.3.2 Chromatic and flow polynomial

Definition 1.3.4. Let G be a graph. A map $\varphi : V(G) \to \{1, 2, \dots, q\}$ is a proper coloring with q colors if $\varphi(u) \neq \varphi(v)$ whenever $(u, v) \in E(G)$.

The number of proper colorings of G with q colors is denoted by $\mathrm{ch}(G, q)$.

Remark 1.3.5. The function $\mathrm{ch}(G, q)$ is polynomial in q. This is called the chromatic polynomial. For further results on chromatic polynomial see the chapter on graph polynomials.

Proposition 1.3.6. *If $e \in E(G)$ is a loop, then $\mathrm{ch}(G, q) = 0$. If $e \in E(G)$ is a bridge, then $\mathrm{ch}(G, q) = \frac{q-1}{q}\mathrm{ch}(G - e, q)$. If $e \in E(G)$ is neither loop, nor bridge, then we have*

$$\mathrm{ch}(G, q) = \mathrm{ch}(G - e, q) - \mathrm{ch}(G/e, q),$$

where G/e denotes the graph obtained from G by contracting the edge e.

Proof. Let us consider the proper colorings of $G - e$. If $e = (u, v)$ then we can distinguish two cases: u and v get different colors then it is even a proper coloring of G. If u and v get the same color then it corresponds to a proper coloring of G/e. Hence

$$\mathrm{ch}(G - e, q) = \mathrm{ch}(G, q) + \mathrm{ch}(G/e, q).$$

□

The chromatic polynomial is a special evaluation of the Tutte-polynomial. Indeed, one only needs to compare the recursion formulas of the two polynomials to prove the following statement.

Theorem 1.3.7. *We have*

$$\mathrm{ch}(G, q) = (-1)^{v(G)-k(G)} q^{k(G)} T_G(1 - q, 0).$$

Another notable special evaluation is the flow polynomial. Given an Abelian group Γ and a graph G let us consider an arbitrary orientation $\mathcal{O}$ of the edges. A function $g : E(G_{\mathcal{O}}) \to \Gamma$ is a flow if for all vertices v we have

$$\sum_{e \in N^+(v)} g(e) = \sum_{e \in N^-(v)} g(e).$$

A nowhere-zero flow is simply a flow without an edge mapped to $0 \in \Gamma$. Note that the number of nowhere-zero flows does not depend on the orientation as changing the orientation of an edge e will correspond to $g(e) \mapsto -g(e)$. Let $C(G, \Gamma)$ be the number of nowhere-zero flows.

Proposition 1.3.8. *If $e \in E(G)$ is a bridge, then $C(G, \Gamma) = 0$. If $e \in E(G)$ is a loop, then $C(G, \Gamma) = (|\Gamma| - 1)C(G - e, q)$. If $e \in E(G)$ is neither loop, nor bridge, then we have*

$$C(G/e, \Gamma) = C(G - e, \Gamma) + C(G, \Gamma)$$

Proof. If e is a bridge then in every flow g we have $g(e) = 0$ as can be seen by adding $\pm g(f)$ on one component of the vertex set of $G - e$. If e is a loop, then $g(e)$ can be any non-zero element. Now consider a nowhere-zero flow of G/e. This naturally extends to a flow of G in a unique way. If on the edge e we need to have a 0 value, then it is nowhere-zero flow of $G - e$. If on the edge e we have a non-zero value, then it gives a nowhere-zero flow of G. This construction works backward. □

The above proposition shows that $C(G, \Gamma)$ depends only on G and $|\Gamma|$, but not on the structure of Γ. For instance, $C(G, \mathbb{Z}_4) = C(G, \mathbb{Z}_2 \times \mathbb{Z}_2)$. Let $C(G, q) = C(G, \mathbb{Z}_q)$. This is the flow polynomial in G. We can also see that the flow polynomial is a special evaluation of the Tutte-polynomial.

Proposition 1.3.9.

$$C_G(q) = (-1)^{e(G) - v(G) + k(G)} T_G(0, 1 - q).$$

1.3.3 Random cluster model

There is a very clear connection between the Tutte-polynomial and the partition function of the random cluster model $Z_{\mathrm{RC}}(G, q, w)$. Namely,

$$T_G(x, y) = (x - 1)^{-k(G)}(y - 1)^{-v(G)} Z_{\mathrm{RC}}(G, (x - 1)(y - 1), y - 1).$$

2. Extremal Regular Graphs

2.1 Extremal regular graphs

2.1.1 General question

Let $P(G)$ be a graph parameter which has size roughly $c^{v(G)}$, where $v(G)$ denotes the number of vertices of the graph G. For instance, the number of spanning trees, the number of acyclic subgraphs (forests), the number of acyclic orientations, the number of (perfect) matchings, the number of homomorphisms of G into a fixed graph H.

The general question of this chapter is the following. What is

$$\sup P(G)^{1/v(G)} \text{ and } \inf P(G)^{1/v(G)}$$

along all d–regular (bipartite) graphs? In what follows let $\mathcal{G}_d$ and $\mathcal{G}_d^b$ denote the family of d–regular graphs and d–regular bipartite graphs, respectively.

It turns out that the answer often (but far from always) involves one of the following three graphs: the complete graph K_{d+1}, the complete bipartite graph $K_{d,d}$, and the infinite d–regular tree $\mathbb{T}_d$. We remark that Y. Zhao [37] published an excellent survey on the topic of this section.

Let us give some examples for each of the above three cases. A sample theorem for the first case is a theorem of Cutler and Radcliffe [11], see Theorem 2.3.2. It asserts that for any d–regular graph G we have

$$I(G, \lambda)^{1/v(G)} \geq I(K_{d+1}, \lambda)^{1/v(K_{d+1})},$$

where $I(G, \lambda)$ denotes the independence polynomial of G. For the second case we offer a consequence of Brégman's theorem, Theorem 10.3.1 and Theorem 10.3.2, which says that for a d–regular (bipartite) graph G we have

$$\text{pm}(G)^{1/v(G)} \leq \text{pm}(K_{d,d})^{1/v(K_{d,d})}.$$

For the third case, we offer a theorem of McKay, Theorem 5.4.1, which says that if $\tau(G)$ denote the number of spanning trees of G, then

$$\tau(G) \leq \frac{c \ln n}{n} \left(\frac{(d-1)^{d-1}}{(d^2 - 2d)^{d/2-1}} \right)^n.$$

At this moment it might be unclear why it is a theorem about the infinite d–regular tree $\mathbb{T}_d$, but it will be more clear later. One of the goals of Chapter 5 is to make such a statement clear.

The following table gives a number of such results and helps to find them in the book.

	All graphs		Bipartite graphs	
	Supremum	Infimum	Supremum	Infimum
$\hom(G,H)$			$K_{d,d}$ (2.2.3)	
independent sets	$K_{d,d}$ (7.2.1)	K_{d+1} (2.3.2)	$K_{d,d}$ (2.2.1)	$\mathbb{T}_d^b$
q-colorings	$K_{d,d}$	K_{d+1} (2.3.1)	$K_{d,d}$ (2.2.3)	$\mathbb{T}_d^b$ (3.3.3)
Ising model $\beta > 0$	K_{d+1}	$\mathbb{T}_d$	$K_{d,d}$ (2.2.3)	$\mathbb{T}_d^b$
Ising model $\beta < 0$	$K_{d,d}$ (2.2.3)		$K_{d,d}$ (2.2.3)	$\mathbb{T}_d^b$
Perfect matchings	$K_{d,d}$ (7.2.4)	0	$K_{d,d}$ (10.3.2)	$\mathbb{T}_d^b$ (9.1.17)
All matchings	$K_{d,d}$		$K_{d,d}$	$\mathbb{T}_d^b$
Eulerian orientations	K_{d+1} (conj.)	$\mathbb{T}_d$	$K_{d,d}$ (conj.)	$\mathbb{T}_d^b$
Spanning trees	$\mathbb{T}_d$ (5.4.1)		$\mathbb{T}_d^b$ (5.4.1)	
Forests	$\mathbb{T}_d$ (conj.)	K_{d+1} (conj.)	$\mathbb{T}_d^b$ (conj.)	$K_{d,d}$ (conj.)

2.2 The case of $K_{d,d}$

Theorem 2.2.1 (Kahn [20]). *Let G be a d–regular bipartite graph. Then for $\lambda \geq 0$ we have*

$$I(G,\lambda)^{1/v(G)} \leq I(K_{d,d},\lambda)^{1/v(K_{d,d})}.$$

It turns out one can drop the condition of bipartiteness in J. Kahn's theorem. We will get back to this interesting phenomenon in Chapter 7, see Theorem 7.2.1.

Without proof we also mention a theorem of Sah, Sawhney, Stoner and Zhao extending Kahn's theorem that was also conjectured by Kahn.

Theorem 2.2.2 (Sah, Sawhney, Stoner and Zhao). *Let G be a graph without isolated vertices. Then*

$$I(G,\lambda) \leq \prod_{(u,v)\in E(G)} I(K_{d(u),d(v)},\lambda)^{1/(d(u)d(v))},$$

where d_u is the degree of a vertex u.

We will not prove Theorem 2.2.1 separately as it is a special case of a theorem of Galvin and Tetali. Note that the proof of this more general theorem is very similar to the original Kahn's proof of Theorem 2.2.1.

Theorem 2.2.3 (Galvin and Tetali [15]). *Let G be a d–regular bipartite graph, and H be a fixed graph. Then*

$$\hom(G,H)^{1/v(G)} \le \hom(K_{d,d},H)^{1/v(K_{d,d})}.$$

Equivalently,

$$t(G,H)^{1/v(G)} \le t(K_{d,d},H)^{1/v(K_{d,d})},$$

or

$$Z(G,A)^{1/v(G)} \le Z(K_{d,d},A)^{1/v(K_{d,d})}.$$

The proof presented below for Theorem 2.2.3 is not the original one. This proof is due to Lubetzky and Zhao [23], and it will be a direct consequence of Finner's generalized Hölder's inequality, see [14]. The original proof will be discussed in Chapter 10.

Lemma 2.2.4 (Generalized Hölder's inequality [14]). *Let $\mu_1,\dots,\mu_n$ be probability measures on $\Omega_1,\dots,\Omega_n$, respectively. Let $\mu=\prod_{j=1}^n \mu_i$ be the product measure on $\prod_{j=1}^n \Omega_j$. Let $A_1,\dots,A_m$ be non-empty subsets of $[n]$, and let $\Omega_A=\prod_{i\in A}\Omega_\ell$ and $\mu_A=\prod_{j\in A}\mu_j$. Let $f_i\in L^{p_i}(\Omega_{A_i},\mu_{A_i})$ with $p_i\ge 1$ for $i=1,\dots,m$ and suppose further that $\sum_{i:j\in A_i}\frac{1}{p_i}\le 1$ for all $j\in[n]$. Then*

$$\int \prod_{i=1}^m |f_i| d\mu \le \prod_{i=1}^m \left(\int |f_i|^{p_i} d\mu_{A_i}\right)^{1/p_i}.$$

In particular, when all $p_i=d$ then $\int\prod |f_i|d\mu \le \prod\left(\int |f_i|^d d\mu_{A_i}\right)^{1/d}$ if every element j is in at most d sets.

Proof. We prove the statement by induction on n. The case $n=0$ is trivial. We have

$$\int_\Omega \prod_{i=1}^m |f_i|d\mu = \int_\Omega \prod_{i:n\in A_i}|f_i| \prod_{i:n\notin A_i}|f_i|d\mu = \int_{\Omega_{[n-1]}}\left(\int_{\Omega_n}\prod_{i:n\in A_i}|f_i|d\mu_n\right)\prod_{i:n\notin A_i}|f_i|d\mu_{[n-1]}.$$

Next we use Hölder's inequality: note that the condition $\sum_{i:n\in A_i}(1/p_i)\le 1$ is stronger than what Hölder's inequality requires, if it is less than 1 we can use an arithmetic-power mean inequality for one of the terms:

$$\int_{\Omega_n}\prod_{i:n\in A_i}|f_i|d\mu_n \le \prod_{i:n\in A_i}\left(\int_{\Omega_n}|f_i|^{p_i}d\mu_n\right)^{1/p_i}.$$

For each i with $n\in A_i$ let

$$f_i^* = \left(\int_{\Omega_n}|f_i|^{p_i}d\mu_n\right)^{1/p_i}.$$

So far we obtained that

$$\int_\Omega\prod_{i=1}^m|f_i|d\mu \le \int_{\Omega_{[n-1]}}\prod_{i:n\in A_i}|f_i^*|\prod_{i:n\notin A_i}|f_i|d\mu_{[n-1]}.$$

Now the functions f_i^* correspond to the sets $A_i^* = A_i \backslash \{n\}$. The assumption $\sum_{i:\ell\in A_i}(1/p_i) \leq 1$ for each $\ell \in [n-1]$ is valid for the new sets too, so by induction hypothesis we have

$$\int_\Omega \prod_{i=1}^m |f_i| d\mu \leq \prod_{i:n\in A_i} \left(\int_{\Omega_{[n-1]}} |f_i^*|^{p_i} d\mu_{[n-1]} \right)^{1/p_i} \prod_{i:n\notin A_i} \left(\int_{\Omega_{[n-1]}} |f_i|^{p_i} d\mu_{[n-1]} \right)^{1/p_i}$$
$$= \prod_{i=1}^m \left(\int |f_i|^{p_i} d\mu_{A_i} \right)^{1/p_i}.$$

□

Now we are ready to finish the second proof of Theorem 2.2.3.

Proof for Theorem 2.2.3. We will prove the statement in the form

$$t(G, W) \leq t(K_{d,d}, W)^{v(G)/(2d)}.$$

Let $G = (A, B, E)$ be a d–regular bipartite graph on $2n$ vertices, i. e., $|A| = |B| = n$. Let us label the elements of A and B by $a_1, \dots, a_n$ and $b_1, \dots, b_n$, respectively. For $b_i \in B$ let $A_i \subseteq [n]$ be the neighbors of the vertices of B. Set

$$g_i(x_{j_1}, \dots, x_{j_d}) = \int \prod_{k=1}^d W(y_i, x_{j_k})\ dy_i,$$

where $A_i = \{a_{j_1}, \dots, a_{j_d}\}$. For a vector $\underline{x} = (x_1, \dots, x_n)$ let $x_A = (x_{j_1}, \dots, x_{j_d})$ if $A = \{j_1, \dots, j_d\}$.

Then

$$t(G, W) = \int \prod_{(i,j)\in E(G)} W(x_i, y_j) dy_1 \dots dy_n dx_1 \dots dx_n = \int \prod_{i=1}^n g_i(x_{A_i}) dx_1 \dots dx_n.$$

Now we can use Lemma 2.2.4 with $p_i = d$, and get that

$$t(G, W) \leq \prod_{i=1}^n ||g_i||_d.$$

Note that

$$||g_i||_d^d = \int \left(\int \prod_{k=1}^d W(y, x_{j_k})\ dy \right)^d\ dx_{j_1} \dots dx_{j_d}$$
$$= \int \prod_{i=1}^d \prod_{k=1}^d W(y_i, x_{j_k})\ dx_{j_1} \dots dx_{j_d} dy_1 \dots dy_d$$
$$= t(K_{d,d}, W).$$

Hence

$$t(G, W) \leq t(K_{d,d}, W)^{n/d}.$$

We are done.

□

2.3 The case of K_{d+1}

In this section we study the extremality of the complete graph.

Theorem 2.3.1. *Let q be a positive integer. For any graph G we have*

$$\mathrm{ch}(G,q) \geq \prod_{v\in V(G)} \mathrm{ch}(K_{d(v)+1},q)^{1/(d(v)+1)}.$$

In particular, for any d–regular graph G we have

$$\mathrm{ch}(G,q)^{1/v(G)} \geq \mathrm{ch}(K_{d+1},q)^{1/v(K_{d+1})}$$

Proof. We can assume that $q \geq \Delta + 1$, otherwise the right hand side is 0, and the left hand side is always non-negative. Let π be a random order of the vertices of G. Let $d_\pi(v)$ denote the number of neighbors of v which preceeds v in the order π. Then

$$\mathrm{ch}(G,q) \geq \prod_{v\in V(G)} (q - d_\pi(v)),$$

since when we color the vertex v at least $q - d_\pi(v)$ colors are available. (Maybe more if we color certain neighbors of v with the same color.) Hence

$$\ln \mathrm{ch}(G,q) \geq \sum_{v\in V(G)} \ln(q - d_\pi(v)).$$

Now let us average it for all permutation π:

$$\ln \mathrm{ch}(G,q) \geq \mathbb{E}_\pi \left(\sum_{v\in V(G)} \ln(q - d_\pi(v)) \right) = \sum_{v\in V(G)} \mathbb{E}_\pi \ln(q - d_\pi(v)).$$

Now

$$\mathbb{E}_\pi \ln(q - d_\pi(v)) = \sum_{j=0}^{d(v)} \frac{1}{d(v)+1} \ln(q-j) = \frac{1}{d(v)+1} \ln \mathrm{ch}(K_{d(v)+1},q)$$

since $d_\pi(v)$ is $0, 1, \dots, d(v)$ with probability $\frac{1}{d(v)+1}$ in a random permutation π. Hence

$$\ln \mathrm{ch}(G,q) \geq \sum_{v\in V(G)} \frac{1}{d(v)+1} \ln \mathrm{ch}(K_{d(v)+1},q).$$

After exponentiation we get that

$$\mathrm{ch}(G,q) \geq \prod_{v\in V(G)} \mathrm{ch}(K_{d(v)+1},q)^{1/(d(v)+1)}.$$

The second claims immediately follows from this. □

For the number of independent sets a similar result was proved by Cutler and Radcliffe [11].

Theorem 2.3.2 (Cutler and Radcliffe [11]). *For any d–regular graph G we have*

$$I(G,\lambda)^{1/v(G)} \geq I(K_{d+1},\lambda)^{1/v(K_{d+1})},$$

where $I(G,\lambda)$ denotes the independence polynomial of G.

Proof. Let $v(G)=n$. We need to prove that

$$I(G,\lambda)^{d+1} \geq I(K_{d+1},\lambda)^n.$$

Note that

$$I(G_1 \cup G_2,\lambda) = I(G_1,\lambda)I(G_2,\lambda)$$

since

$$i_k(G_1 \cup G_2) = \sum_{r=0}^{k} i_r(G_1)i_{k-r}(G_2)$$

which is true because an independent set of $G_1 \cup G_2$ can be uniquely decomposed as an independent set of G_1 and an independent set of G_2, and every such union is an independent set of $G_1 \cup G_2$. This means that

$$I(G,\lambda)^{d+1} = I((d+1)G,\lambda),$$

where $(d+1)G$ is the disjoint union of $d+1$ copies of G. Similarly,

$$I(K_{d+1},\lambda)^n = I(nK_{d+1},\lambda).$$

Let $H_1=(d+1)G$ and $H_2=nK_{d+1}$. Both graphs have $n(d+1)$ vertices and we need to prove that $I(H_1,\lambda) \geq I(H_2,\lambda)$. We will actually prove that $i_k(H_1) \geq i_k(H_2)$ for all k. In fact, even the following more stronger statement is true:

$$\frac{i_k(H_1)}{i_{k-1}(H_1)} \geq \frac{i_k(H_2)}{i_{k-1}(H_2)}$$

for all $k \geq 1$. Since $i_0(H_1)=i_0(H_2)=1$ this would indeed imply that $i_k(H_1) \geq i_k(H_2)$ for all k.

For a d–regular graph H on $n(d+1)$ vertices let us consider the set

$$T = \{(S_1,S_2) \mid |S_1| = k-1, |S_2| = k, S_1 \subseteq S_2, S_1, S_2 \text{ independent sets in } H\}.$$

Clearly, we can choose S_2 in $i_k(H)$ ways, and once we have chosen S_2 we choose S_1 in k ways since any subset of $k-1$ vertices will automatically be an independent set of H. Whence

$$|T| = k \cdot i_k(H).$$

We can also give a lower bound to T. First we choose S_1 in $i_{k-1}(H)$ ways. Now we have to add one more vertex to S_1, we can do it in at least $n(d+1)-(k-1)(d+1)$ ways since there are $n(d+1)$ vertices and $|S_1 \cup N(S_1)| \leq (k-1)+d(k-1)$. Hence

$$|T| \geq (n-k+1)(d+1)i_{k-1}(H).$$

Note that in case of $H_2 = nK_{d+1}$ we have equality in this argument. Hence

$$k \cdot i_k(H) = |T| \geq (n-k+1)(d+1) i_{k-1}(H).$$

Thus applying it for $H = H_1$ we get that

$$\frac{i_k(H_1)}{i_{k-1}(H_1)} \geq \frac{(n-k+1)(d+1)}{k} = \frac{i_k(H_2)}{i_{k-1}(H_2)}$$

which proves our claim.

□

Without proof we also mention a theorem of Sah, Sawhney, Stoner and Zhao.

Theorem 2.3.3 (Sah, Sawhney, Stoner and Zhao [30]). *Let G be a graph. Then for $\lambda \geq 0$ we have*

$$I(G,\lambda) \geq \prod_{v \in V(G)} I(K_{d(v)+1},\lambda)^{1/(d(v)+1)},$$

where $d(v)$ denotes the degree of a vertex v.

2.4 The case of $\mathbb{T}_d$

In this section we only mention a few important theorem that we will prove later.

The following theorem is due to A. Schrijver. We will prove it via stable polynomials, see Theorem 9.1.17 and its proof. We will also prove a bit more general result it via graph covers and empirical measures, see Theorem 7.6.1.

Theorem 2.4.1 (Schrijver [31]). *Let $\mathrm{pm}(G)$ denote the number of perfect matchings. Then for a d–regular bipartite graph G we have*

$$\mathrm{pm}(G)^{1/v(G)} \geq \left(\frac{(d-1)^{d-1}}{d^{d-2}}\right)^{1/2}.$$

One interesting point about this theorem that there is no finite d–regular graphs for which

$$\mathrm{pm}(G)^{1/v(G)} = \left(\frac{(d-1)^{d-1}}{d^{d-2}}\right)^{1/2},$$

but if (G_n) is a sequence of bipartite d–regular graphs such that the girth $g(G_n) \to \infty$, that is, the size of the shortest cycle tends to infinity, then

$$\lim_{n\to\infty} \mathrm{pm}(G_n)^{1/v(G_n)} = \left(\frac{(d-1)^{d-1}}{d^{d-2}}\right)^{1/2}.$$

Similar theorem holds true for $M(G,\lambda)^{1/v(G)}$, in particular, for the number of all matchings. It means that in this case the extremal graph is not finite, but the infinite d–regular tree $\mathbb{T}_d$. We will make this statement more precise when we introduce the concept of Benjamini–Schramm convergence. The phenomenon that the infinite d–regular tree can play the role of the extremal graph is also very strongly related to the concept of Bethe–approximation that we will study later.

Similar theorem was proved for the number of spanning trees. This will be proved in Chapter 5.

Theorem 2.4.2 (McKay [25])**.** *If G is a d–regular graph on n vertices, and $\tau(G)$ denotes the number of its spanning trees, then*

$$\tau(G) \leq \frac{c \ln n}{n} \left(\frac{(d-1)^{d-1}}{(d^2-2d)^{d/2-1}} \right)^n .$$

3. Sidorenko's conjecture

3.1 Sidorenko's conjecture

Recall that

$$t(G,H) = \frac{\hom(G,H)}{v(H)^{v(G)}}.$$

This is called the homomorphism density, and it is the probability that a random map from $V(G)$ to $V(H)$ is a homomorphism. The main topic of this chapter is the following conjecture.

Conjecture 3.1.1 (Sidorenko [32])**.** For every bipartite graph G with $e(G)$ edges and every graph H we have

$$t(G,H) \geq t(K_2,H)^{e(G)}$$

For sake of convenience we give some alternative forms of Sidorenko's conjecture. For homomorphisms the conjecture says that for a bipartite graph G we have

$$\hom(G,H) \geq v(H)^{v(G)} \left(\frac{\hom(K_2,H)}{v(H)^2}\right)^{e(G)}.$$

For the partition function $Z(G,A)$ the conjecture says that that for a bipartite graph G we have

$$Z(G,A) \geq q^{v(G)} \left(\frac{\sum_{1\leq i,j\leq q} a_{ij}}{q^2}\right)^{e(G)},$$

or more generally for $Z(G,A,\nu)$ we get that

$$Z(G,A,\nu) \geq \left(\sum_{i=1}^{q} \nu(i)\right)^{v(G)} \left(\frac{\sum_{1\leq i,j\leq q} \nu(i)\nu(j)a_{ij}}{\left(\sum_{i=1}^{q} \nu(i)\right)^2}\right)^{e(G)}.$$

Finally for graphons the conjecture says that for a bipartite graph G we have

$$t(G,W) \geq t(K_2,W)^{e(G)} = \left(\int_{\Omega^2} W(x,y)\right)^{e(G)}.$$

It might look that the different forms are not entirely equivalent as for instance $Z(G,A)$ allows $a_{ij} \neq 0,1$, but actually the different forms are completely equivalent. The main idea behind the equivalence is to show that for any graphon W there is a sequence of graphs

G_i such that $\lim_{i\to\infty} t(G, W_{G_i}) = t(G, W)$. Once this is shown it is not hard to deduce the partition function version from the graphon. (One has to do some renormalization though since a_{ij} can be bigger than 1 while $W(x, y) \leq 1$.)

There are two essentially different approach to attack Sidorenko's conjecture. In the more classical approach we try to find families of bipartite graphs G such that the inequality holds true for all H. We say that a bipartite graph H has Sidorenko's property if

$$t(G, H) \geq t(K_2, H)^{e(G)}$$

for all graph H. For instance, it is known that G has the Sidorenko's property if G is a cycle of even length, tree, complete bipartite graph or a hypercube. In the other approach we fix H and prove that for all bipartite graphs G the inequality is true. For many graphs H it turns out that we can even drop the condition on the bipartiteness of G. Below we show many examples for such theorems.

3.2 Graphs with Sidorenko's property

Proposition 3.2.1. *Let ℓ be even. Then for the cycle C_ℓ we have*

$$t(C_\ell, H) \geq t(K_2, H)^\ell$$

for any graph H.

Proof. Let $v(H) = n$. Recall that

$$\hom(C_\ell, H) = \sum_{i=1}^{n} \rho_i^\ell,$$

where $\rho_1 \geq \cdots \geq \rho_n$ are the eigenvalues of the adjacency matrix of the graph H. Since ℓ is even we have $\rho_i^\ell \geq 0$. Hence

$$\hom(C_\ell, H) = \sum_{i=1}^{n} \rho_i^\ell \geq \rho_1^\ell \geq \left(\frac{2e(H)}{n}\right)^\ell,$$

where we used that the largest eigenvalue is greater than the average degree. Dividing both sides by n^ℓ we get that

$$t(C_\ell, H) \geq t(K_2, H)^\ell.$$

□

Proposition 3.2.2. *Let $K_{s,t}$ be the complete bipartite graph. Then*

$$t(K_{s,t}, H) \geq t(K_2, H)^{st}$$

for all graph H.

Proof. Let $v(H) = n$. A homomorphism of $K_{s,t}$ to H is a sequence $(v_1, \dots, v_s, w_1, \dots, w_t)$ such that $(v_i, w_j) \in E(H)$. For a sequence $S = (v_1, v_2, \dots, v_s)$ let $N(S)$ be the set of common neighbors of $v_1, \dots, v_s$. Since $w_j \in N(S)$ we have

$$\hom(K_{s,t}, H) = \sum_S N(S)^t,$$

where the summation goes for all S. Hence

$$\frac{\hom(K_{s,t}, H)}{n^s} = \frac{1}{n^s} \sum_S N(S)^t \geq \left(\frac{1}{n^s} \sum_S N(S) \right)^t.$$

Note that $\sum_S N(S)$ counts the sequence $(v_1, v_2, \dots, v_s, u)$ where $u \in N(S)$. Of course, we can first fix u and then choose each v_i from the $d(u)$ possible neighbors of u. Hence

$$\sum_S N(S) = \sum_{u \in V(H)} d(u)^s.$$

Hence

$$\frac{\hom(K_{s,t}, H)}{n^s} \geq \left(\frac{1}{n^s} \sum_S N(S) \right)^t = \left(\frac{1}{n^s} \sum_{u \in V(H)} d(u)^s \right)^t =$$

$$= \left(\frac{1}{n^{s-1}} \cdot \frac{1}{n} \sum_{u \in V(H)} d(u)^s \right)^t \geq \left(\frac{1}{n^{s-1}} \left(\frac{1}{n} \sum_{u \in V(H)} d(u) \right)^s \right)^t = \left(\frac{1}{n^{s-1}} \left(\frac{2e(H)}{n} \right)^s \right)^t.$$

After dividing by n^t we get that

$$t(K_{s,t}, H) = \frac{\hom(K_{s,t}, H)}{n^{s+t}} \geq \left(\frac{2e(H)}{n^2} \right)^{st} = t(K_2, H)^{st}.$$

$\square$

Our next goal is to prove that hypercubes have Sidorenko's property.

Theorem 3.2.3 (Hatami [18])**.** *Hypercubes have the Sidorenko's property.*

The proof presented here is not the original one, but due to J. Kim, C. Lee and J. Lee [21]. We start with some lemmas on box products of graphs.

Definition 3.2.4. Let T and G be two graphs. The graph $T \square G$ is defined as follows: $V(T \square G) = V(T) \times V(G)$ and $((u_1, v_1), (u_2, v_2)) \in E(T \square G)$ if $(u_1 = u_2$ and $(v_1, v_2) \in E(G))$ or $((u_1, u_2) \in E(T)$ and $v_1 = v_2)$. For graphs H and T let $\Psi_T(H)$ denote the graph with vertex set $\hom(T, H)$ and $\varphi, \psi \in \hom(T, H)$ are adjacent if $(\varphi(v), \psi(v)) \in E(H)$ for all $v \in V(T)$.

Lemma 3.2.5. *For any graphs T, G and H we have*

$$\hom(T \square G, H) = \hom(G, \Psi_T(H)).$$

Proof. Fix a vertex u then one can consider the image of the vertices $\{(u, v_i) \mid v_i \in V(T)\}$ as a homomorphism from T to H. Moreover, for adjacent vertices $u_1, u_2 \in V(G)$ the corresponding two homomorphisms have to be adjacent in $\Psi_T(H)$. This construction is clearly a one-to-one map from the set of homomorphisms from $T \square G$ to H to the set of homomorphisms from G to $\Psi_T(H)$. $\square$

Lemma 3.2.6. *If G has the Sidorenko's property then $K_2 \square G$ has the Sidorenko property.*

Proof.

$$\begin{aligned}
\hom(G \square K_2, H) &= \hom(G, \Psi_{K_2}(H)) \\
&\geq v(\Psi_{K_2}(H))^{v(G)} \left(\frac{\hom(K_2, \Psi_{K_2}(H))}{v(\Psi_{K_2}(H))^2} \right)^{e(G)} \\
&= \hom(K_2, H)^{v(G)} \left(\frac{\hom(K_2 \square K_2, H)}{\hom(K_2, H)^2} \right)^{e(G)} \\
&= \hom(K_2, H)^{v(G) - 2e(G)} \hom(C_4, H)^{e(G)} \\
&\geq \hom(K_2, H)^{v(G) - 2e(G)} \left(v(H)^4 \left(\frac{\hom(K_2, H)}{v(H)^2} \right)^4 \right)^{e(G)} \\
&= \frac{\hom(K_2, H)^{v(G) + 2e(G)}}{v(H)^{4e(G)}}.
\end{aligned}$$

In the first and third steps we used the previous lemma. In the fourth step we used the fact that $K_2 \square K_2 = C_4$ satisfies the Sidorenko's conjecture by Proposition 3.2.1. After dividing by $v(H)^{v(G \square K_2)} = v(H)^{2v(G)}$ and noting that $e(G \square K_2) = 2e(G) + v(G)$ we get that

$$t(G \square K_2, H) \geq t(K_2, H)^{e(G \square K_2)}.$$

$\square$

Proof of Theorem 3.2.3. The hypercube Q_n is $K_2 \square \dots \square K_2$ so the theorem follows from Lemma 3.2.6 with induction.

$\square$

3.3 Sidorenko's conjecture with fixed target graph

Our next goal is to show that certain graphs satisfy Sidorenko's conjecture as a target graph. Most notably, complete graphs have this property.

Theorem 3.3.1. *For any bipartite graph G we have*

$$t(G, K_q) \geq t(K_2, K_q)^{e(G)}.$$

Since $\hom(G, K_q) = \mathrm{ch}(G, q)$, the first part of Theorem 3.3.3 is exactly Theorem 3.3.1. The second part of this theorem is a slight extension which will be important for us later when we study extremality of the infinite d–regular tree for certain graph parameters.

Theorem 3.3.2 (Csikvári and Lin [10])**.** *Let $G = (A, B, E)$ be a bipartite graph. Let c be a uniform random proper coloring of G with q colors. Then,*

$$\mathbb{P}(c(u) = c(v)) \leq \frac{1}{q}$$

if u and v are in different parts of the bipartite graph, and

$$\mathbb{P}(c(u) = c(v)) \geq \frac{1}{q}$$

if u and v are in the same part of the bipartite graph. The former part is equivalent with the statement that for any bipartite graph H with a missing edge e between the two parts we have

$$\frac{\mathrm{ch}(G + e, q)}{\mathrm{ch}(G, q)} \geq \frac{q-1}{q}. \tag{3.1}$$

Theorem 3.3.3 (Csikvári and Lin [10])**.** *For any bipartite graph G on n vertices and $e(G)$ edges we have*

$$\mathrm{ch}(G, q) \geq q^n \left(\frac{q-1}{q}\right)^{e(G)}.$$

In particular, for a d–regular graph G we have

$$\mathrm{ch}(G, q)^{1/v(G)} \geq q \left(\frac{q-1}{q}\right)^{d/2}.$$

Furthermore, if the graph G contains εn vertex disjoint cycles of length at most ℓ, then there is some $c_q(\varepsilon, \ell) > 1$ such that

$$\mathrm{ch}(G, q) \geq c_q(\varepsilon, \ell)^n q^n \left(\frac{q-1}{q}\right)^{e(G)}.$$

Proof of Theorem 3.3.2. Note that by the symmetry of colors we have

$$\mathbb{P}(c(u) = c(v) = 1) = \frac{1}{q}\mathbb{P}(c(u) = c(v)),$$

and

$$\mathbb{P}(c(u) = 1, c(v) = 2) = \frac{1}{q(q-1)}\mathbb{P}(c(u) \neq c(v)).$$

Hence the claim

$$\mathbb{P}(c(u) = c(v)) \leq \frac{1}{q}$$

is equivalent with

$$\mathbb{P}(c(u) = c(v) = 1) \leq \mathbb{P}(c(u) = 1, c(v) = 2)$$

if u and v are in different parts of the bipartite graph. Equivalently,

$$\mathbb{P}(c(u) = c(v) = 1) + \mathbb{P}(c(u) = c(v) = 2) \leq \mathbb{P}(c(u) = 1, c(v) = 2) + \mathbb{P}(c(u) = 2, c(v) = 1).$$

Now let us conditioning on every colors different from colors 1 and 2. So assume that we know the colors of every vertex except the colors of those who gets color 1 or 2. Let us consider the partial colorings c in which the color of a vertex is fixed if it is not 1 or 2, and the uncolored vertices will be assigned color 1 or 2. Note that when we color the uncolored vertices by 1 and 2 we only have to make sure that the components of the induced subgraph of uncolored vertices must be properly colored. Now if u and v are in different components then by flipping the coloring on the component of u we see that the color of u and v will be the same or different in the same number of colorings which extends the partial coloring. On the other hand, if they are in the same component then u and v must have different color. Hence the claim follows.

The claim when u and v are in the same part works exactly the same way, only this time if u and v are in the same component then their color is the same.

Now if G is a bipartite graph with a missing edge $e = (u, v) \notin E(G)$ between the two parts. Then the probability that in a proper coloring of G, the vertices u and v will get different colors is

$$\frac{\mathrm{ch}(G + e, q)}{\mathrm{ch}(G, q)} \geq 1 - \frac{1}{q} \geq \frac{q-1}{q}.$$

So this is just rephrasing the first claim. □

Proof of Theorem 3.3.3. Let $e(G) = m$, and let $E(G) = \{e_1, e_2, \ldots, e_m\}$ then by the last claim we have

$$\begin{aligned}\frac{\mathrm{ch}(G, q)}{\mathrm{ch}(G - \{e_1, \ldots, e_m\}, q)} &= \frac{\mathrm{ch}(G, q)}{\mathrm{ch}(G - e_1, q)} \frac{\mathrm{ch}(G - e_1, q)}{\mathrm{ch}(G - \{e_1, e_2\}, q)} \cdots \frac{\mathrm{ch}(G - \{e_1, \ldots, e_{m-1}\}, q)}{\mathrm{ch}(G - \{e_1, e_2, \ldots, e_m\}, q)} \\ &\geq \left(\frac{q-1}{q}\right)^m.\end{aligned}$$

Note that $G - \{e_1, e_2, \ldots, e_m\}$ is simply the empty graph on $|V(G)| = n$ vertices so $\mathrm{ch}(G - \{e_1, e_2, \ldots, e_m\}, q) = q^n$. Hence

$$\mathrm{ch}(G, q) \geq q^n \left(\frac{q-1}{q}\right)^m.$$

To prove the second statement, let S be the union of εn vertex disjoint cycles of length at most ℓ together with the remaining isolated vertices. In other words, $S = C_1 \cup C_2 \cup \cdots \cup C_k \cup (n - e(S))K_1$, where C_i is a cycle of even length $t_i \leq \ell$, and $k = \varepsilon n$. Let $E(G) \setminus E(S) = \{e_1, e_2, \ldots, e_r\}$, so $r = e(G) - e(S)$. Then,

$$\frac{\mathrm{ch}(G, q)}{\mathrm{ch}(S, q)} = \frac{\mathrm{ch}(G, q)}{\mathrm{ch}(G - e_1, q)} \frac{\mathrm{ch}(G - e_1, q)}{\mathrm{ch}(G - \{e_1, e_2\}, q)} \cdots \frac{\mathrm{ch}(G - \{e_1, \ldots, e_{r-1}\}, q)}{\mathrm{ch}(G - \{e_1, e_2, \ldots, e_r\}, q)}$$

$$\geq \left(\frac{q-1}{q}\right)^r.$$

We will need the fact that if C_ℓ is a cycle of length ℓ then

$$\mathrm{ch}(C_\ell, q) = (q-1)^\ell + (-1)^\ell(q-1).$$

This can be seen by induction on ℓ by using the recursion

$$\mathrm{ch}(G, q) = \mathrm{ch}(G-e, q) - \mathrm{ch}(G/e, q), \quad \text{for any } e \in E(G).$$

Alternatively, if the Reader is familiar with spectral graph theory then we can use the observation that $\mathrm{ch}(C_\ell, q) = \hom(C_\ell, K_q)$. In general, $\hom(C_\ell, H)$ counts the number of closed walks of length ℓ in a graph H. This is also $\sum_{i=1}^n \lambda_i^\ell$, where λ_i's are the eigenvalues of the graph H. The eigenvalues of K_q is $q-1$ and -1 with multiplicity $q-1$. This gives that $\hom(C_\ell, K_q) = (q-1)^\ell + (q-1)(-1)^\ell$. In our case, the cycles have even lengths so we can omit the term $(-1)^\ell$.

Note that $e(S) = |C_1| + \cdots + |C_k|$, and as we have seen

$$\mathrm{ch}(C_i, q) = (q-1)^{t_i} + (q-1) = (q-1)^{t_i}\left(1 + \frac{1}{(q-1)^{t_i-1}}\right) \geq (q-1)^{t_i}\left(1 + \frac{1}{(q-1)^{\ell-1}}\right).$$

Hence

$$\mathrm{ch}(S, q) \geq q^{n-e(S)}(q-1)^{e(S)}\left(1 + \frac{1}{(q-1)^{\ell-1}}\right)^k.$$

Then

$$\mathrm{ch}(G, q) \geq \mathrm{ch}(S, q)\left(\frac{q-1}{q}\right)^r \geq q^{n-e(S)}(q-1)^{e(S)}\left(1 + \frac{1}{(q-1)^{\ell-1}}\right)^{\varepsilon n}\left(\frac{q-1}{q}\right)^r.$$

Hence

$$\mathrm{ch}(G, q) \geq c_q(\varepsilon, \ell)^n q^n \left(\frac{q-1}{q}\right)^{e(G)},$$

where

$$c_q(\varepsilon, \ell) = \left(1 + \frac{1}{(q-1)^{\ell-1}}\right)^{\varepsilon}.$$

□

Remark 3.3.4. In the chapter on local convergence we will see that Theorem 3.3.3 is tight in an asymptotic sense.

4. Graph Polynomials

4.1 Adjacency matrix and eigenvalues

All graphs in this chapter are simple if otherwise not stated, i. e., there are no loops and multiple edges. (Actually, it will be otherwise stated in some proof in the section on the Laplacian-matrix, where it will be convenient to allow multiple edges.)

Apart from the last section on Laplacian-matrix we are concerning with the adjacency matrix of a graph G. The adjacency matrix $A(G)$ of a simple graph $G = (V, E)$ is defined as follows: it is a symmetric matrix of size $|V| \times |V|$ labelled by the vertices of the graph G, and

$$A(G)_{u,v} = \begin{cases} 1 & \text{if } (u, v) \in E(G), \\ 0 & \text{if } (u, v) \notin E(G). \end{cases}$$

If the graph G is clear from the context we will simple write A instead of $A(G)$.

It is important to understand what it means that we multiply a vector $\underline{x} \in \mathbb{R}^V$ with $A(G)$:

$$(A\underline{x})_u = \sum_{v \in V(G)} A_{u,v} x_v = \sum_{v \in N_G(u)} x_v,$$

where $N_G(u)$ is the set of neighbors of u in the graph G. So multiplication with $A(G)$ simply means that we add up the values of the vector on the neighbors of a vertex u and we write this sum in place of u. In particular, if $\underline{x}$ is an eigenvector of A, i. e., $A\underline{x} = \lambda\underline{x}$ then for all vertex u we have

$$\lambda x_u = \sum_{v \in N_G(u)} x_v.$$

Recall from linear algebra that since A is a real symmetric matrix all eigenvalues are real, and we can choose a basis consisting of orthonormal eigenvectors. Note that if $A\underline{x} = \lambda\underline{x}$ and $A\underline{y} = \mu\underline{y}$ and $\lambda \neq \mu$ then $\underline{x}$ and $\underline{y}$ is immediately orthogonal to each other. If $\lambda = \mu$ it is not necessarily true, but we can still choose orthogonal eigenvectors from this eigenspace.

The goal of this chapter is to collect those ideas and facts from spectral graph theory that we will use later. Disclaimer: this chapter is not a brief introduction to spectral graph theory, it is much less. In fact, the main scope of this chapter is the counting of spanning trees.

4.1.1 Walks

We will use the fact many times that if A is a $n \times n$ real symmetric matrix then there exists a basis of $\mathbb{R}^n$ consisting of eigenvectors which we can choose[1] to be orthonormal. Let $\underline{u}_1, \ldots, \underline{u}_n$ be the orthonormal eigenvectors belonging to $\rho_1 \geq \cdots \geq \rho_n$: we have $A\underline{u}_i = \rho_i \underline{u}_i$, and $(\underline{u}_i, \underline{u}_j) = \delta_{ij}$.

In this section we relate the eigenvalues with the number of walks in the graph G.

Definition 4.1.1. A walk of length k in a graph G is a sequence of vertices and edges $u_0 e_1 u_1 e_2 u_2 \ldots u_{k-1} e_k u_k$ such that $u_i \in V(G)$, and e_i is an edge $u_{i-1}u_i$. If the graph is simple then we omit the edges e_i from the list as it is determined by u_{i-1}, u_i. The starting vertex of the walk is u_0, the end vertex is u_k. If $u_0 = u_k$ we speak about a closed walk. The number of walks of length k starting at vertex u, and ending at vertex v is $W_k(u, v)$, if $u = v$ we simply write $W_k(u)$. The number of closed walks of length k is denoted by $W_k(G)$.

Let us start with some elementary observations.

Proposition 4.1.2. *Let G be a simple graph with n vertices and $e(G)$ edges. Then the eigenvalues of its adjacency matrix A satisfies $\sum_{i=1}^n \rho_i = 0$ and $\sum_{i=1}^n \rho_i^2 = 2e(G)$. In general, $\sum_{i=1}^n \rho_i^\ell$ counts the number $W_\ell(G)$ of closed walks of length ℓ.*

Proof. Since G has no loop we have

$$\sum_{i=1}^n \rho_i = \mathrm{Tr} A = 0.$$

Since G has no multiple edges, the diagonal of A^2 consists of the degrees of G. Hence

$$\sum_{i=1}^n \rho_i^2 = \mathrm{Tr} A^2 = \sum d_i = 2e(G).$$

The third statement also follows from the fact $\mathrm{Tr} A^\ell$ is nothing else than the number $W_\ell(G)$ of closed walks of length ℓ. □

Proposition 4.1.3. *Let A be a symmetric matrix with eigenvalues $\rho_1, \ldots, \rho_n$ and corresponding orthonormal eigenvectors $\underline{u}_1, \ldots, \underline{u}_n$. Let $U = (\underline{u}_1, \ldots, \underline{u}_n)$ and $S = \mathrm{diag}(\rho_1, \ldots, \rho_n)$. Then*

$$A = USU^T$$

or equivalently

$$A = \sum_{k=1}^n \rho_k \underline{u}_k \underline{u}_k^T.$$

[1]For the matrix I, any basis will consists of eigenvectors as every vectors are eigenvectors, but of course they won't be orthonormal immediately.

Consequently, we have

$$A^\ell = \sum_{k=1}^{n} \rho_k^\ell \underline{u}_k \underline{u}_k^T.$$

Proof. First of all, note that $U^T = U^{-1}$ as the vectors u_i are orthonormal. Let $B = USU^T$. Let $\underline{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$, where the i'th coordinate is 1. Then

$$B\underline{u}_i = USU^T u_i = US\underline{e}_i = (\rho_1 \underline{u}_1, \dots, \rho_n \underline{u}_n)\underline{e}_i = \rho_i \underline{u}_i = A\underline{u}_i.$$

So A and B coincides on a basis, hence $A = B$. □

Proposition 4.1.4. *Let A be the adjacency matrix of a graph G with orthonormal eigenvectors $\underline{u}_1, \dots \underline{u}_n$. Let $\underline{u}_k = (u_k(1), u_k(2), \dots, u_k(n))$. Let $c_k(i,j) = u_k(i)u_k(j)$. The number of walks $W_\ell(i,j)$ of length ℓ between vertex i and j can be obtained as $\sum_{k=1}^n c_k(i,j)\rho_k^\ell$.*

Remark 4.1.5. Note that $c_k(i,i) = u_k(i)^2 \geq 0$. Furthermore, $\sum_{k=1}^n c_k(i,i) = 1$ for any $i \in V(G)$.

Proof of Proposition 4.1.4. Indeed, $W_\ell(i,j) = (A^\ell)_{ij}$. On the other hand,

$$\left(\sum_{k=1}^{n} \rho_k^\ell \underline{u}_k \underline{u}_k^T \right)_{ij} = \sum_{k=1}^{n} u_k(i) u_k(j) \rho_k^\ell.$$

□

Definition 4.1.6. The characteristic polynomial of the adjacency matrix is

$$\varphi(G,x) = \det(xI - A).$$

Proposition 4.1.7. *Let G be a graph and $v \in V(G)$. Then*

$$\frac{\varphi(G-v,x)}{\varphi(G,x)} = \sum_{\ell=0}^{\infty} W_\ell(v) x^{-(\ell+1)}.$$

Proof. As before, let A be the adjacency matrix of the graph G with eigenvalues ρ_k and orthonormal eigenvectors $\underline{u}_1, \dots \underline{u}_n$. Let $\underline{u}_k = (u_k(1), u_k(2), \dots, u_k(n))$.

We will show that

$$\frac{\varphi(G-v,x)}{\varphi(G,x)} = \sum_{k=1}^{n} \frac{u_k(v)^2}{x - \rho_k}. \tag{4.1}$$

This will imply the statement since

$$\sum_{k=1}^{n} \frac{u_k(v)^2}{x-\rho_k} = x^{-1} \sum_{k=1}^{n} \frac{u_k(v)^2}{1 - \rho_k \cdot x^{-1}} = \sum_{\ell=0}^{\infty} \left(\sum_{k=1}^{n} u_k(v)^2 \rho_k^\ell \right) x^{-(\ell+1)} = \sum_{\ell=0}^{\infty} W_\ell(v) x^{-(\ell+1)}.$$

To prove the identity 4.1 we show that it is true for all $x \notin \{\rho_1, \dots, \rho_n\}$. This implies the statement since after multiplication by $\phi(G,x)$ we have two polynomials that are equal at infinitely many places.

Let us consider the matrix $(xI - A)^{-1}$. By the usual way of computing the inverse of a matrix we get that

$$(xI - A)^{-1}_{vv} = \frac{\varphi(G - v, x)}{\varphi(G, x)}.$$

On the other hand,

$$\underline{u}_k = (xI - A)^{-1}(xI - A)\underline{u}_k = (xI - A)^{-1}(x - \rho_k)\underline{u}_k$$

thus

$$(xI - A)^{-1}\underline{u}_k = \frac{1}{x - \rho_k}\underline{u}_k.$$

Then by Proposition 4.1.3 we get that

$$(xI - A)^{-1} = \sum_{k=1}^{n} \frac{1}{x - \rho_k}\underline{u}_k\underline{u}_k^T.$$

Comparing the elements at (v, v) we get that

$$\frac{\varphi(G - v, x)}{\varphi(G, x)} = (xI - A)^{-1}_{vv} = \sum_{k=1}^{n} \frac{u_k(v)^2}{x - \rho_k}.$$

□

4.1.2 Largest eigenvalue

In this section we turn our attention to the study of the largest eigenvalue and its eigenvector.

Many of the things described in this section is just the Frobenius–Perron theory specialized for our case. On the other hand, we will cheat. Our cheating is based on the fact that we will only work with symmetric matrices, and so we can do some shortcuts in the arguments.

Proposition 4.1.8. *We have*

$$\rho_1 = \max_{||\underline{x}||=1} \underline{x}^T A\underline{x} = \max_{\underline{x}\neq 0} \frac{\underline{x}^T A\underline{x}}{||x||^2}.$$

Further, if for some vector $\underline{x}$ we have $\underline{x}^T A\underline{x} = \rho_1||\underline{x}||^2$, then $A\underline{x} = \rho_1\underline{x}$.

Proof. Let us write $\underline{x}$ in the basis of $\underline{u}_1, \dots \underline{u}_n$: $\underline{x} = \alpha_1\underline{u}_1 + \cdots + \alpha_n\underline{u}_n$. Then $||\underline{x}||^2 = \sum_{i=1}^n \alpha_i^2$ and $\underline{x}^T A\underline{x} = \sum_{i=1}^n \rho_i\alpha_i^2$. From this we immediately see that

$$\underline{x}^T A\underline{x} = \sum_{i=1}^{n} \rho_i\alpha_i^2 \le \rho_1 \sum_{i=1}^{n} \alpha_i^2 = \rho_1||\underline{x}||^2.$$

On the other hand, $\underline{u}_1^T A \underline{u}_1 = \rho_1 ||\underline{u}_1||^2$. Hence

$$\rho_1 = \max_{||\underline{x}||=1} \underline{x}^T A \underline{x} = \max_{\underline{x} \neq 0} \frac{\underline{x}^T A \underline{x}}{||\underline{x}||^2}.$$

Now assume that we have $\underline{x}^T A \underline{x} = \rho_1 ||\underline{x}||^2$ for some vector $\underline{x}$. Assume that $\rho_1 = \cdots = \rho_k > \rho_{k+1} \geq \cdots \geq \rho_n$, then in the above computation we only have equality if $\alpha_{k+1} = \cdots = \alpha_n = 0$. Hence

$$\underline{x} = \alpha_1 \underline{u}_1 + \cdots + \alpha_k \underline{u}_k,$$

and so

$$A\underline{x} = \rho_1 \underline{x}.$$

□

Proposition 4.1.9. *Let A be a non-negative symmetric matrix. There exists a non-zero vector $\underline{x} = (x_1, \ldots, , x_n)$ for which $A\underline{x} = \rho_1 \underline{x}$ and $x_i \geq 0$ for all i.*

Proof. Let $\underline{u}_1 = (u_{11}, u_{12}, \ldots, u_{1n})$. Let us consider $\underline{x} = (|u_{11}|, |u_{12}|, \ldots, |u_{1n}|)$. Then $||\underline{x}|| = ||\underline{u}_1|| = 1$ and

$$\underline{x}^T A \underline{x} \geq \underline{u}_1^T A \underline{u}_1 = \rho_1.$$

Then $\underline{x}^T A \underline{x} = \rho_1$ and by the previous proposition we have $A\underline{x} = \rho_1 \underline{x}$. Hence $\underline{x}$ satisfies the conditions. □

Proposition 4.1.10. *Let G be a connected graph, and let A be its adjacency matrix. Then*

(a) If $A\underline{x} = \rho_1 \underline{x}$ and $\underline{x} \neq \underline{0}$ then no entries of $\underline{x}$ is 0.

(b) The multiplicity of ρ_1 is 1.

(c) If $A\underline{x} = \rho_1 \underline{x}$ and $\underline{x} \neq \underline{0}$ then all entries of $\underline{x}$ have the same sign.

(d) If $A\underline{x} = \rho \underline{x}$ for some ρ and $x_i \geq 0$, where $\underline{x} \neq \underline{0}$ then $\rho = \rho_1$.

Proof. (a) Let $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (|x_1|, \ldots, |x_n|)$. As before we have $||\underline{y}|| = ||\underline{x}||$, and

$$\underline{y}^T A \underline{y} \geq \underline{x}^T A \underline{x} = \rho_1 ||\underline{x}||^2 = \rho_1 ||\underline{y}||^2.$$

Hence

$$A\underline{y} = \rho_1 \underline{y}.$$

Let $H = \{i \mid y_i = 0\}$ and $V \setminus H = \{i \mid y_i > 0\}$. Assume for contradiction that H is not empty. Note that $V \setminus H$ is not empty either as $\underline{x} \neq \underline{0}$. On the other hand, there cannot be any edge between H and $V \setminus H$: if $i \in H$ and $j \in V \setminus H$ and $(i, j) \in E(G)$, then

$$0 = \rho_1 y_i = \sum_j a_{ij} y_j \geq y_j > 0$$

contradiction. But if there is no edge between H and $V\backslash H$ then G would be disconnected, which contradicts the condition of the proposition. So H must be empty.

(b) Assume that $A\underline{x}_1 = \rho_1\underline{x}_1$ and $A\underline{x}_2 = \rho_1\underline{x}_2$, where $\underline{x}_1$ and $\underline{x}_2$ are independent eigenvectors. Note that by part (a), the entries of $\underline{x}_1$ is not 0, so we can choose a constant c such that the first entry of $\underline{x} = \underline{x}_2 - c\underline{x}_1$ is 0. Note that $A\underline{x} = \rho_1\underline{x}$ and $\underline{x} \neq 0$ since $\underline{x}_1$ and $\underline{x}_2$ were independent. But then $\underline{x}$ contradicts part (a).

(c) If $A\underline{x} = \rho_1\underline{x}$, and $\underline{y} = (|x_1|, \dots, |x_n|)$ then we have seen before that $A\underline{y} = \rho_1\underline{y}$. By part (b), we know that $\underline{x}$ and $\underline{y}$ must be linearly dependent so $\underline{x} = \underline{y}$ or $\underline{x} = -\underline{y}$. Together with part (a), namely that there is no 0 entry, this proves our claim.

(d) Let $A\underline{u}_1 = \rho_1\underline{u}_1$. By part (c), all entries have the same sign, we can choose it to be positive by replacing $\underline{u}_1$ with $-\underline{u}_1$ if necessary. Assume for contradiction that $\rho \neq \rho_1$. Note that if $\rho \neq \rho_1$ then $\underline{x}$ and $\underline{u}_1$ are orthogonal, but this cannot happen as all entries of both $\underline{x}$ and $\underline{u}_1$ are non-negative, further they are positive for $\underline{u}_1$, and $\underline{x} \neq \underline{0}$. This contradiction proves that $\rho = \rho_1$.

□

So part (c) enables us to recognize the largest eigenvalue from its eigenvector: this is the only eigenvector consisting of only positive entries (or actually, entries of the same sign).

Proposition 4.1.11. *(a) Let H be a subgraph of G. Then $\rho_1(H) \leq \rho_1(G)$.*
(b) Further, if G is connected and H is a proper subgraph then $\rho_1(H) < \rho_1(G)$.

Proof. (a) Let $\underline{x}$ be an eigenvector of length 1 of the adjacency matrix of H such that it has only non-negative entries. Then

$$\rho_1(H) = \underline{x}^T A(H)\underline{x} \leq \underline{x}^T A(G)\underline{x} \leq \max_{||\underline{z}||=1} \underline{z}^T A(G)\underline{z} = \rho_1(G).$$

In the above computation, if H has less number of vertices than G, then we complete $\underline{x}$ with 0's in the remaining vertices and we denote the obtained vector with $\underline{x}$ too in order to make sense for $\underline{x}^T A(G)\underline{x}$.

(b) Assume for contradiction that $\rho_1(H) = \rho_1(G)$. Then we have equality everywhere in the above computation. In particular $\underline{x}^T A(G)\underline{x} = \rho_1(G)$. This means that $\underline{x}$ is eigenvector of $A(G)$ too. Since G is connected $\underline{x}$ must be a (or rather "the") vector with only positive entries by part (a) of the above proposition. But then $\underline{x}^T A(H)\underline{x} < \underline{x}^T A(G)\underline{x}$, a contradiction. □

Proposition 4.1.12. *(a) We have* $|\rho_n| \leq \rho_1$.
(b) Let G *be a connected graph and assume that* $-\rho_n = \rho_1$. *Then* G *is bipartite.*
(c) G *is a bipartite graph if and only if its spectrum is symmetric to* 0.

Proof. (a) Let $\underline{x} = (x_1, \dots, x_n)$ be a unit eigenvector belonging to ρ_n, and let $\underline{y} = (|x_1|, \dots, |x_n|)$. Then

$$|\rho_n| = \left|\underline{x}^T A \underline{x}\right| = \left|\sum a_{ij} x_i x_j\right| \leq \sum a_{ij} |x_i||x_j| = \underline{y}^T A \underline{y} \leq \max_{||\underline{z}||=1} \underline{z}^T A \underline{z} = \rho_1.$$

(Another solution can be given based on the observation that $0 \leq TrA^\ell = \sum \rho_i^\ell$. If $|\rho_n| > \rho_1$ then for large enough odd ℓ we get that $\sum \rho_i^\ell < 0$.) (b) Since $\rho_1 \geq \dots \geq \rho_n$, the condition can only hold if $\rho_1 \geq 0 \geq \rho_n$. Again let $\underline{x} = (x_1, \dots, x_n)$ be a unit eigenvector belonging to ρ_n, and let $\underline{y} = (|x_1|, \dots, |x_n|)$. Then

$$\rho_1 = |\rho_n| = \left|\underline{x}^T A \underline{x}\right| = \left|\sum a_{ij} x_i x_j\right| \leq \sum a_{ij} |x_i||x_j| = \underline{y}^T A \underline{y} \leq \max_{||\underline{z}||=1} \underline{z}^T A \underline{z} = \rho_1.$$

We need to have equality everywhere. In particular, $\underline{y}$ is the positive eigenvector belonging to ρ_1, and all $a_{ij} x_i x_j$ have the same signs which can be only negative since $\rho_n \leq 0$. Hence every edges must go between the sets $V^- = \{i \mid x_i < 0\}$ and $V^+ = \{i \mid x_i > 0\}$. This means that G is bipartite.

(c) First of all, if G is a bipartite graph with color classes A and B then the following is a linear bijection between the eigenspace of the eigenvalue ρ and the eigenspace of the eigenvalue $-\rho$: if $A\underline{x} = \rho\underline{x}$ then let $\underline{y}$ be the vector which coincides with $\underline{x}$ on A, and -1 times $\underline{x}$ on B. It is easy to check that this will be an eigenvector belonging to $-\rho$.

Next assume that the spectrum is symmetric to 0. We prove by induction on the number of vertices that G is bipartite. Since the spectrum of the graph G is the union of the spectrum of the components there must be a component H with smallest eigenvalue $\rho_n(H) = \rho_n(G)$. Note that $\rho_1(G) = |\rho_n(G)| = |\rho_n(H)| \leq \rho_1(H) \leq \rho_1(G)$ implies that $\rho_1(H) = -\rho_n(H)$. Since H is connected we get that H is bipartite and its spectrum is symmetric to 0. Then the spectrum of $G \setminus H$ has to be also symmetric to 0. By induction $G \setminus H$ must be bipartite. Hence G is bipartite. □

Proposition 4.1.13. *Let* Δ *be the maximum degree, and let* $\overline{d}$ *denote the average degree. Then*

$$\max(\sqrt{\Delta}, \overline{d}) \leq \rho_1 \leq \Delta.$$

Proof. Let $\underline{v} = (1, 1, \dots, 1)$. Then

$$\rho_1 \geq \frac{\underline{v}^T A \underline{v}}{||\underline{v}||^2} = \frac{2e(G)}{n} = \overline{d}.$$

If the largest degree is Δ then G contains $K_{1,\Delta}$ as a subgraph. Hence

$$\rho_1(G) \geq \rho_1(K_{1,\Delta}) = \sqrt{\Delta}.$$

Finally, let $\underline{x}$ be an eigenvector belonging to ρ_1. Let x_i be the entry with largest absolute value. Then

$$|\rho_1||x_i| = \left|\sum_j a_{ij}x_j\right| \leq \sum_j a_{ij}|x_j| \leq \sum_j a_{ij}|x_i| \leq \Delta|x_i|.$$

Hence $\rho_1 \leq \Delta$. $\square$

Proposition 4.1.14. *Let G be a d-regular graph. Then $\rho_1 = d$ and its multiplicity is the number of components. Every eigenvector belonging to d is constant on each component.*

Proof. The first statement already follows from the previous propositions, but it also follows from the second statement so let us prove this statement. Let $\underline{x}$ be an eigenvector belonging to d. We show that it is constant on a connected component. Let H be a conncted component, and let $c = \max_{i \in V(H)} x_i$, let $V_c = \{i \in V(H) \mid x_i = c\}$ and $V(H) \setminus V_c = \{i \in V(H) \mid x_i < c\}$. If $V(H) \setminus V_c$ were not empty then there exists an edge $(i,j) \in E(H)$ such that $i \in V_c$, $j \in V(H) \setminus V_c$. Then

$$dc = dx_i = \sum_{k \in N(i)} x_k \leq x_j + \sum_{k \in N(i), k \neq j} x_k < c + (d-1)c = dc,$$

contradiction. So $\underline{x}$ is constant on each component. $\square$

Proposition 4.1.15. *(a) Assume that the edges of the graph G has an orientation such that every vertex has indegree at most D^+, and outdegree at most D^-. Then the following inequality holds true for the largest eigenvalue $\rho_1(G)$ of the adjacency matrix of G:*

$$\rho_1(G) \leq 2\sqrt{D^+D^-}.$$

(b) Let T be a tree on at least 3 vertices with maximum degree D. Show that for the largest eigenvalue $\rho_1(T)$ of the adjacency matrix of T we have

$$\rho_1(T) \leq 2\sqrt{D-1}.$$

Proof. (a) Let $\underline{x} = (x_1, \dots, x_n)$ be a non-negative eigenvector belonging to $\rho_1(G)$. Furthermore, let d_i^+ denote the in-degree of vertex i and d_j^- denote the out-degree of vertex j. Let us consider the following inequalities: let $(i,j) \in E(G)$ be an edge such that j is oriented towards i:

$$\frac{2x_ix_j}{\sqrt{D^+D^-}} \leq \frac{2x_ix_j}{\sqrt{d_i^+d_j^-}} \leq \frac{x_i^2}{d_i^+} + \frac{x_j^2}{d_j^-}.$$

Now let us add these inequalities for all (oriented) edges. Then

$$\frac{1}{\sqrt{D^+D^-}}\underline{x}^TA\underline{x} \leq \sum_{(i,j) \in E(G)} \left(\frac{x_i^2}{d_i^+} + \frac{x_j^2}{d_j^-}\right) = \sum_i x_i^2 + \sum_j x_j^2 = 2||x||^2.$$

Since $\underline{x}^T A \underline{x} = \rho_1(G)||x||^2$ we have $\rho_1(G) \leq 2\sqrt{D^+ D^-}$.

Second proof for part (a):

$$\rho_1(G) = 2 \sum_{(i,j)\in E(G)} x_i x_j = 2\sum_{i=1}^n x_i \left(\sum_{j:i\to j} x_j\right) \leq 2\sum_{i=1}^n x_i \sqrt{d_i^-}\left(\sum_{j:i\to j} x_j^2\right)^{1/2} \leq$$

$$\leq 2\sqrt{D^-}\sum_{i=1}^n x_i \left(\sum_{j:i\to j} x_j^2\right)^{1/2} \leq 2\sqrt{D^-}\left(\sum_{i=1}^n x_i^2\right)^{1/2}\left(\sum_{i=1}^n \sum_{j:i\to j} x_j^2\right)^{1/2} =$$

$$= 2\sqrt{D^-}\cdot 1 \cdot \left(\sum_{j=1}^n d_j^+ x_j^2\right)^{1/2} \leq 2\sqrt{D^-}\sqrt{D^+}\left(\sum_{j=1}^n x_j^2\right)^{1/2} = 2\sqrt{D^-}\sqrt{D^+}.$$

(b) Let us pick a vertex v of degree 1 and orient every edges towards this vertex. (Note that in a tree there is exactly one path between any two vertices so it makes sense to say such a thing that "orient every edges towards this vertex".) Then the out-degree of every vertex is exactly 1 except the chosen vertex v which is 0, consequently every in-degree is at most $D-1$ except the indegree of v which is 1. Note that since $n \geq 3$ we have $D \geq 2$, so $D-1 \geq 1$. By part (a) we have

$$\rho_1(G) \leq 2\sqrt{D^+D^-} = 2\sqrt{D-1}.$$

□

4.2 Matching polynomial

Definition 4.2.1. Let G be a graph on $v(G) = n$ vertices and let $m_k(G)$ denote the number of matchings of size k. Then the matching polynomial $\mu(G,x)$ is defined as follows:

$$\mu(G,x) = \sum_{k=0}^{\lfloor v/2\rfloor} (-1)^k m_k(G) x^{n-2k}.$$

Note that $m_0(G) = 1$. Another way to define $\mu(G,x)$ is the following:

$$\mu(G,x) = \sum_{M\in\mathcal{M}(G)} (-1)^{|M|} x^{n-2|M|}.$$

Sometimes it will be more convenient to work with the so-called matching generating function

$$M(G,\lambda) = \sum_{k=0}^{\lfloor v/2\rfloor} m_k(G)\lambda^k.$$

This is also the partition function of the monomer-dimer model at fugacity λ.

Clearly, the matching polynomial and the matching generating function encode the same information.

Proposition 4.2.2 ([19, 16]). *(a) Let* $u \in V(G)$ *then*

$$\mu(G,x) = x\mu(G-u,x) - \sum_{v \in N(u)} \mu(G-\{u,v\},x).$$

(b) For $e = (u,v) \in E(G)$ *we have*

$$\mu(G,x) = \mu(G-e,x) - \mu(G-\{u,v\},x).$$

(c) For $G = G_1 \cup G_2 \cup \cdots \cup G_k$ *we have*

$$\mu(G,x) = \prod_{i=1}^{k} \mu(G_i,x).$$

(d) We have

$$\mu'(G,x) = \sum_{u \in V(G)} \mu(G-u,x).$$

Proof. (a) By comparing the coefficient of x^{n-2k} we need to prove that

$$m_k(G) = m_k(G-u) + \sum_{v \in N(u)} m_{k-1}(G-\{u,v\}).$$

This is indeed true since we can count the number of k-matchings of G as follows: there $m_k(G-u)$ k-matchings which do not contain u, and if a k-matching contains u then there is a unique $v \in N(u)$ such that the edge (u,v) is in the matching, and the remaining $k-1$ edges are chosen from $G-\{u,v\}$.

(b) By comparing the coefficient of x^{n-2k} we need to prove that

$$m_k(G) = m_k(G-e) + m_{k-1}(G-\{u,v\}).$$

This is indeed true since the number of k-matchings not containing e is $m_k(G-e)$, and the number of k-matchings containing $e=(u,v)$ is $m_{k-1}(G-\{u,v\})$.

(c) It is enough to prove the claim when $G = G_1 \cup G_2$, for more components the claim follows by induction. By comparing the coefficient of x^{n-2k} we need to prove that

$$m_k(G) = \sum_{r=0}^{k} m_r(G_1)m_{k-r}(G_2).$$

This is indeed true since a k-matching fof G uniquely determine an r-matching of G_1 and a $(k-r)$-matching of G_2 for some $0 \le r \le k$.

(d) This follows from the fact that

$$(m_k(G)x^{n-2k})' = (n-2k)m_k(G)x^{n-1-2k} = \sum_{u \in V(G)} m_k(G-u)x^{n-1-2k}$$

since we can compute the cardinality of the set

$$\{(M,u) \mid u \notin V(M), |M| = k\}$$

in two different ways. $\square$

Theorem 4.2.3 (Heilmann and Lieb [19]). *All zeros of the matching polynomial $\mu(G,x)$ are real.*

Proof. We will prove the following two statements by induction on the number of vertices.

(i) All zeros of $\mu(G,x)$ are real.

(ii) For an x with $\mathrm{Im}(x) > 0$ we have

$$\mathrm{Im}\frac{\mu(G,x)}{\mu(G-u,x)} > 0$$

for all $u \in V(G)$.

Note that in (ii) we already use the claim (i) inductively, namely that $\mu(G-u,x)$ doesn't vanish for an x with $\mathrm{Im}x > 0$. On the other hand, claim (ii) for G implies claim (i). So we need to check claim (i).

By the recursion formula we have

$$\frac{\mu(G,x)}{\mu(G-u,x)} = \frac{x\mu(G-u,x) - \sum_{v\in N(u)} \mu(G-\{u,v\},x)}{\mu(G-u,x)} = x - \sum_{v\in N(u)} \frac{\mu(G-\{u,v\},x)}{\mu(G-u,x)}.$$

By induction we have

$$\mathrm{Im}\frac{\mu(G-u,x)}{\mu(G-\{u,v\},x)} > 0$$

for $\mathrm{Im}(x) > 0$. Hence

$$-\mathrm{Im}\frac{\mu(G-\{u,v\},x)}{\mu(G-u,x)} > 0$$

which gives that

$$\mathrm{Im}\frac{\mu(G,x)}{\mu(G-u,x)} > 0.$$

□

Remark 4.2.4. One can also deduce from this proof that the zeros of $\mu(G,x)$ and $\mu(G-u,x)$ interlace each other just like the zeros of a real-rooted polynomial and its derivative.

Definition 4.2.5. Let G be graph with a given vertex u. The *path-tree* $T(G,u)$ is defined as follows. The vertices of $T(G,u)$ are the paths[2] in G which start at the vertex u and two paths joined by an edge if one of them is a one-step extension of the other.

[2]In statistical physics, paths are called self-avoiding walks.

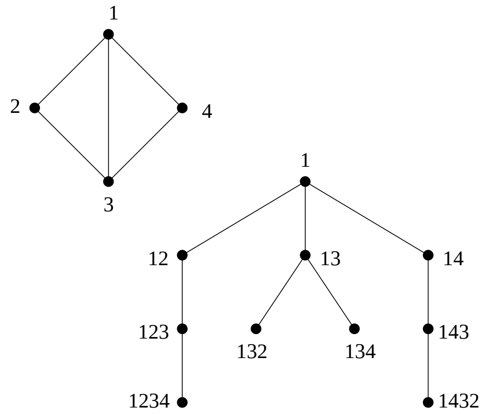


Figure 4.1: A path-tree from the vertex 1.

Proposition 4.2.6. *Let G be a graph with a root vertex u. Let $T(G,u)$ be the corresponding path-tree in which the root is again denoted by u for sake of convenience. Then*

$$\frac{\mu(G-u,x)}{\mu(G,x)}=\frac{\mu(T(G,u)-u,x)}{\mu(T(G,u),x)},$$

and $\mu(G,x)$ divides $\mu(T(G,u),x)$. In fact,

$$\mu(T(G,u),x)=\mu(G,x)\prod_{H}\mu(H,x)^{\alpha_H},$$

where H are induced subgraphs of G, and α_H is some non-negative integer.

Proof. The proof of this proposition is again by induction using part (a) of Proposition 4.2.2. Indeed,

$$\frac{\mu(G,x)}{\mu(G-u,x)}=\frac{x\mu(G-u,x)-\sum_{v\in N(u)}\mu(G-\{u,v\},x)}{\mu(G-u,x)}=$$

$$=x-\sum_{v\in N(u)}\frac{\mu(G-\{u,v\},x)}{\mu(G-u,x)}=x-\sum_{v\in N(u)}\frac{\mu(T(G-u,v)-v,x)}{\mu(T(G-u,v),x)}$$

$$=x\frac{\prod_{v\in N(u)}\mu(T(G-u,v),x)-\sum_{v\in N(u)}\mu(T(G-u,v)-v,x)\prod_{v'\in N(u)\setminus\{v\}}\mu(T(G-u,v'),x)}{\prod_{v\in N(u)}\mu(T(G-u,v),x)}$$

$$=\frac{x\mu(T(G,u)-u,x)-\sum_{v\in N(u)}\mu(T(G,u)-\{u,v\},x)}{\mu(T(G,u)-u,x)}=\frac{\mu(T(G,u),x)}{\mu(T(G,u)-u,x)}.$$

In the first step we used the recursion formula, and in the third step we used the induction step to the graph $G-u$ and root vertex v. Here it is an important observation that $T(G-u,v)$ is exactly the branch of the tree $T(G,u)$ that we get if we delete the vertex u from $T(G,u)$ and consider the subtree rooted at the path uv.

The second part of the claim follows by induction using

$$\mu(T(G,u),x)=\mu(G,x)\frac{\mu(T(G,u)-u,x)}{\mu(G-u,x)}$$

together with the observation that a connected component of $T(G,u)-u$, say rooted at the vertex uv, is path-tree of $G-u$ with root vertex v. □

Proposition 4.2.7 ([19, 16]). *For a forest T, the matching polynomial $\mu(T,x)$ coincides with the characteristic polynomial $\varphi(T,x)=\det(xI-A_T)$.*

Proof. Indeed, when we expand the $\det(xI-A)$ we only get non-zero terms when the cycle decomposition of the permutation consists of cycles of length at most 2; but these terms correspond to the terms of the matching polynomial.

□

Remark 4.2.8. Clearly, Propositions 4.2.6 and 4.2.7 together give a new proof of the Heilmann-Lieb theorem since $\mu(G,x)$ divides $\mu(T(G,u),x)=\varphi(T(G,u),x)$ whose zeros are real since they are the eigenvalues of a symmetric matrix.

Proposition 4.2.9 ([19, 16]). *If the largest degree Δ is at least 2, then all zeros of the matching polynomial lie in the interval $[-2\sqrt{\Delta-1}, 2\sqrt{\Delta-1}]$.*

First proof. First we show that if u is a vertex of degree at most $\Delta-1$, then for any $x\geq 2\sqrt{\Delta-1}$ we have

$$\frac{\mu(G,x)}{\mu(G-u,x)}\geq\sqrt{\Delta-1}.$$

We prove this statement by induction on the number of vertices. This is true if $G=K_1$, so we can assume that $v(G)\geq 2$. Then

$$\frac{\mu(G,x)}{\mu(G-u,x)}=\frac{x\mu(G-u,x)-\sum_{v\in N_G(u)}\mu(G-\{u,v\},x)}{\mu(G-u,x)}$$
$$=x-\sum_{v\in N_G(u)}\frac{\mu(G-\{u,v\},x)}{\mu(G-u,x)}\geq x-(\Delta-1)\frac{1}{\sqrt{\Delta-1}}\geq\sqrt{\Delta-1}.$$

We used the fact that $v\in N_G(u)$ has degree at most $\Delta-1$ in the graph $G-u$.

Then for any vertex u we have

$$\frac{\mu(G,x)}{\mu(G-u,x)}=\frac{x\mu(G-u,x)-\sum_{v\in N_G(u)}\mu(G-\{u,v\},x)}{\mu(G-u,x)}$$
$$=x-\sum_{v\in N_G(u)}\frac{\mu(G-\{u,v\},x)}{\mu(G-u,x)}\geq x-\Delta\frac{1}{\sqrt{\Delta-1}}>0$$

since $v\in N_G(u)$ has degree at most $\Delta-1$ in the graph $G-u$. This shows $\mu(G,x)\neq 0$ if $x\geq 2\sqrt{\Delta-1}$. Since the zeros of the matching polynomial are symmetric to 0 we get that all zeros lie in the interval $(-2\sqrt{\Delta-1}, 2\sqrt{\Delta-1})$. □

Second proof. By Propositions 4.2.6 and 4.2.7 we know that $\mu(G,x)$ divides $\mu(T(G,u),x)=\varphi(T(G,u),x)$. The largest degree of $T(G,u)$ is also at most Δ so by Proposition 4.1.15 we have $\rho_1(T(G,u))\leq 2\sqrt{\Delta-1}$. Since the zeros of the matching polynomial are symmetric to 0 we get that all zeros lie in the interval $[-2\sqrt{\Delta-1}, 2\sqrt{\Delta-1}]$. □

Proposition 4.2.10. *Let*

$$\frac{\mu(G-u,x)}{\mu(G,x)}=\sum_k a_k(G,u)x^{-(k+1)}.$$

Then $a_k(G,u)$ counts the number of closed walks of length k in the tree $T(G,u)$ from u to u.

Proof. This proposition follows from Proposition 4.2.6 and 4.2.7 and the fact that

$$\frac{\varphi(H-u,x)}{\varphi(H,x)}=\sum_k W_k(H,u)x^{-(k+1)},$$

where $W_k(H,u)$ counts the number of closed walks of length k from u to u in a graph H. Indeed,

$$\frac{\mu(G-u,x)}{\mu(G,x)}=\frac{\mu(T(G,u)-u,x)}{\mu(T(G,u),x)}=\frac{\varphi(T(G,u)-u,x)}{\varphi(T(G,u),x)}=\sum_k W_k(T(G,u),u)x^{-k}.$$

Here $W_k(T(G,u),u)=a_k(G,u)$ by definition. □

Remark 4.2.11. A walk in the tree $T(G,u)$ from u can be imagined as follows. Suppose that in the graph G a worm is sitting at the vertex u at the beginning. Then at each step the worm can either grow or pull back its head. When it grows it can move its head to a neighboring unoccupied vertex while keeping its tail at vertex u. At each step the worm occupies a path in the graph G. A closed walk in the tree $T(G,u)$ from u to u corresponds to the case when at the final step the worm occupies only vertex u. C. Godsil calls these walks tree-like walks in the graph G.

Proposition 4.2.12. *(a) Let*

$$\frac{\mu'(G,x)}{\mu(G,x)}=\sum_k a_k(G)x^{-(k+1)}.$$

Then $a_k(G)$ counts the number of closed tree-like walks of length k.

(b) If $\mu(G,x)=\prod_{i=1}^{v(G)}(x-\alpha_i)$ then for all $k\geq 1$ we have

$$a_k(G)=\sum_{i=1}^{v(G)}\alpha_i^k.$$

Theorem 4.2.13 ([19, 16])**.** *The following identities hold true.*

(a) We have

$$\frac{\mu(G-u,x)\mu(G,y)-\mu(G-u,y)\mu(G,x)}{y-x}=\sum_{P\in\mathcal{P}_u}\mu(G-P,x)\mu(G-P,y).$$

(b) We have

$$\mu(G-u,x)\mu(G-v,x)-\mu(G,x)\mu(G-\{u,v\},x)=\sum_{P\in\mathcal{P}_{u,v}}\mu(G-P,x)^2.$$

Proof. First we prove part (a). The claim can be checked for the one vertex graph by hand. Now suppose that the statemnet is true for all graphs with fewer vertices than G has. Then

$$\frac{\mu(G-u,x)\mu(G,y)-\mu(G-u,y)\mu(G,x)}{y-x}=$$

$$=\frac{1}{y-x}\mu(G-u,x)\left(y\mu(G-u,y)-\sum_{v\in N(u)}\mu(G-\{u,v\},y)\right)-$$
$$-\frac{1}{y-x}\mu(G-u,y)\left(x\mu(G-u,x)-\sum_{v\in N(u)}\mu(G-\{u,v\},x)\right)=$$
$$=\mu(G-u,x)\mu(G-u,y)+$$
$$+\frac{1}{y-x}\sum_{v\in N(u)}\left(\mu(G-u,x)\mu(G-\{u,v\},y)-\mu(G-u,y)\mu(G-\{u,v\},x)\right)=$$
$$=\mu(G-u,x)\mu(G-u,y)+\sum_{v\in N(u)}\sum_{P\in\mathcal{P}(G-u)_v}\mu((G-u)-P,x)\mu((G-u)-P,y)=$$
$$=\sum_{P\in\mathcal{P}_u}\mu(G-P,x)\mu(G-P,y).$$

In the third step we used induction to the graph $G-u$ and its special vertex v for all $v\in N(u)$.

Next we prove part (b). Recall that one can write $\mu(G,x)$ as follows:

$$\mu(G,x)=\sum_{M\in\mathcal{M}(G)}(-1)^{|M|}x^{n-2|M|}.$$

Hence

$$\mu(G-u,x)\mu(G-v,x)=\left(\sum_{M\in\mathcal{M}(G-u)}(-1)^{|M|}x^{n-1-2|M|}\right)\left(\sum_{N\in\mathcal{M}(G-v)}(-1)^{|N|}x^{n-1-2|N|}\right)$$
$$=\sum_{\substack{R=M\cup N\\ M\in\mathcal{M}(G-u),\ N\in\mathcal{M}(G-v)}}(-1)^{|R|}x^{2n-2-|R|}.$$

and

$$\mu(G,x)\mu(G-\{u,v\},x)=\left(\sum_{M\in\mathcal{M}(G)}(-1)^{|M|}x^{n-2|M|}\right)\left(\sum_{N\in\mathcal{M}(G-\{u,v\})}(-1)^{|N|}x^{n-2-2|N|}\right)$$
$$=\sum_{\substack{R=M\cup N\\ M\in\mathcal{M}(G),\ N\in\mathcal{M}(G-\{u,v\})}}(-1)^{|R|}x^{2n-2-|R|}.$$

A configuration R which is the union of two matchings is always the union of paths, cycles of even length and "double edges" (an edge being in both M and N). Note that if u or

v is covered by R then they are necessarily the end vertices of a path in both cases. Let $c(R)$ denote the number of connected components of R different from a double edge. In what follows: $v(P)$ denotes the number of vertices of a path, and $|P|$ denotes the number of edges of P. Clearly, $|P| = v(P) - 1$.

We would like to count how many ways we can get such a configuration R in the first and in the second case. We need to distinguish six cases: (1) u and v are not covered by R, (2) u is covered by R, but v is not covered by R, (3) v is covered by R, but u is not covered by R, (4) u and v are both covered by R, but they are not in the same connected component, (5) u and v are both covered by R, and they are in the same connected component, a path P, which has even number of edges, (6) u and v are both covered by R, and they are in the same connected component, a path P, which has even number of edges. It is easy to check that in case 1 the configuration R can be obtained in $2^{c(R)}$ ways in both sums as we have two choices to decompose a component to two matchings. In case 2 and case 3 we have $2^{c(R)-1}$ choices in both cases as we have 2 choices if the component does not contain u or v, and only one choice at the component covering u or v. In case 4 we have $2^{c(R)-2}$ choices in both cases. In case 5 there are $2^{c(R)-1}$ ways to decompose R such that $R = M \cup N$ and M does not cover u, N does not cover v, but there is no decomposition for $M \cup N$ such that N does not cover neither u nor v. In case 6 the situation is opposite: there are $2^{c(R)-1}$ ways to decompose R such that $R = M \cup N$ and N does not cover neither u nor v, but there is no decomposition for $M \cup N$ such that M does not cover u and N does not cover v.

Let $\mathcal{R}$ be the set of configurations R such that R has a connected component P whose end vertices are u and v. According to the previous paragraph we have

$$\mu(G-u,x)\mu(G-v,x) - \mu(G,x)\mu(G-\{u,v\},x) = \sum_{R \in \mathcal{R}} 2^{c(R)-1}(-1)^{|P_{u,v}|}(-1)^{|R|}x^{2n-2-|R|}.$$

Now let us turn to the other side:

$$\begin{aligned}
\sum_{P \in \mathcal{P}_{u,v}} \mu(G-P,x)^2 &= \sum_{P \in \mathcal{P}_{u,v}} \left(\sum_{M \in \mathcal{M}(G-P)} (-1)^{|M|} x^{n-v(P)-2|M|} \right)^2 \\
&= \sum_{P \in \mathcal{P}_{u,v}} \sum_{M,N \in \mathcal{M}(G-P)} (-1)^{|M|+|N|} x^{2n-2v(P)-2|M|-2|N|} \\
&= \sum_{P \in \mathcal{P}_{u,v}} \sum_{M,N \in \mathcal{M}(G-P)} (-1)^{|M|+|N|} x^{2n-2(|P|+1)-2|M|-2|N|} \\
&= \sum_{\substack{R = P \cup M \cup N \\ P \in \mathcal{P}_{u,v}}} (-1)^{|P|}(-1)^{|R|} x^{2n-2-|R|} \\
&= \sum_{R \in \mathcal{R}} 2^{c(R)-1}(-1)^{|P_{u,v}|}(-1)^{|R|}x^{2n-2-|R|}.
\end{aligned}$$

Hence

$$\mu(G-u,x)\mu(G-v,x)-\mu(G,x)\mu(G-\{u,v\},x)=\sum_{P\in\mathcal{P}_{u,v}}\mu(G-P,x)^2.$$

□

Remark 4.2.14. From part (a) one can give a third proof for the real-rootedness of the matching polynomial. Namely, suppose for contradiction that there is a graph G with non-real zeros z and $\overline{z}$, take the smallest such graph. Apply the identity of part (a) to $x=z$ and $y=\overline{z}$, then the left hand side is 0 while the right hand side is the sum of terms $|\mu(G-P,z)|^2$, and they are non-zero, because $G-P$ has no non-real zero, contradiction.

4.3 Chromatic polynomial

Definition 4.3.1. Let G be a graph. A map $\varphi: V(G)\to\{1,2,\dots,q\}$ is a proper coloring with q colors if $\varphi(u)\neq\varphi(v)$ whenever $(u,v)\in E(G)$.

The number of proper colorings of G with q colors is denoted by $\mathrm{ch}(G,q)$.

Proposition 4.3.2. *The function* $\mathrm{ch}(G,q)$ *is polynomial in* q*.*

Proof. If we use exactly k colors then it corresponds to a decomposition of the vertex set into k independent sets. So if $a_k(G)$ denotes the number of decompositions of the vertex set into k independent sets then there are $a_k(G)q(q-1)\dots(q-k+1)$ such proper colorings. Hence

$$\mathrm{ch}(G,q)=\sum_{k=1}^{n}a_k(G)q(q-1)\dots(q-k+1).$$

This is clearly a polynomial in q. □

Proposition 4.3.3. *If* $e\in E(G)$ *is a loop, then* $\mathrm{ch}(G,q)=0$*. If* $e\in E(G)$ *is a bridge, then* $\mathrm{ch}(G,q)=\frac{q-1}{q}\mathrm{ch}(G-e,q)$*. If* $e\in E(G)$ *is neither loop, nor bridge, then we have*

$$\mathrm{ch}(G,q)=\mathrm{ch}(G-e,q)-\mathrm{ch}(G/e,q),$$

where G/e *denotes the graph obtained from* G *by contracting the edge* e*.*

Proof. Let us consider the proper colorings of $G-e$. If $e=(u,v)$ then we can distinguish two cases: u and v get different colors then it is even a proper coloring of G. If u and v get the same color then it corresponds to a proper coloring of G/e. Hence

$$\mathrm{ch}(G-e,q)=\mathrm{ch}(G,q)+\mathrm{ch}(G/e,q).$$

□

Proposition 4.3.4. *For an $A \subseteq E(G)$ let $k(A)$ denote the number of components of the graph $(V(G), A)$. Then*

$$\mathrm{ch}(G,q) = \sum_{A \subseteq E(G)} (-1)^{|A|} q^{k(A)}.$$

Proof. Let $\mathcal{B}$ the set of all colorings (not just the proper ones) of $V(G)$. Let $\mathcal{B}_e$ the set of all colorings of $V(G)$, where the end vertices of e get the same color. Then the number of proper colorings of G is

$$\mathrm{ch}(G,q) = |\mathcal{B} \setminus \cup_{e \in E(G)} \mathcal{B}_e|.$$

By inclusion-exclusion principle we have

$$|\mathcal{B} \setminus \cup_{e \in E(G)} \mathcal{B}_e| = |\mathcal{B}| - \sum_{e \in E(G)} |\mathcal{B}_e| + \sum_{e_1, e_2 \in E(G)} |\mathcal{B}_{e_1} \cap \mathcal{B}_{e_2}| - \dots .$$

Note that for some $A \subseteq E(G)$ we have

$$| \cap_{e \in A} \mathcal{B}_e| = q^{k(A)}.$$

Hence

$$\mathrm{ch}(G,q) = \sum_{A \subseteq E(G)} (-1)^{|A|} q^{k(A)}.$$

□

Proposition 4.3.5. *Let*

$$\mathrm{ch}(G,q) = \sum_{k=0}^{n-1} (-1)^k c_{n-k} q^{n-k}.$$

Then $c_i \geq 0$.

Proof. We prove this claim by induction on the number of edges. For the empty graph O_n on n vertices we have $\mathrm{ch}(O_n, q) = q^n$, so the claim is true. Then from

$$\mathrm{ch}(G,q) = \mathrm{ch}(G-e,q) - \mathrm{ch}(G/e)$$

we have

$$c_{n-k}(G) = c_{n-k}(G-e) + c_{n-k}(G/e).$$

(Note that G/e has $n-1$ vertices!) Hence by induction

$$c_{n-k}(G) = c_{n-k}(G-e) + c_{n-k}(G/e) \geq 0.$$

□

Remark 4.3.6. June Huh proved that the sequence $(c_k(G))_{k=1}^n$ is log-concave, consequently unimodal. The coefficient $c_{n-k}(G)$ has also a combinatorial meaning: it is the number of edge sets of size k not containing any broken cycle. A broken cycle is defined as follows: take any ordering of the edges, and a broken cycle is a cycle minus the highest index edge.

Theorem 4.3.7 (Alan Sokal [33]). *Let G be a graph of largest degree Δ. Let* $\mathrm{ch}(G, z)$ *be the chromatic polynomial of the graph G. Then the absolute value of any zero of* $\mathrm{ch}(G, z)$ *is at most $c\Delta$ where c is a fix constant not depending on G or Δ. The constant c is at most* 8.

5. Empirical Measures

The main idea of this chapter is the following: if there is some polynomial $f(G,z)$ of degree $v(G)$ associated to the graph G, then we can consider

$$f(G,z)=\prod_{i=1}^{v(G)}(z-\alpha_i),$$

and the corresponding measure on $\mathbb{C}$:

$$\rho_G=\frac{1}{v(G)}\sum_{i=1}^{v(G)}\delta_{\alpha_i},$$

where δ_α is the Dirac-delta measure with mass 1 on α. So ρ_G is the probability measure on $\mathbb{C}$ given by the uniform distribution on the zeros of the polynomial $f(G,z)$. Recall that

$$\int h(z)\,d\rho_G(z)=\frac{1}{v(G)}\sum_{i=1}^{v(G)}h(\alpha_i).$$

Now if we were able to show that for every fix k and every Benjamini–Schramm convergent graph sequence (G_i) the sequence

$$p_k(G_i):=\frac{1}{v(G_i)}\sum_{j=1}^{v(G_i)}\alpha_j^k=\int z^k\,d\rho_{G_i}(z)$$

is convergent then we get automatically the convergence of those graph parameters $p(G)$ for which there is an $h(z)$ such that

$$p(G)=\int h(z)\,d\rho_G(z),$$

where $h(z)$ is a polynomial or –and this is important– can be approximated by a polynomial. (Indeed, if $h(z)=\sum_{k=0}^m c_kz^k$ then $\int h(z)\,d\rho_G(z)=\sum_{k=0}^m c_kp_k(G)$, and so the convergence of the sequences $(p_k(G_i))_{i=1}^\infty$ implies the convergence of $(p(G_i))_{i=1}^\infty$.)

For the sake of simplicity we will say that the measures ρ_{G_i} are holomorphic convergent if for every k the sequence

$$p_k(G_i)=\int z^k\,d\rho_{G_i}(z)$$

is convergent.

For further reference we study certain graph polynomials in more detail.

Example 5.0.1. Let $\phi(G, x)$ be the characteristic polynomial of the adjacency matrix of the graph G. Let $\rho_G^{(s)}$ be probability measure on $\mathbb{C}$ given by the uniform distribution on its zeros, i. e., the eigenvalues of the graph G. This is the so-called empirical spectral measure of the graph G.

Example 5.0.2. Let $L(G, x)$ be the characteristic polynomial of the Laplacian matrix of the graph G. Let $\rho_G^{(\ell)}$ be probability measure on $\mathbb{C}$ given by the uniform distribution on its zeros, i. e., the Laplacian eigenvalues of the graph G. This is the so-called empirical Laplacian spectral measure of the graph G.

Example 5.0.3. Let $\mu(G, x)$ be the matching polynomial of the graph G. Let $\rho_G^{(m)}$ be probability measure on $\mathbb{C}$ given by the uniform distribution on its zeros. This is the so-called matching measure of the graph G.

Example 5.0.4. Let $\mathrm{ch}(G, x)$ be the chromatic polynomial of the graph G. Let $\rho_G^{(c)}$ be probability measure on $\mathbb{C}$ given by the uniform distribution on its zeros. This is the so-called chromatic measure of the graph G.

5.1 Benjamini–Schram convergence and convergence of measures

In this section we study the so-called Benjamini–Schramm convergence.

Definition 5.1.1. Let L be a probability distribution on (infinite) connected rooted graphs; we will call L a *random rooted graph*. For a finite connected rooted graph α and a positive integer r, let $\mathbb{P}(L, \alpha, r)$ be the probability that the r-ball centered at a random root vertex chosen from the distribution L is isomorphic to α.

For a finite graph G, a finite connected rooted graph α and a positive integer r, let $\mathbb{P}(G, \alpha, r)$ be the probability that the r-ball centered at a uniform random vertex of G is isomorphic to α.

We say that a bounded-degree graph sequence (G_i) is *Benjamini–Schramm convergent* if for all finite rooted graphs α and $r > 0$, the probabilities $\mathbb{P}(G_i, \alpha, r)$ converge. Furthermore, we say that (G_i) *Benjamini-Schramm converges to* L, if for all positive integers r and finite rooted graphs α, $\mathbb{P}(G_i, \alpha, r) \to \mathbb{P}(L, \alpha, r)$.

The Benjamini–Schramm convergence is also called *local convergence* as it primarily grasps the local structure of the graphs (G_i).

Example 5.1.2. Note that if (G_i) is a sequence of d–regular graphs such that the girth $g(G_i) \to \infty$ then it is Benjamini–Schramm convergent and we can even see its limit object: the rooted infinite d-regular tree $\mathbb{T}_d$, so the corresponding random rooted graph L is simply the distribution which takes a rooted infinite d-regular tree with probability

1. When L is a certain rooted infinite graph with probability 1 then we simply say that this rooted infinite graph is the limit without any further reference on the distribution.

Example 5.1.3. There are other very natural graph sequences which are Benjamini–Schramm convergent, for instance if we take larger and larger boxes in the d-dimensional grid $\mathbb{Z}^d$ then it will converge to the rooted $\mathbb{Z}^d$.

Before we proceed let us point out one important property of the Benjamini–Schramm convergence. It turns out that a graph sequence (G_i) is Benjamini–Schramm convergent if and only if for every **connected** graph H the sequence

$$\frac{\hom(H,G_i)}{v(G_i)}$$

is convergent. Intuitively it is clear why these sequences are convergent: if H has k vertices and we know the statistics of all k balls around the vertices then we are able to count the number of homomorphisms from H to G_i, because H is connected. So the functions $p_H(.) = \hom(H,.)/v(.)$ play the role of coordinating functions of this convergence. It turns out that instead of $\hom(H,G_i)$ one can use $\mathrm{ind}(H,G_i)$, the number of induced homomorphisms, or simply $\mathrm{sub}(H,G_i)$, the number of induced subgraphs of G_i isomorphic to H. In fact, later it will be indeed more convenient to work with the quantities $\mathrm{sub}(H,G_i)$.

Theorem 5.1.4. *If $(G_i)_i$ is a sequence of Benjamini–Schramm convergent graphs, then the sequence of spectral measures $\rho^{(s)}_{G_i}$ converges weakly.*

Proof. Since we defined Benjamini–Schramm convergence only for bounded degree graph sequences we can assume that Δ is an upper bound for all degrees of the graphs $(G_i)_i$. By Propositions 4.1.12 and 4.1.13 this implies that all eigenvalues of all $(G_i)_i$ lie in the interval $[-\Delta,\Delta]$. We need to prove that for every bounded continuous function $f(z)$, the sequence of integrals

$$\int f(z)\, d\rho^{(s)}_{G_i}(z)$$

converges. By Weierstrass' theorem every continuous function $f(z)$ can be approximated in maximum norm by a polynomial in a bounded interval. This implies that it is enough to check that for every fixed k the sequence

$$\int z^k\, d\rho^{(s)}_{G_i}(z)$$

converges. Let $\rho_1,\dots,\rho_{v(G)}$ be the eigenvalues of the graph G. Note that

$$\int z^k\, d\rho^{(s)}_{G}(z) = \frac{1}{v(G)}\sum_{k=1}^{v(G)} \rho_i^k$$

by definition. By Proposition 4.1.2 the $\sum_{k=1}^{v(G)} \rho_i^k = W_k(G)$ counts the number of closed walks of length k. In return the quantity $\frac{W_k(G)}{v(G)}$ can be determined by the statistics of the k-balls: in a rooted graph α let $W_k(\alpha)$ be the number of closed walks of length k from the root. Then

$$\frac{W_k(G)}{v(G)} = \sum_{\alpha} \mathbb{P}_G(B_k(o) \simeq \alpha) W_k(\alpha).$$

Now we see that if for every fixed rooted graph α the sequence $\mathbb{P}_{G_i}(B_k(o) \simeq \alpha)$ converges then so the sequence $\frac{W_k(G_i)}{v(G_i)}$. Hence the sequence

$$\int z^k \, d\rho_{G_i}^{(s)}(z)$$

converges. This finishes the proof. □

The following theorem can be proved exactly the same way. The only difference that instead of Propositions 4.1.12 and 4.1.13 we need to cite Propositions 4.2.9 ensuring that all zeros of all matching polynomials lie in the interval $[-2\sqrt{\Delta - 1}, 2\sqrt{\Delta - 1}]$, and instead of Proposition 4.1.2 we need to cite Proposition 4.2.12 concerning the combinatorial meaning of the power sums of the zeros of the matching polynomial.

Theorem 5.1.5. *If $(G_i)_i$ is a sequence of Benjamini–Schramm convergent graphs, then the sequence of matching measures $\rho_{G_i}^{(m)}$ converges weakly.*

5.2 Moments and homomorphism numbers

In the previous section we have seen that knowing the combinatorial meaning of the power sum of the zeros of a graph polynomial can be used to establishing convergence results for empirical measures. Since it is rather ad hoc whether we can find such a combinatorial meaning it would be desirable to find something more robust way to prove such convergence results. This is the scope of this section. We will see that practically the multiplicativity of the studied graph polynomial ensures such a convergence result.

Theorem 5.2.1. *Let f be a monic multiplicative graph polynomial of exponential type. Assume that f has bounded roots.*

Let (G_i) be a Benjamini–Schramm convergent graph sequence. Let $K \subset \mathbb{C}$ be a compact set containing all roots of $f(G_i, x)$ for all i, such that $\mathbb{C} \setminus K$ is connected.

(a) *For a graph G, let ρ_G be the uniform distribution on the roots of $f(G, x)$. Then for every continuous function $g : K \to \mathbb{R}$ that is harmonic on the interior of K, the sequence*

$$\int_K g(z) d\rho_{G_n}(z)$$

converges.

Moreover, for any open set $\Omega \subseteq \mathbb{R}^d$ and any continuous function $g : K \times \Omega \to \mathbb{R}$ that is harmonic on the interior of K for any fixed $\xi \in \Omega$ and harmonic on Ω for any fixed $z \in K$, the sequence

$$\int_K g(z, \xi) d\rho_{G_n}(z)$$

converges, locally uniformly in $\xi \in \Omega$, to a harmonic function on Ω.

(b) *For $\xi \in \mathbb{C} \setminus K$, let*

$$t_n(\xi) = \frac{\log |f(G_n, \xi)|}{v(G_n)}.$$

Then $t_n(\xi)$ converges to a harmonic function locally uniformly on $\mathbb{C} \setminus K$.

The integral in (a) is called a *harmonic moment* of the roots of $f(G_i, z)$. Its convergence is referred to as the 'Benjamini–Schramm continuity' or 'estimability' of the moment. The fraction in (b) is called the *entropy per vertex* or the *free energy* at ξ.

Note that part (b) immediately follows from part (a) since

$$t_n(\xi) = \frac{\log |f(G_n, \xi)|}{|V(G_n)|} = \frac{1}{v(G_n)} \sum_{i=1}^{v(G_n)} \log |\xi - \alpha_i(G_n)| = \int \log |\xi - z| d\rho_{G_n}(z)$$

and $\log |\xi - z|$ is a harmonic function on the interior of K and continuous on K for $\xi \notin K$.

The phenomenon described in Theorem 5.2.1 was discovered, in the case of the chromatic polynomial, by M. Abért and T. Hubai. The main achievement of their paper [1] was to state and prove Theorem 5.2.1 for the chromatic polynomial (in a slightly different form: they used a disc in place of our set K, they used holomorphic functions in place of our harmonic ones, they took $\log f$ in place of $\log |f|$ in (b) — all these are inessential differences). They thus answered a question of Borgs and generalized a result of Borgs, Chayes, Kahn and Lovász [4], who had proved convergence of $(\log \mathrm{ch}(G_n, q))/v(G_n)$ at large positive integers q.

One may naively hope for the weak convergence of the measures μ_{G_i} arising from the roots, i. e., one may hope that for any bounded continuous function $h(z)$ the sequence

$$\int h(z)\, d\mu_{G_i}(z)$$

converges. As remarked by Abért and Hubai, this is easily seen to be false already for the chromatic polynomial. Indeed, paths and circuits together form a Benjamini–Schramm convergent sequence, but the chromatic measures of paths tend to the Dirac measure at 1, and the chromatic measures of circuits tend to the uniform measure on the unit circle centered at 1. So it was a crucial observation that the useful relaxation is to consider the convergence of the holomorphic moments.

To prove the convergence of the holomorphic moments, Abért and Hubai [1] showed that for a finite graph G and for every k, the power sum

$$v(G)p_k(G) = \sum_{i=1}^{v(G)} \alpha_i^k = v(G) \int_K z^k d\rho_G(z)$$

can be expressed as a fixed linear combination of the functions $\hom(H_i, G)$, where the H_i are non-empty **connected** finite graphs. In terms of our remark about the functions $\hom(H, .)/v(.)$ as being coordinate functions for the Benjamini–Schramm convergence, this immediately implies the convergence of $(p_k(G_i))$ for Benjamini–Schramm convergent sequence (G_i).

To show this, they determined an exact expression for the power sum $v(G)p_k(G)$ of the zeros in terms of homomorphism numbers. Our approach is a bit different: without determining the exact expression, we use multiplicativity of the graph polynomial to show that the coefficient of $\mathrm{sub}(H, G)$ for non-connected H must be 0. (In fact, one can determine the exact expression too with a little extra work.) Below we give the heart of the argument.

Theorem 5.2.2. *Let $f(G, z)$ be a monic multiplicative graph polynomial, and assume that for every fixed k there is a finite set $\mathcal{H}_k$ of finite graphs and constants $c_k(H)$ such that for every graph G we have*

$$\sum_{i=1}^{v(G)} \alpha_i^k = \sum_{H \in \mathcal{H}_k} c_k(H)\mathrm{sub}(H, G),$$

where $\alpha_1, \ldots, \alpha_{v(G)}$ are the zeros of $f(G, z)$. Then $c_k(H) = 0$ whenever H is not connected.

Proof. Suppose for contradiction that there is some disconnected $H \in \mathcal{H}_k$ such that $c_k(H) \neq 0$. Choose such a disconnected H with smallest number of vertices, and suppose that $H = H_1 \cup H_2$, where H_1 and H_2 are non-empty graphs. Let

$$f(H_1, z) = \prod_{i=1}^{v(H_1)} (z - \alpha_i(H_1)),$$

and similarly,

$$f(H_2, z) = \prod_{i=1}^{v(H_2)} (z - \alpha_i(H_2)),$$

$$f(H, z) = \prod_{i=1}^{v(H)} (z - \alpha_i(H)),$$

Then by the multiplicativity of the polynomial $f(G, z)$ we have

$$f(H, z) = f(H_1, z)f(H_2, z) = \prod_{i=1}^{v(H_1)} (z - \alpha_i(H_1)) \cdot \prod_{i=1}^{v(H_2)} (z - \alpha_i(H_2)).$$

Hence
$$\sum_{i=1}^{v(H)} \alpha_i(H)^k = \sum_{i=1}^{v(H_1)} \alpha_i(H_1)^k + \sum_{i=1}^{v(H_2)} \alpha_i(H)^k.$$
On the other hand,
$$\sum_{i=1}^{v(H)} \alpha_i(H)^k = \sum_{H' \in \mathcal{H}_k} c_k(H')\text{sub}(H', H),$$
and
$$\sum_{i=1}^{v(H_1)} \alpha_i(H_1)^k + \sum_{i=1}^{v(H_2)} \alpha_i(H)^k = \sum_{H' \in \mathcal{H}_k} c_k(H')(\text{sub}(H', H_1) + \text{sub}(H', H_2)).$$
Hence
$$\sum_{H' \in \mathcal{H}_k} c_k(H')\text{sub}(H', H) = \sum_{H' \in \mathcal{H}_k} c_k(H')(\text{sub}(H', H_1) + \text{sub}(H', H_2)).$$
Now if H' is connected then we have
$$\text{sub}(H', H) = \text{sub}(H', H_1) + \text{sub}(H', H_2).$$
Note that if $H' \in \mathcal{H}_k$ and H' is disconnected then $v(H') \geq v(H)$ by the choice of H, and if furthermore $H' \not\simeq H$ then $\text{sub}(H', H) = 0$ just as $\text{sub}(H', H_1) = \text{sub}(H', H_2) = 0$ since $v(H') \geq v(H) > v(H_1), v(H_2)$. This means that
$$\sum_{H' \in \mathcal{H}_k} c_k(H')\text{sub}(H', H) = c_k(H)\text{sub}(H, H) + \sum_{H' \in \mathcal{H}_k} c_k(H')(\text{sub}(H', H_1) + \text{sub}(H', H_2)).$$
By comparing it with our previous equation we get that
$$c_k(H)\text{sub}(H, H) = 0.$$
Since $\text{sub}(H, H) = 1$ we get that $c_k(H) = 0$ which contradicts our assumption. Hence for all disconnected graph H we have $c_k(H) = 0$.

□

Remark 5.2.3. It is clear from the proof that the multiplicativity was indeed crucial. It is also clear why we used $\text{sub}(H, G)$ instead of $\text{hom}(H, G)$. Simply we used that $\text{sub}(H, G) = 0$ if $v(H) \geq v(G)$ and $H \not\simeq G$. This is not true for the homomorphism numbers. Note that it does not mean that the analogous statement is not true for homomorphism numbers. Actually it is true, it follows from this theorem!

In terms of Theorem 5.2.2 we only need to ensure the existence of $\mathcal{H}_k$ such that for every graph G we have
$$\sum_{i=1}^{v(G)} \alpha_i^k = \sum_{H \in \mathcal{H}_k} c_k(H)\text{sub}(H, G),$$

where $\alpha_1, \dots, \alpha_{v(G)}$ are the zeros of $f(G,z)$. It turns out that if

$$f(G,z) = \sum_{i=0}^{v(G)} a_i(G) z^i,$$

then it is enough to require that for all $i \leq k$, the coefficient

$$a_{v(G)-i} = \sum_{H \in \mathcal{G}_i} d_k(H) \mathrm{sub}(H,G),$$

for some finite set $\mathcal{G}_i$. This is true, because we can use the Newton-Waring identities (also known as Newton–Girard formulae) to express

$$\sum_{i=1}^{v(G)} \alpha_i^k$$

in terms of the coefficients $a_{v(G)-i}$ for $i \leq k$. Seemingly there is an extra problem here: when we use the Newton-Waring identities not just terms of the form $\mathrm{sub}(H,G)$ appear, but also terms of the form $\mathrm{sub}(H_1,G)\mathrm{sub}(H_2,G)\dots\mathrm{sub}(H_t,G)$. It turns out that these products can be rewritten as linear combinations of $\mathrm{sub}(H',G)$ again, so we don't have to worry about it.

As the above discussion suggests we don't really need that $f(G,z)$ is of exponential-type, instead we need some bounds on the zeros and that the coefficients of the graph polynomial can be expressed as subgraph counting. This is indeed true and we give the corresponding general theorem. For a family $\mathcal{H}$ of graphs let

$$\mathbb{C}\mathcal{H} = \left\{ \sum_{H \in \mathcal{H}} c_H \mathrm{sub}(H,\cdot) \mid c_H \in \mathbb{C} \right\},$$

where all sums are finite, and let $\mathcal{G}$ be the family of all graphs.

Theorem 5.2.4. *Let f be a multiplicative monic graph polynomial with bounded zeros. We also assume that*

$$f(G,x) = \sum_{k=0}^{n} (-1)^k e_k(G) x^{n-k},$$

where $n = v(G)$ and all coefficients $e_k \in \mathbb{C}\mathcal{G}$. Let $(G_n)_n$ be a Benjamini–Schramm convergent graph sequence. Let $K \subset \mathbb{C}$ be a compact set containing all zeros of $f(G_n,x)$ for all n, such that $\mathbb{C} \setminus K$ is connected. Then the statements (a) and (b) of Theorem 5.2.1 hold.

One might wonder why we stated our original theorem in terms of exponential-type graph polynomials. It turns out that this class is surprisingly large, yet it automatically satisfies the condition $e_k \in \mathbb{C}\mathcal{G}$, and it is easy to check the boundedness condition on the zeros. Nevertheless, Theorem 5.2.4 is indeed a much more general theorem, for instance it covers the characteristic polynomial of the adjacency matrix, $\phi(G,x) = \det(xI - A_G)$. Guus Regts [28] also used this more general theorem to prove a convergence result about tensor networks.

5.3 Kesten-McKay measure

Definition 5.3.1. The Kesten-McKay measure is defined as follows. Let $\omega = 2\sqrt{d-1}$, and

$$f_d(x) = \frac{d\sqrt{\omega^2 - x^2}}{2\pi(d^2 - x^2)}\chi_{[-\omega,\omega]}.$$

Then the Kesten-McKay measure is the measure with density function $f_d(x)$.

We will see soon that it is the spectral measure of the infinite d–regular tree $\mathbb{T}_d$.

Definition 5.3.2. Let r_j denote the number of closed walks of length j from the root vertex o to the root vertex o in the infinite d–regular tree $\mathbb{T}_d$.

Theorem 5.3.3 (McKay [24])**.** *Let $\omega = 2\sqrt{d-1}$. Let $\rho_{\mathbb{T}_d}$ be the Kesten-McKay measure. Then*

$$r_j = \int_{-\omega}^{\omega} x^j \cdot \frac{d\sqrt{4(d-1) - x^2}}{2\pi(d^2 - x^2)}\, dx = \int_{-\omega}^{\omega} x^j \, d\rho_{\mathbb{T}_d}(x)$$

for all j.

Remark 5.3.4. Exact formulas for r_k are known, but we will not use them. If k is odd then $r_k = 0$, and for $k \geq 1$ we have

$$\begin{aligned} r_{2k} &= \sum_{j=1}^{k} \binom{2k-j}{k} \frac{j}{2k-j} d^j (d-1)^{k-j} = d \sum_{j=0}^{k-1} \binom{2k}{j} \frac{k-j}{k} (d-1)^j \\ &= \sum_{j=1} \binom{2k}{j} \frac{2k-2j+1}{2k-j+1} (d-1)^j. \end{aligned}$$

Theorem 5.3.5. *If $(G_i)_i$ is a sequence of d–regular graphs such that $g(G_i) \to \infty$, then the sequence of spectral measures $\rho_{G_i}^{(s)}$ converge weakly to the Kesten-McKay measure. The same statement is true for the matching measures.*

Proof. This immediately follows from Theorem 5.3.3 and the way we proved Theorems 5.1.4 and 5.1.5. □

There are a few special integrals that play important roles in certain proofs.

Proposition 5.3.6. *We have (a)*

$$\int_{-\omega}^{\omega} \ln(d-x)\, d\rho_{\mathbb{T}_d}(x) = \ln\left(\frac{(d-1)^{d-1}}{(d^2-2d)^{d/2-1}}\right),$$

(b)

$$\int_{-\omega}^{\omega} \ln|x|\, d\rho_{\mathbb{T}_d}(x) = \frac{1}{2}\ln\left(\frac{(d-1)^{d-1}}{d^{d-2}}\right).$$

(c)

$$\int \frac{1}{2}\ln(1+tz^2)\,d\rho_{\mathbb{T}_d}(z) = \frac{1}{2}\ln S_d(t),$$

where

$$S_d(t) = \frac{1}{\eta_t^2}\left(\frac{d-1}{d-\eta_t}\right)^{d-2} \quad \textit{with} \quad \eta_t = \frac{\sqrt{1+4(d-1)t}-1}{2(d-1)t}.$$

Part (a) is an important ingredient of the proof of McKay about the upper bound on the number of spanning trees of a d–regular graph. Part (b) appears in one of the proofs of Schrijver's theorem about the lower bound on the number of perfect matchings of a d–regular bipartite graph.

5.4 McKay's theorem

It is not very difficult exercise to prove that for a d–regular graph we have

$$\tau(G) \le \frac{e}{n} d^{n-1}.$$

It turns out that one can prove a much better upper bound. This is the content of the following theorem of McKay.

Theorem 5.4.1 (McKay [25]). *Let G be a d–regular graph on n vertices. Let $\tau(G)$ denote the number of spanning trees of G. Then*

$$\tau(G) \le \frac{c\ln n}{n}\left(\frac{(d-1)^{d-1}}{(d^2-2d)^{d/2-1}}\right)^n.$$

The constant $\frac{(d-1)^{d-1}}{(d^2-2d)^{d/2-1}}$ is best possible as we will see soon.

We will decompose the proof into several lemmas. The proof given here is a combination of McKay's [25] original proof and the proof of Chung and Yau [6].

Let $\rho_1 \ge \rho_2 \ge \dots \ge \rho_n$ be the eigenvalues of the adjacency matrix $A(G)$ of the graph G. Furthermore, let

$$\varphi_G(x) = \det(xI - A(G)) = \prod_{i=1}^{n}(x-\rho_i)$$

be the characteristic polynomial of the adjacency matrix.

Lemma 5.4.2. *If G is a d–regular graph and $\lambda_1 \ge \dots \ge \lambda_{n-1} \ge \lambda_n = 0$ are the Laplacian-eigenvalues and $\rho_1 \ge \rho_2 \ge \dots \ge \rho_n$ are the eigenvalues of the adjacency matrix $A(G)$ then $\lambda_i = d - \rho_{n+1-i}$. Furthermore,*

$$\tau(G) = \frac{1}{n}\prod_{i=2}^{n}(d-\rho_i).$$

Proof. The first claim simply follows from the observation that $L(G) = dI - A(G)$. The second claim follows from the first one and the fact that

$$\tau(G) = \frac{1}{n}\prod_{i=1}^{n-1}\lambda_i.$$

□

Recall that r_j denotes the number of closed walks of length j from the root vertex o to the root vertex o in the infinite d–regular tree $\mathbb{T}_d$.

Lemma 5.4.3. *Let G be a d-regular graph on n vertices, and let $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_n$ be the eigenvalues of its adjacency matrix. Let*

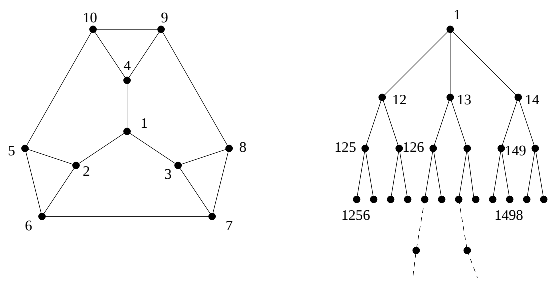
$$p_k = \sum_{i=1}^{n} \rho_i^k.$$

Then we have

$$p_k \geq n r_k.$$

Proof. Recall that p_{2k} counts the number of closed walks of length k. We will show that for any vertex v, the number of closed walks $W_k(v)$ of length k starting and ending at v is at least as large as the number of closed walks starting and ending at some root vertex of the infinite d–regular tree $\mathbb{T}_d$. When k is odd, the statemnet is trivial as $W_{2k+1}(v) \geq 0$ while $r_{2k+1} = 0$. So we only need to prove that $W_{2k}(v) \geq r_{2k}$.

Let us consider the following infinite d-regular tree, its vertices are labeled by the walks starting at the vertex v which never steps immediately back to a vertex from where it came. Such walks are called non-backtracking walks. For instance, 149831 is such a walk, but 1494 is not a backtracking walk since after 9 we immediately stepped back to 4. We connect two non-backtracking walks in the tree if one of them is a one-step extension of the other.



Note that every closed walk in the tree corresponds to a closed walk in the graph: for instance, $1, 14, 149, 14, 1$ corresponds to $1, 4, 9, 4, 1$. (In some sense, these are the "genuinely" closed walks.) On the other hand, there are closed walks in the graph G, like

149831, which are not closed anymore in the tree. Let r_{2k} denote the number of closed closed walks from a given a root vertex in the infinite d–regular tree. So far we know that

$$p_{2k} = \sum_{v \in V(G)} W_{2k}(v) \geq n r_{2k}.$$

□

Next we give a lower and an upper bound to r_{2j}.

Lemma 5.4.4. *We have*

$$\frac{\binom{2j}{j}}{j+1} d(d-1)^{j-1} \leq r_{2j} \leq (4(d-1))^j.$$

Remark 5.4.5. Chung and Yau proved the following stronger upper bound

$$r_{2j} \leq \frac{4^j d(d-1)^{j+1}}{j\sqrt{\pi j}(d-2)^2}.$$

Both the above lower bound and the stronger upper bound is of size $c(d)\frac{4^j d(d-1)^{j-1}}{j^{3/2}}$ with different constant $c(d)$.

Proof. The lower bound follows from the fact that each closed walk in d–regular tree can be considered as a generalized Dyck-path where every down-step can be chosen d or $d-1$ different ways according we are in the root or not and every up-step is determined as the unique step towards the root. For the upper bound observe that for $x \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$ we have $x^{2j} \leq (2\sqrt{d-1})^{2j}$, whence

$$r_{2j} = \int x^{2j} d\rho_{\mathbb{T}_d}(x) \leq \int (2\sqrt{d-1})^{2j} d\rho_{\mathbb{T}_d}(x) = (2\sqrt{d-1})^{2j} \int 1 d\rho_{\mathbb{T}_d}(x) = (2\sqrt{d-1})^{2j}.$$

□

Now we are ready to prove McKay's theorem.

Proof. For the sake of simplicity let

$$C_d = \frac{(d-1)^{d-1}}{(d^2-2d)^{d/2-1}}.$$

Furthermore, let $p_j = nr_j + w_j$. We have seen that $w_j \geq 0$.

$$\begin{aligned}
\tau(G) &= \frac{1}{n}\prod_{i=2}^{n}(d-\rho_i) = \frac{1}{n}\exp\left(\sum_{i=2}^{n}\ln(d-\rho_i)\right) \\
&= \frac{1}{n}\exp\left((n-1)\ln d + \sum_{i=2}^{n}\ln\left(1-\frac{\rho_i}{d}\right)\right) \\
&= \frac{1}{n}\exp\left((n-1)\ln d - \sum_{i=2}^{n}\sum_{j=1}^{\infty}\frac{1}{j}\left(\frac{\rho_i}{d}\right)^j\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\exp\left((n-1)\ln d - \sum_{j=1}^{\infty}\frac{1}{jd^j}\sum_{i=2}^{n}\rho_i^j\right)\\
&= \frac{1}{n}\exp\left((n-1)\ln d - \sum_{j=1}^{\infty}\frac{1}{jd^j}(p_j - d^j)\right)\\
&= \frac{1}{n}\exp\left((n-1)\ln d - \sum_{j=1}^{\infty}\frac{1}{jd^j}(nr_j + w_j - d^j)\right)\\
&= \frac{1}{n}\exp\left(-\ln d + n\left(\ln d - \sum_{j=1}^{\infty}\frac{r_j}{jd^j}\right) + \sum_{j=1}^{\infty}\frac{1}{jd^j}(d^j - w_j)\right)\\
&= \frac{1}{dn}\exp\left(n\left(\int_{-\omega}^{\omega}\ln(d-x)\,d\rho_{\mathbb{T}_d(x)}\right) + \left(\sum_{j=1}^{\infty}\frac{1}{jd^j}(d^j - w_j)\right)\right)\\
&= \frac{1}{dn}C_d^n\exp\left(\sum_{j=1}^{\infty}\frac{1}{jd^j}(d^j - w_j)\right).
\end{aligned}
$$

Let us fix an integer β which we will choose later. Then

$$
\sum_{j=1}^{\infty}\frac{1}{jd^j}(d^j - w_j) = \sum_{j=1}^{2\beta-1}\frac{1}{jd^j}(d^j - w_j) + \sum_{j=2\beta}^{\infty}\frac{1}{jd^j}(d^j - w_j).
$$

Then

$$
\sum_{j=1}^{2\beta-1}\frac{1}{jd^j}(d^j - w_j) \le \sum_{j=1}^{2\beta-1}\frac{1}{j} \le \ln(2\beta) + 1.
$$

While for the other sum we have

$$
\sum_{j=2\beta}^{\infty}\frac{1}{jd^j}(d^j - w_j) = \sum_{j=2\beta}^{\infty}\frac{1}{jd^j}(d^j - p_j + nr_j) = \sum_{j=2\beta}^{\infty}\frac{1}{jd^j}\left(-\sum_{i=2}^{n}\rho_i^j + nr_j\right) =
$$

$$
= -\sum_{i=2}^{n}\sum_{j=2\beta}^{\infty}\frac{\rho_i^j}{jd^j} + n\sum_{j=2\beta}^{\infty}\frac{r_j}{jd^j} \le n\sum_{j=2\beta}^{\infty}\frac{r_j}{jd^j}.
$$

since $\sum_{j=2\beta}^{\infty}\frac{u^j}{j} \ge 0$ for all $-1 < u < 1$. Hence

$$
\begin{aligned}
\sum_{j=2\beta}^{\infty}\frac{1}{jd^j}(d^j - w_j) &\le n\sum_{j=2\beta}^{\infty}\frac{r_j}{jd^j}\\
&= n\sum_{k=\beta}^{\infty}\frac{r_{2k}}{2kd^{2k}}\\
&\le n\sum_{k=\beta}^{\infty}\frac{1}{2k}\frac{1}{d^{2k}}(4(d-1))^k\\
&\le n\left(\frac{4(d-1)}{d^2}\right)^{\beta}\sum_{k=\beta}^{\infty}\left(\frac{4(d-1)}{d^2}\right)^{k-\beta}
\end{aligned}
$$

$$
\begin{aligned}
&= n\left(\frac{4(d-1)}{d^2}\right)^{\beta}\left(\frac{d}{d-2}\right)^{2} \\
&\leq 9n\left(\frac{4(d-1)}{d^2}\right)^{\beta}.
\end{aligned}
$$

Now let us choose β in such a way that

$$
9n\left(\frac{4(d-1)}{d^2}\right)^{\beta} \leq 1.
$$

Then $\beta < C\log_d(n)$. Putting all together we get that

$$
\tau(G) \leq \frac{1}{dn}C_d^n \exp\left(\sum_{j=1}^{\infty}\frac{1}{jd^j}(d^j - w_j)\right) \leq \frac{1}{dn}C_d^n \exp(\ln(2\beta)+2) \leq \frac{2e^2C\log_d(n)}{dn}C_d^n.
$$

We are done.

□

6. Correlation Inequalities

6.1 Gibbs-measures

For any counting problem there is a natural probability meaasure μ by putting uniform distribution on the studied object. We already meet this idea when we studied the probability that and edge e is in a random perfect matching or not. Naturally we can take weight function into account when we introduce the new measure. For instance, if we take a random independent set I we can introduce the measure

$$\mathbb{P}_{G,\lambda}(I) = \frac{\lambda^{|I|}}{I(G,\lambda)},$$

where $I(G,\lambda) = \sum_{k=0}^{n} i_k(G)\lambda^k$.

Once we have such a measure we can ask that certain events Q_1, Q_2 are correlated or not, i. e., what the relation of $\mathbb{P}_\mu(Q_1 \cap Q_2)$ and $\mathbb{P}_\mu(Q_1)\mathbb{P}_\mu(Q_2)$ is. These are natural questions from a probabilistic point of view and as we will see these questions are also relevant for various other problems. For instance, knowing a certain correlation inequality can predict what kind of extremal graph theoretic results might hold.

6.2 Positive correlation

Definition 6.2.1. For $\underline{x}, \underline{y} \in \{0,1\}^n$ let $\underline{x} \vee \underline{y}$ be the vector for which $(\underline{x} \vee \underline{y})_i = \max(x_i, y_i)$, and let $\underline{x} \wedge \underline{y}$ be the vector for which $(\underline{x} \wedge \underline{y})_i = \min(x_i, y_i)$.

Theorem 6.2.2 (Ahlswede and Daykin [2])**.** *Let $f_1, f_2, f_3, f_4 : \{0,1\}^n \to \mathbb{R}_+$ satisfying the inequality*

$$f_1(\underline{x})f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y})f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. Let

$$F_i = \sum_{\underline{x} \in \{0,1\}^n} f_i(\underline{x})$$

for $i = 1, 2, 3, 4$. Then

$$F_1 \cdot F_2 \leq F_3 \cdot F_4.$$

Proof. We prove the statement by induction on n. For $n = 1$ the condition of the theorem gives that

$$f_1(0)f_2(0) \leq f_3(0)f_4(0).$$

$$f_1(0)f_2(1) \leq f_3(1)f_4(0).$$

$$f_1(1)f_2(0) \leq f_3(1)f_4(0).$$

$$f_1(1)f_2(1) \leq f_3(1)f_4(1).$$

We need to prove that

$$(f_1(0) + f_1(1))(f_2(0) + f_2(1)) \leq (f_3(0) + f_3(1))(f_4(0) + f_4(1)).$$

If $f_3(1) = 0$ or $f_4(0) = 0$ then $f_3(1)f_4(0) \leq f_3(0)f_4(1)$ and the claim is trivially true:

$$(f_1(0)+f_1(1))(f_2(0)+f_2(1)) \leq f_3(0)f_4(0)+2f_3(1)f_4(0)+f_3(1)f_4(1) \leq (f_3(0)+f_3(1))(f_4(0)+f_4(1)).$$

So we can assume that $f_3(1) \neq 0$ and $f_4(0) \neq 0$. Then

$$(f_3(0) + f_3(1))(f_4(0) + f_4(1)) \geq \left(\frac{f_1(0)f_2(0)}{f_4(0)} + f_3(1)\right)\left(f_4(0) + \frac{f_1(1)f_2(1)}{f_3(1)}\right).$$

So it would be enough to prove that

$$\left(\frac{f_1(0)f_2(0)}{f_4(0)} + f_3(1)\right)\left(f_4(0) + \frac{f_1(1)f_2(1)}{f_3(1)}\right) \geq (f_1(0) + f_1(1))(f_2(0) + f_2(1)).$$

This is equivalent with

$$(f_1(0)f_2(0)+f_3(1)f_4(0))(f_3(1)f_4(0)+f_1(1)f_2(1)) \geq f_3(1)f_4(0)(f_1(0)+f_1(1))(f_2(0)+f_2(1)).$$

This is in turn equivalent with

$$(f_3(1)f_4(0) - f_1(0)f_2(1))(f_3(1)f_4(0) - f_1(1)f_2(0)) \geq 0$$

which is true by the assumptions of the theorem. This proves the case $n = 1$.

Now suppose that the claim is true till $n - 1$ and we wish to prove it for n. Set $f_i'(\underline{x}) : \{0,1\}^{n-1} \to \mathbb{R}_+$ for $i = 1, 2, 3, 4$ as follows:

$$f_i'(\underline{x}) = f_i(\underline{x}, 0) + f_i(\underline{x}, 1).$$

First we show that f_i' satisfies the inequality

$$f_1'(\underline{x})f_2'(\underline{y}) \leq f_3'(\underline{x} \vee \underline{y})f_4'(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^{n-1}$. This is of course true: for a fixed $\underline{x}, \underline{y} \in \{0,1\}^{n-1}$ let us apply the case $n = 1$ to the functions

$$g_1(u) = f_1(\underline{x}, u) \quad g_2(u) = f_2(\underline{y}, u) \quad g_3(u) = f_3(\underline{x} \vee \underline{y}, u) \quad g_4(u) = f_4(\underline{x} \wedge \underline{y}, u),$$

where $u \in \{0,1\}$. Then the functions g_i satisfy

$$g_1(u_1)g_2(u_2) \leq g_3(u_1 \vee u_2)g_4(u_1 \wedge u_2)$$

for all $u_1, u_2 \in \{0,1\}$ by the assumption on f. By the case $n = 1$ we know that

$$(g_1(0) + g_1(1))(g_2(0) + g_2(1)) \leq (g_3(0) + g_3(1))(g_4(0) + g_4(1)).$$

In other words,

$$f_1'(\underline{x})f_2'(\underline{y}) \leq f_3'(\underline{x} \vee \underline{y})f_4'(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^{n-1}$. Then by induction we get that for $F_i' = \sum_{\underline{x} \in \{0,1\}^{n-1}} f_i'(\underline{x})$ we have

$$F_1' \cdot F_2' \leq F_3' \cdot F_4'.$$

But of course $F_i' = F_i$ whence

$$F_1 \cdot F_2 \leq F_3 \cdot F_4.$$

□

Theorem 6.2.3. *Let $f_1, f_2, f_3, f_4 : \{0,1\}^n \to \mathbb{R}_+$ satisfying the inequality*

$$f_1(\underline{x})f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y})f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. Let $f_1', f_2', f_3', f_4'\{0,1\}^k \to \mathbb{R}_+$ be defined by

$$f_i'(\underline{x}) = \sum_{\underline{u} \in \{0,1\}^{n-k}} f_i(\underline{x}, \underline{u}).$$

Then for all $\underline{x}, \underline{y} \in \{0,1\}^k$ we have

$$f_1'(\underline{x})f_2'(\underline{y}) \leq f_3'(\underline{x} \vee \underline{y})f_4'(\underline{x} \wedge \underline{y})$$

Proof. This immediately follows from Theorem 6.2.2. For fixed $\underline{x}, \underline{y} \in \{0,1\}^k$ define $g_1, g_2, g_3, g_4 : \{0,1\}^{n-k} \to \mathbb{R}_+$

$$g_1(\underline{u}) = f_1(\underline{x}, \underline{u}), \quad g_2(\underline{u}) = f_2(\underline{y}, \underline{u}), \quad g_3(\underline{u}) = f_3(\underline{x} \vee \underline{y}, \underline{u}), \quad g_4(\underline{u}) = f_4(\underline{x} \wedge \underline{y}, \underline{u}).$$

Then for any $\underline{u}, \underline{v} \in \{0,1\}^{n-k}$ we have

$$g_1(\underline{u})g_2(\underline{v}) \leq g_3(\underline{u} \vee \underline{v})g_4(\underline{u} \wedge \underline{v})$$

by the assumption on the functions f_1, f_2, f_3, f_4. Then for

$$f_i'(\underline{x}) = G_i = \sum_{\underline{u} \in \{0,1\}^{n-k}} g_i(\underline{u}) = \sum_{\underline{u} \in \{0,1\}^{n-k}} f_i(\underline{x}, \underline{u})$$

we have

$$f_1'(\underline{x})f_2'(\underline{y}) = G_1G_2 \leq G_3G_4 = f_3'(\underline{x} \vee \underline{y})f_4'(\underline{x} \wedge \underline{y}).$$

□

Definition 6.2.4. For $\underline{x}, \underline{y} \in \{0,1\}^n$ we say that $\underline{x} \geq \underline{y}$ if for all $i \in [n]$ we have $x_i \geq y_i$.

A function $f : \{0,1\}^n \to \mathbb{R}^+$ is monotone increasing if $f(\underline{x}) \geq f(\underline{y})$ for all $\underline{x} \geq \underline{y}$ and it is monotone decreasing if $f(\underline{x}) \leq f(\underline{y})$ for all $\underline{x} \geq \underline{y}$.

In general, for a poset (or lattice) L a function $f : L \to \mathbb{R}^+$ is monotone increasing if $f(x) \geq f(x)$ for all $x \geq_L y$ and it is monotone decreasing if $f(x) \leq f(y)$ for all $x \geq_L y$.

Theorem 6.2.5. *A function $\mu : \{0,1\}^n \to \mathbb{R}^+$ is log-supermodular if*

$$\mu(\underline{x})\mu(\underline{y}) \leq \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. Then for a log-supermodular $\mu : \{0,1\}^n \to \mathbb{R}^+$ and monotone increasing (decreasing) functions $f, g : \{0,1\}^n \to \mathbb{R}^+$ we have

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{x \in \{0,1\}^n} \mu(\underline{x})g(\underline{x})\right) \leq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})f(\underline{x})g(\underline{x})\right)\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})\right).$$

Furthermore, if $f : \{0,1\}^n \to \mathbb{R}^+$ is monotone increasing and $g : \{0,1\}^n \to \mathbb{R}^+$ is monotone decreasing then

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})g(\underline{x})\right) \geq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})f(\underline{x})g(\underline{x})\right)\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})\right).$$

Proof. First suppose that both f and g are monotone increasing. Let us apply Theorem 6.2.2 for the following theorems:

$$f_1(\underline{x}) = \mu(\underline{x})f(\underline{x}), \quad f_2(\underline{x}) = \mu(\underline{x})g(\underline{x}), \quad f_3(\underline{x}) = \mu(\underline{x})f(\underline{x})g(\underline{x}), \quad f_4(\underline{x}) = \mu(\underline{x}).$$

We need to check that

$$f_1(\underline{x})f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y})f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. This is indeed true:

$$\begin{aligned} f_1(\underline{x})f_2(\underline{y}) &= \mu(\underline{x})f(\underline{x})\mu(\underline{y})g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})f(\underline{x})g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})f(\underline{x} \vee \underline{y})g(\underline{x} \vee \underline{y}) \\ &= f_3(\underline{x} \vee \underline{y})f_4(\underline{x} \wedge \underline{y}). \end{aligned}$$

In the first inequality we used the log-supermodularity of μ, and in the second inequality we used that both f and g are monotone increasing. Then by Theorem 6.2.2 we have $F_1 \cdot F_2 \leq F_3 \leq F_4$, i. e.,

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{x \in \{0,1\}^n} \mu(\underline{x})g(\underline{x})\right) \leq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})f(\underline{x})g(\underline{x})\right)\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x})\right).$$

If f and g are both monotone decreasing then set

$$f_1(\underline{x}) = \mu(\underline{x})f(\underline{x}), \quad f_2(\underline{x}) = \mu(\underline{x})g(\underline{x}), \quad f_3(\underline{x}) = \mu(\underline{x}), \quad f_4(\underline{x}) = \mu(\underline{x})f(\underline{x})g(\underline{x}).$$

Again we need to check that

$$f_1(\underline{x})f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y})f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. This is indeed true:

$$\begin{aligned} f_1(\underline{x})f_2(\underline{y}) &= \mu(\underline{x})f(\underline{x})\mu(\underline{y})g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})f(\underline{x})g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})f(\underline{x} \wedge \underline{y})g(\underline{x} \wedge \underline{y}) \\ &= f_3(\underline{x} \vee \underline{y})f_4(\underline{x} \wedge \underline{y}). \end{aligned}$$

From this we can conclude again that

$$\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{x\in\{0,1\}^n} \mu(\underline{x})g(\underline{x})\right) \leq \left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})g(\underline{x})\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})\right).$$

If f is monotone increasing, and g is monotone decreasing then let $M = \max_{\underline{x}\in\{0,1\}^n} g(\underline{x})$, and consider the function $g'(\underline{x}) = M - g(\underline{x})$. Then $g'(\underline{x}) \geq 0$ and monotone increasing. Whence

$$\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{x\in\{0,1\}^n} \mu(\underline{x})g'(\underline{x})\right) \leq \left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})g'(\underline{x})\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})\right).$$

By writing the definition of $g(\underline{x}) = M - g'(\underline{x})$ into it we get that

$$\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{x\in\{0,1\}^n} \mu(\underline{x})(M - g(\underline{x}))\right) \leq \left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})(M - g(\underline{x}))\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})\right).$$

After subtracting $M(\sum \mu(\underline{x}))(\sum \mu(\underline{x})f(\underline{x}))$ and multiplying with -1 we get that

$$\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})g(\underline{x})\right) \geq \left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})g(\underline{x})\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})\right).$$

□

Theorem 6.2.6. *Let L be a distributive lattice. A function $\mu : L \to \mathbb{R}^+$ is log-supermodular if*

$$\mu(x)\mu(y) \leq \mu(x \vee y)\mu(x \wedge y)$$

for all $x, y \in L$. For a log-supermodular $\mu : L \to \mathbb{R}^+$ and monotone increasing (decreasing) functions $f, g : L \to \mathbb{R}^+$ we have

$$\left(\sum_{x\in L}\mu(x)f(x)\right)\left(\sum_{x\in L}\mu(x)g(x)\right) \le \left(\sum_{x\in L}\mu(x)f(x)g(x)\right)\left(\sum_{x\in L}\mu(x)\right).$$

Furthermore, if $f : L \to \mathbb{R}^+$ is monotone increasing and $g : L \to \mathbb{R}^+$ is monotone decreasing then

$$\left(\sum_{x\in L}\mu(x)f(x)\right)\left(\sum_{x\in L}\mu(x)g(x)\right) \ge \left(\sum_{x\in L}\mu(x)f(x)g(x)\right)\left(\sum_{x\in L}\mu(x)\right).$$

Proof. This theorem follows from Theorem 6.2.5 since every distributive lattice L is a sublattice of some $\{0,1\}^n$. So all we need to do is to define μ on $\{0,1\}^n \setminus L$ to be 0, and to extend f and g in a monotone increasing way. (This last step is only needed formally since $\mu(\underline{x})f(\underline{x}), \mu(\underline{x})g(\underline{x}), \mu(\underline{x})f(\underline{x})g(\underline{x})$ are all 0 anyway for $\underline{x} \in \{0,1\}^n \setminus L$.) The extended μ will remain log-supermodular since $\mu(x)\mu(y) \neq 0$ then $x, y \in L$ and then $x \vee y, x \wedge y \in L$ so $\mu(x)\mu(y) \le \mu(x \vee y)\mu(x \wedge y)$, and if $\mu(x)\mu(y) = 0$ then the inequality holds true trivially. □

In the next few results we give examples of various log-supermodular measures.

Theorem 6.2.7. *Assume that the function $\mu : \{0,1\}^n \to \mathbb{R}_+$ is log-supermodular. Then the function $\mu' : \{0,1\}^k \to \mathbb{R}_+$ defined by*

$$\mu'(\underline{x}) = \sum_{\underline{u}\in\{0,1\}^{n-k}} \mu(\underline{x}, \underline{u})$$

is also log-supermodular.

Proof. This theorem is an immediate application of Theorem 6.2.3 applied to
$f_1 = f_2 = f_3 = f_4 = \mu$.

□

Theorem 6.2.8. *For probabilities $p_1, \dots, p_n$ let*

$$\mathbb{P}_p(A) = \prod_{i\in A} p_i \prod_{j\notin A}(1 - p_j).$$

Let $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$ be monotone increasing, and $\mathcal{C}, \mathcal{D} \subseteq 2^{[n]}$ be monotone decreasing set families. For a set family $\mathcal{S}$ set

$$\mathbb{P}_p(\mathcal{S}) = \sum_{S\in\mathcal{S}} \mathbb{P}_p(S).$$

Then we have

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \ge \mathbb{P}_p(\mathcal{A}) \cdot \mathbb{P}_p(\mathcal{B}),$$
$$\mathbb{P}_p(\mathcal{C} \cap \mathcal{D}) \ge \mathbb{P}_p(\mathcal{C}) \cdot \mathbb{P}_p(\mathcal{D}),$$
$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{C}) \le \mathbb{P}_p(\mathcal{A}) \cdot \mathbb{P}_p(\mathcal{C}).$$

Proof. We can associate the characteristic vector $\underline{1}_A \in \{0,1\}^n$ with a set A. Let

$$\mu(\underline{x}) = \prod_{i=1}^{n} p_i^{x_i}(1-p_i)^{1-x_i}.$$

Then $\mathbb{P}_p(A) = \mu(\underline{1}_A)$. Then

$$\mu(\underline{x})\mu(\underline{y}) = \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})$$

or equivalently $\mathbb{P}_p(A)\mathbb{P}_p(B) = \mathbb{P}_p(A \cup B)\mathbb{P}_p(A \cap B)$. Furthermore, let f be the characteristic functions of the family of sets $\mathcal{A}$, i. e., $f(1_A) = 1$ if $A \in \mathcal{A}$ and 0 otherwise. Similarly, let g be the characteristic functions of the family of sets $\mathcal{B}$. Then f and g are monotone increasing functions. The inequality

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \geq \mathbb{P}_p(\mathcal{A}) \cdot \mathbb{P}_p(\mathcal{B})$$

is simply Theorem 6.2.5 applied to μ, f and g. The other parts of the theorem follows similarly.

□

Proposition 6.2.9. *Let $A_1, \dots, A_m$ be fixed subsets of $N = \{1, 2, \dots, n\}$. Let S be a random subset of N such that $\mathbb{P}(i \in S) = p_i$ independently of each other. Then we have*

$$\mathbb{P}_p[A_k \not\subseteq S \textit{ for each } 1 \leq k \leq m] \geq \prod_{i=1}^{m} \mathbb{P}_p[A_k \not\subseteq S].$$

Proof. Let B_k be the event that $A_k \not\subseteq S$. Then $\cap_{k \in I} B_k$ is a monotone decreasing event for all $I \subseteq \{1, 2, \dots, m\}$. Hence by induction on m we get that

$$\mathbb{P}_p[A_k \not\subseteq S \text{ for each } 1 \leq k \leq m] = \mathbb{P}_p\left[\bigcap_{k=1}^{m} B_k\right] \geq \prod_{i=1}^{m} \mathbb{P}_p[B_k].$$

□

Theorem 6.2.10. *Given a graph G on the vertex set $[n]$, and a parameter $\beta > 0$ with a vector $\underline{B} = (B_1, \dots, B_n)$. For a $\underline{x} = (x_1, \dots, x_n) \in \{-1, 1\}^n$ set*

$$\mathbb{P}_{\beta,\underline{B}}(\underline{x}) = \frac{1}{Z} \exp\left(\beta \sum_{(i,j) \in E(G)} x_i x_j + \sum_{i=1}^{n} B_i x_i\right),$$

where Z is some normalizing constant. For vectors $\underline{x}$ and $\underline{y}$ let $(\underline{x} \wedge \underline{y})_i = \min(x_i, y_i)$ and $(\underline{x} \vee \underline{y})_i = \max(x_i, y_i)$. Then we have

$$\mathbb{P}_{\beta,\underline{B}}(\underline{x}) \cdot \mathbb{P}_{\beta,\underline{B}}(\underline{y}) \leq \mathbb{P}_{\beta,\underline{B}}(\underline{x} \wedge \underline{y}) \cdot \mathbb{P}_{\beta,\underline{B}}(\underline{x} \vee \underline{y}).$$

Proof. Clearly, the statement is equivalent with

$$\left(\beta \sum_{(i,j)\in E(G)} x_i x_j + \sum_{i=1}^{n} B_i x_i\right) + \left(\beta \sum_{(i,j)\in E(G)} y_i y_j + \sum_{i=1}^{n} B_i y_i\right) \leq$$

$$\leq \left(\beta \sum_{(i,j)\in E(G)} \min(x_i, y_i)\min(x_j, y_j) + \sum_{i=1}^{n} B_i \min(x_i, y_i)\right)$$

$$+ \left(\beta \sum_{(i,j)\in E(G)} \max(x_i, y_i)\max(x_j, y_j) + \sum_{i=1}^{n} B_i \max(x_i, y_i)\right).$$

It is enough to prove the statement term by term. Note that $x_i + y_i = \min(x_i, y_i) + \max(x_i, y_i)$ so we only need to prove that

$$x_i x_j + y_i y_j \leq \min(x_i, y_i)\min(x_j, y_j) + \max(x_i, y_i)\max(x_j, y_j).$$

If $x_i \leq y_i$ and $x_j \leq y_j$ then this statement holds true with equality. Of course, the same is true if $x_i \geq y_i$ and $x_j \geq y_j$. If $x_i \leq y_i$ and $x_j \geq y_j$ then we need to prove that

$$x_i x_j + y_i y_j \leq x_i y_j + x_j y_i$$

which is equivalent with

$$(x_i - y_i)(y_j - x_j) \geq 0$$

which is true by assumption. Again the same is true if $x_i \geq y_i$ and $x_j \leq y_j$. □

Theorem 6.2.11. *Let $G = (A, B, E)$ be a bipartite graph, and let $\lambda \geq 0$. Let $\mathcal{I}(G)$ be the set of independent sets of G. Let $\boldsymbol{I}$ be a random independent set of G such that for an independent set I of G we have*

$$\mathbb{P}_\lambda(\boldsymbol{I} = I) = \frac{\lambda^{|I|}}{I(G, \lambda)}$$

where

$$I(G, \lambda) = \sum_I \lambda^{|I|}.$$

Then for all $u, v \in A$ we have

$$\mathbb{P}_\lambda[u, v \in \boldsymbol{I}] \geq \mathbb{P}_\lambda[u \in \boldsymbol{I}]\mathbb{P}_\lambda[v \in \boldsymbol{I}],$$

and for $u \in A$ and $v \in B$ we have

$$\mathbb{P}_\lambda[u, v \in \boldsymbol{I}] \leq \mathbb{P}_\lambda[u \in \boldsymbol{I}]\mathbb{P}_\lambda[v \in \boldsymbol{I}].$$

Proof. Consider the following function $\mu : \{0,1\}^{A \cup B}$:

$$\mu(\underline{x}, \underline{y}) = \exp\left(\ln(\lambda)\left(\sum_{u \in A} x_u + \sum_{v \in B}(1 - y_v)\right)\right) \prod_{(u,v) \in E(G)} (1 - x_u(1 - y_v)).$$

First we show the connection between μ and $\mathbb{P}_\lambda$. For $(\underline{x}, \underline{y}) \in \{0,1\}^{A \cup B}$ set

$$S = \{u \in A \mid x_u = 1\} \cup \{v \in B \mid y_v = 0\}.$$

Note that if S is not an independent set then there exists a $(u,v) \in E(G)$ such that $x_u = 1, y_v = 0$. Then $1 - x_u(1 - y_v) = 0$ and $\mu(\underline{x}, \underline{y}) = 0$. If S is an independent set for all $(u,v) \in E(G)$ we have $1 - x_u(1 - y_v) = 1$, and

$$\exp\left(\ln(\lambda)\left(\sum_{u \in A} x_u + \sum_{v \in B}(1 - y_v)\right)\right) = \lambda^{|S|}.$$

So up to the normalization constant $I(G, \lambda)$ and $\mathbb{P}_\lambda$ are the same. Next we show that $\mu(\underline{x}, \underline{y})$ is log-supermodular. It is clear that if $(\underline{x}, \underline{y})$ and $(\underline{x}', \underline{y}')$ are two vectors then

$$\left(\sum_{u \in A} x_u + \sum_{v \in B}(1 - y_v)\right) + \left(\sum_{u \in A} x'_u + \sum_{v \in B}(1 - y'_v)\right) =$$

$$= \left(\sum_{u \in A} \min(x_u, x'_u) + \sum_{v \in B}(1 - \min(y_v, y'_v))\right) + \left(\sum_{u \in A} \max(x_u, x'_u) + \sum_{v \in B}(1 - \max(y_v, y'_v))\right).$$

So we only need to prove that

$$(1 - x_u(1 - y_v))(1 - x'_u(1 - y'_v)) \leq (1 - \min(x_u, x'_u)(1 - \min(y_v, y'_v))(1 - \max(x_u, x'_u)(1 - \max(y_v, y'_v)).$$

One can do it by checking 16 cases, but it is possible to speed up the checking by some observations. The right hand side is non-negative so we only need to exclude the cases where the left hand side is 1 (and the right hand side is 0). If $x_u = x'_u = 1$ then $y_v = y'_v = 0$ and then the right hand side is 1. If $x_u = x'_u = 0$ then the right hand side is again 1. Similarly, if $y_v = y'_v = 0$ then $x_u = x'_u = 0$, or $y_v = y'_v = 1$, then right hand side is again 1. So we only need to check when one of x_u and x'_u is 1, the other 0, and one of y_v and y'_v is 1, the other 0. By symmetry we can assume that $x_u = 0, x'_u = 1$: if $y_v = 0, y'_v = 1$ then both sides is 1, and if $y_v = 1, y'_v = 0$ then the left hand side is 0, but the right hand side is still 1. Hence the inequality is indeed true.

Finally, fix a u and v as in the statement in the theorem. Now we can apply Theorem 6.2.5 in the first case to the functions

$$\mu(\underline{x}, \underline{y}), \quad f(\underline{x}, \underline{y}) = x_u, \quad g(\underline{x}, \underline{y}) = x_v.$$

Clearly,

$$\sum_{(\underline{x}, \underline{y})} \mu(\underline{x}, \underline{y}) x_u x_v = I(G, \lambda) \mathbb{P}_\lambda[u, v \in \mathbf{I}].$$

In the second case we can apply Theorem 6.2.5 to the functions.

$$\mu(\underline{x}, \underline{y}), \quad f(\underline{x}, \underline{y}) = x_u, \quad g(\underline{x}, \underline{y}) = 1 - y_v.$$

In the first case, both f and g are monotone increasing, in the second case f is monotone increasing and g is monotone decreasing. After dividing by $I(G, \lambda)^2$ we get that for all $u, v \in A$ we have

$$\mathbb{P}_\lambda[u, v \in \mathbf{I}] \geq \mathbb{P}_\lambda[u \in \mathbf{I}]\mathbb{P}_\lambda[v \in \mathbf{I}],$$

and for a $u \in A$ and $v \in B$ we have

$$\mathbb{P}_\lambda[u, v \in \mathbf{I}] \leq \mathbb{P}_\lambda[u \in \mathbf{I}]\mathbb{P}_\lambda[v \in \mathbf{I}].$$

□

7. Graph Covers

7.1 Two ideas, one method

The scope of this chapter is the study of graph covers. Let us start with the definition.

Definition 7.1.1. Let G be a graph. Then H is a 2-lift of G, if $V(H) = V(G) \times \{0,1\}$, and for every $(u,v) \in E(G)$, exactly one of the following two pairs are edges of H: $((u,0),(v,0))$ and $((u,1),(v,1)) \in E(H)$, or $((u,0),(v,1))$ and $((u,1),(v,0)) \in E(H)$. If $(u,v) \notin E(G)$, then none of $((u,0),(v,0))$, $((u,1),(v,1))$, $((u,0),(v,1))$ and $((u,1),(v,0))$ are edges in H.

More generally one can define a k-lift (or k-cover) of a graph as follows. The vertex set of a k-lift H is $V(H) = V(G) \times \{0,1,\dots,k-1\}$, and if $(u,v) \in E(G)$ then we choose a perfect matching between the vertices (u,i) and (v,j) for $0 \le i,j \le k-1$. If $(u,v) \notin E(G)$, then there is no edges between (u,i) and (v,j) for $0 \le i,j \le k-1$.

There are two notable special cases of a 2-lift. When all edges are of the form $((u,0),(v,0))$ and $((u,1),(v,1))$ then we simply get two disjoint copies of the graph G, so it is just $G \cup G$. When all edges are of the form $((u,0),(v,1))$ for some $(u,v) \in E(G)$ then we get $G \times K_2$. It will turn out that these special 2-lifts often play the role of an extremal graph among 2-lifts of a graph.

There are two essentially different method connected to graph covers. The first idea is the so-called bipartite double cover trick. Here we compare the graphs $G \cup G$ and $G \times K_2$ in order to transfer results about bipartite graphs to all graphs. See the next section for details. The other idea related to graph covers is a bit more involved. The idea very briefly is the following: assume that some graph parameter $p(G)$ and for any 2-lift H of G we have

$$p(G) \ge p(H).$$

A key observation made by Linial [22] (see Lemma 7.5.1 below) is that for any graph G one can construct a sequence of graphs $G = G_0, G_1, G_2, \dots$ such that G_{i+1} is a 2-lift of G_i, and the girth $g(G_i)$ tend to infinity. This way we get two things

$$p(G) \ge p(G_1) \ge p(G_2) \ge \dots$$

and the sequence $g(G_i)$ tends to infinity. When the graph G is d–regular then all (G_i) are d–regular too, and if G is bipartite then so all G_i. In other words, the constructed sequence (G_i) is Benjamini–Schramm convergent to the infinite d–regular tree in case of a d–regular graph G. In general, (G_i) converges to a distribution on the rooted universal cover trees of G. If it were true that

$$\lim_{i\to\infty} p(G_i)$$

exists and we can compute it then we would get a general lower bound for $p(G)$. This is exactly the problem we attacked with some success when we studied empirical measures. Actually, there are other methods giving such results, especially in locally tree-like graphs, exactly the case we need. For instance, it is known that the limit exists and it is computed for $p(G) = Z(G, A_{\mathrm{Is}(\beta)})^{1/v(G)}$ if $\beta > 0$, or $\beta < 0$ and G_i's are bipartite, or for $I(G)^{1/v(G)}$ again when G_i's are bipartite. For these models these are exactly the cases when we were able to prove an inequality of type $p(G) \geq p(H)$.

Another variant of the same idea is the following: suppose that we can prove that $p(G) \geq p(H_k)$ for all k-covers of G. Then

$$p(G) \geq \mathbb{E}p(H_k),$$

where the average $\mathbb{E}$ goes for all k-cover of G. The computation of $\mathbb{E}p(H_k)$ is still not very easy, but in general is quite easy to compute

$$\lim_{k\to\infty} \mathbb{E}p(H_k)$$

or some variant of it like $\lim_{k\to\infty} \left(\mathbb{E}p(H_k)^k\right)^{1/k}$.

7.2 Bipartite double cover

Recall that Theorem 2.2.1 asserts that if G is a d–regular bipartite graph, and $I(G)$ denotes the number of independent sets of G, then

$$I(G)^{1/v(G)} \leq I(K_{d,d})^{1/v(K_{d,d})}.$$

The goal of this section to introduce a simple yet powerful idea of Zhao to show that the bipartiteness condition is not necessary in this statement. The general theme is that one can transfer many results on bipartite graphs to general graphs by using the biprirtite double cover graph $G \times K_2$. In the particular case of independent sets Zhao proved the following result.

Theorem 7.2.1 (Zhao [36])**.** *For any graph G we have*

$$I(G)^2 \leq I(G \times K_2).$$

Consequently, for any d–regular graph G we have

$$I(G)^{1/v(G)} \leq I(G \times K_2)^{1/v(G \times K_2)} \leq I(K_{d,d})^{1/v(K_{d,d})},$$

where the second inequality follows from Theorem 2.2.1.

We postpone its proof to Section 7.3, where we will prove a slightly more general result. In the same vain one can prove a similar theorem about perfect matchings. This result was discovered and rediscovered many times.

Theorem 7.2.2. *For any graph G we have*

$$\mathrm{pm}(G)^2 \leq \mathrm{pm}(G \times K_2).$$

Again we postpone its proof to Section 7.3, where we will prove a slightly more general result. As an application of this result one can combine it with Brégman's theorem to get the following result. This theorem was first proved by Kahn and Lovász in an unpublished manuscript. Then Alon and Friedland [3] noticed that with the above result one can deduce it from Brégman's theorem.

Theorem 7.2.3 (Kahn and Lovász)**.** *Let G be a graph on n vertices with degree sequence $d_1, \dots, d_n$. Then*

$$\mathrm{pm}(G) \leq \prod_{i=1}^{n} (d_i!)^{1/(2d_i)}.$$

Proof. We have

$$\mathrm{pm}(G)^2 \leq \mathrm{pm}(G \times K_2) \leq \prod_{i=1}^{n} (d_i!)^{1/d_i}.$$

The second inequality is the direct application of Brégman's theorem. □

An immediate corollary of this result that in Theorem 10.3.2 we can drop the condition on the bipartiteness of the graph G:

Theorem 7.2.4. *Let* $\mathrm{pm}(G)$ *denote the number of perfect matchings. Then for a d–regular graph G we have*

$$\mathrm{pm}(G)^{1/v(G)} \leq \mathrm{pm}(K_{d,d})^{1/v(K_{d,d})}$$

7.3 Independent sets, matchings and 2-covers

In this section we consider the case $A = A_{\mathrm{ind}}$, i. e., we are counting independent sets. This section is completely elementary.

Theorem 7.3.1. *Let G be a graph, and let H be a 2-lift of G. Then*

$$i_k(H) \leq i_k(G \times K_2),$$

where $i_k(.)$ denotes the number of independent sets of size k.

Remark 7.3.2. This statement gives a generalization of Zhao's result, namely

$$i_k(G \cup G) \leq i_k(G \times K_2).$$

On the other hand, if G is bipartite then $G \times K_2 = G \cup G$ in which case it gives that

$$i_k(G \cup G) \geq i_k(H)$$

for any 2-lift H.

Proof. Let I be any independent set of a 2-lift of G. Let us consider the projection of I to G, then it will consist of vertices and "double-vertices" (i.e, when two vertices map to the same vertex). Let $\mathcal{R}$ be the set of these configurations. Then

$$i_k(H) = \sum_{R \in \mathcal{R}} |\phi_H^{-1}(R)|$$

and

$$i_k(G \times K_2) = \sum_{R \in \mathcal{R}} |\phi_{G \times K_2}^{-1}(R)|,$$

where ϕ_H and $\phi_{G \times K_2}$ are the projections from H and $G \times K_2$ to G. Note that

$$|\phi_{G \times K_2}^{-1}(R)| = 2^{k(R)},$$

where $k(R)$ is the number of connected components of R different from a double-vertex. Indeed, in each component we can lift the vertices such a way that the image belongs to exactly one bipartite class. The projection of a double-vertex must be a connected component on its own. On the other hand,

$$|\phi_H^{-1}(R)| \leq 2^{k(R)},$$

since in each component if we know the inverse image of one vertex then we immediately know the inverse images of all other vertices. Clearly, there is no equality in general. Hence

$$|\phi_H^{-1}(R)| \leq |\phi_{G \times K_2}^{-1}(R)|$$

and consequently,

$$i_k(H) \leq i_k(G \times K_2).$$

□

Theorem 7.3.3. *Let G be a graph, and let H be a 2-lift of G. Then*

$$m_k(H) \leq m_k(G \times K_2),$$

where $m_k(.)$ denotes the number of matchings of size k. In particular, if G is bipartite then

$$m_k(H) \leq m_k(G \cup G).$$

Proof. Let M be any matching of a 2-lift of G. Let us consider the projection of M to G, then it will consist of paths, cycles and "double-edges" (i.e, when two edges project to the same edge). Let $\mathcal{R}$ be the set of these configurations. Then

$$m_k(H) = \sum_{R \in \mathcal{R}} |\phi_H^{-1}(R)|$$

and

$$m_k(G \times K_2) = \sum_{R \in \mathcal{R}} |\phi_{G \times K_2}^{-1}(R)|,$$

where ϕ_H and $\phi_{G \times K_2}$ are the projections from H and $G \times K_2$ to G. Note that

$$|\phi_{G \times K_2}^{-1}(R)| = 2^{k(R)},$$

where $k(R)$ is the number of paths and cycles of R. Indeed, in each path or cycle we can lift the edges in two different ways. The projection of a double-edge is naturally unique. On the other hand,

$$|\phi_H^{-1}(R)| \leq 2^{k(R)},$$

since in each path or cycle if we know the inverse image of one edge then we immediately know the inverse images of all other edges. Clearly, there is no equality for cycles in general. Hence

$$|\phi_H^{-1}(R)| \leq |\phi_{G \times K_2}^{-1}(R)|$$

and consequently,

$$m_k(H) \leq m_k(G \times K_2).$$

The second part follows from the first part since for a bipartite graph G we have $G \times K_2 \simeq G \cup G$. □

7.4 Ruozzi's theorem

Theorem 7.4.1 (Ruozzi [29])**.** *Suppose that for the factor graph* $\mathcal{G} = (F, V, E, \mathcal{X} = \{0,1\}, (g_a)_{a \in F})$ *all functions* g_a *are log-supermodular. Then for any* k*-cover* $\mathcal{H}$ *of* $\mathcal{G}$ *we have* $Z(\mathcal{H}) \leq Z(\mathcal{G}^k)$*.*

Since the proof of Theorem 7.4.1 is quite involved, it might be interesting to see that one can get a much easier proof for 2-covers. This variant of Ruozzi's theorem does not require all g_a to be log-supermodular, only log-supermodularity of $g_{\mathcal{H}} = \prod_{a \in F(\mathcal{H})} g_a$, and its proof is much simpler. Unfortunately, this proof seems to work only for 2-covers.

Theorem 7.4.2. *Let* $\mathcal{G}$ *be a factor graph, and let* $\mathcal{H}$ *be its* 2*-cover. Suppose that the function* $g_{\mathcal{H}} = \prod_{a \in F(\mathcal{H})} g_a$ *is a log-supermodular function. Then we have* $Z(\mathcal{H}) \leq Z(\mathcal{G}^2)$*.*

Proof. Let u be a variable node of $\mathcal{G}$, and let u' be its pair in the lifts $\mathcal{G}^2$ and $\mathcal{H}$. For $i, j \in \{0,1\}^2$ consider the following quantities:

$$Z_{ij}(\mathcal{G}^2) = \sum_{\substack{\underline{x} \in \{0,1\}^n \\ x_u = i, x_{u'} = j}} f(\mathcal{G}^2, \underline{x}) \quad Z_{ij}(\mathcal{H}) = \sum_{\substack{\underline{x} \in \{0,1\}^n \\ x_u = i, x_{u'} = j}} g(\mathcal{H}, x).$$

Note that $Z_{00}(\mathcal{G}^2), Z_{11}(\mathcal{G}^2), Z_{00}(\mathcal{H}), Z_{11}(\mathcal{H})$ can be considered as the partition function of 2-lifts of $G - u$ by simply replacing those functions g_a which contains x_u by g'_a and define $g_a(\underline{x}_{\partial a \setminus u}, j)$, where $j = 0$ or 1 according to which Z_{jj} we consider. The obtained function $g_{\mathcal{H} - \{u, u'\}}$ is still log-supermodular. By induction we get that

$$Z_{00}(\mathcal{G}^2) \geq Z_{00}(\mathcal{H}) \quad \text{and} \quad Z_{11}(\mathcal{G}^2) \geq Z_{11}(\mathcal{H}).$$

Note that

$$Z_{01}(\mathcal{G}^2) = Z_{10}(\mathcal{G}^2) = \sqrt{Z_{00}(\mathcal{G}^2) Z_{11}(\mathcal{G}^2)}.$$

On the other hand, we have

$$Z_{01}(\mathcal{H}) = Z_{10}(\mathcal{H}) \leq \sqrt{Z_{00}(\mathcal{H}) Z_{11}(\mathcal{H})}.$$

This is true since if $g = g_{\mathcal{H}}$ is a log-supermodular function on $\{0,1\}^n$ then for any $k \leq n$ the function $h : \{0,1\}^k \to \mathbb{R}$ defined by

$$h(\underline{y}) = \sum_{\underline{x} \in \{0,1\}^{n-k}} f(\underline{y}, \underline{x})$$

is also log-supermodular. Hence

$$Z_{01}(\mathcal{H}) Z_{10}(\mathcal{H}) \leq Z_{00}(\mathcal{H}) Z_{11}(\mathcal{H}).$$

Now

$$Z(\mathcal{H}) = Z_{00}(\mathcal{H}) + Z_{11}(\mathcal{H}) + Z_{01}(\mathcal{H}) + Z_{10}(\mathcal{H})$$

$$
\begin{aligned}
&\le Z_{00}(\mathcal{H}) + Z_{11}(\mathcal{H}) + 2\sqrt{Z_{00}(\mathcal{H})Z_{11}(\mathcal{H})} \\
&\le Z_{00}(\mathcal{G}^2) + Z_{11}(\mathcal{G}^2) + 2\sqrt{Z_{00}(\mathcal{G}^2)Z_{11}(\mathcal{G}^2)} \\
&= Z_{00}(\mathcal{G}^2) + Z_{11}(\mathcal{G}^2) + Z_{01}(\mathcal{G}^2) + Z_{10}(\mathcal{G}^2) \\
&= Z(\mathcal{G}^2).
\end{aligned}
$$

Hence $Z(\mathcal{G}^2) \ge Z(\mathcal{H})$.

□

7.5 Large girth phenomena

The following lemma of Linial will be particularly useful for us when we connect the graph cover ideas with local convergence.

Lemma 7.5.1 (Linial [22]). *For a graph H let $g(H)$ denote the girth of the graph H, i. e., the length of the shortest cycle. Then for any graph G, there exists a graph sequence $(G_i)_{i=0}^{\infty}$ such that $G_0 = G$, G_i is a 2-lift of G_{i-1} for $i \ge 1$, and $g(G_i) \to \infty$.*

Proof. We will show that there exists a sequence (G_i) of 2-lifts such that for any k, there exists an $N(k)$ such that for $j > N(k)$, the graph G_j has no cycle of length at most k. Clearly, if H has no cycle of length at most $k-1$, then any 2-lift of it has the same property. So it is enough to prove that if H has no cycle of length at most $k-1$, then there exists an H' obtained from H by a sequence of 2-lifts without cycle of length at most k. We show that if the girth $g(H) = k$, then there exists a lift of H with less number of k-cycles than H. Let X be the random variable counting the number of k-cycles in a random 2-lift of H. Every k-cycle of H lifts to two k-cycles or a $2k$-cycle with probability $1/2$ each, so $\mathbb{E}X$ is exactly the number of k-cycles of H. But $H \cup H$ has two times as many k-cycles than H, so there must be a lift with strictly fewer k-cycles than H has. Choose this 2-lift and iterate this step to obtain an H' with girth at least $k+1$. □

7.6 Matchings

Theorem 7.6.1 (Csikvári [9]). *Let G be a d–regular bipartite graph on $v(G) = 2n$ vertices, and let $m_k(G)$ denote the number of matchings of size k. Let $0 \le p \le 1$, then*

$$\sum_{k=0}^{n} m_k(G) \left(\frac{p}{d}\left(1 - \frac{p}{d}\right)\right)^k (1-p)^{2(n-k)} \ge \left(1 - \frac{p}{d}\right)^{nd}.$$

Proof. Let $M(G,t) = \sum_{k=0}^{n} m_k(G)t^k$. In Theorem 7.3.3 we have seen that if G is a bipartite graph and H is a 2-lift of G then $m_k(H) \le m_k(G \cup G)$. Consequently,

$$M(H,t) \le M(G \cup G, t) = M(G,t)^2.$$

This implies that

$$\frac{\ln M(H,t)}{v(H)} \leq \frac{\ln M(G,t)}{v(G)}.$$

Let us introduce the notation

$$p_t(G) = \frac{\ln M(G,t)}{v(G)}.$$

By Lemma 7.5.1 we can choose a sequence of 2–lifts (G_i) such that $g(G_i) \to \infty$. Next we show that it implies that the sequence $\frac{\ln M(G_i,t)}{v(G_i)}$ is convergent. Note that if G is a graph on $v(G) = 2n$ vertices and

$$\mu(G,x) = \sum_{k=0}^{\lfloor v(G)/2 \rfloor} (-1)^k m_k(G) x^{v(G)-2k} = \prod_{i=1}^{v(G)} (x - \lambda_i),$$

or equivalently

$$M(G,t) = \sum_{k=0}^{\lfloor v(G)/2 \rfloor} m_k(G) t^k = \prod_{\lambda_i > 0} (1 + \lambda_i^2 t)$$

then

$$\frac{1}{v(G)} \ln M(G,t) = \frac{1}{v(G)} \sum_{i=1}^{v(G)/2} \ln(1 + \lambda_i^2 t) = \int \frac{1}{2} \ln(1 + tz^2)\, d\rho_G^{(m)}(z).$$

Now we see that $\frac{1}{2}\ln(1+tz^2)$ is a bounded continuous function on the interval $[-2\sqrt{\Delta - 1}, 2\sqrt{\Delta - 1}]$ for every fixed t. Since the sequence of measures $(\rho_{G_i}^{(m)})$ is weakly convergent this implies that $\frac{\ln M(G_i,t)}{v(G_i)}$ is convergent for every Benjamini–Schramm convergent graph sequence. In fact, in the particular case when the sequence (G_i) satisfies $g(G_i) \to \infty$, we can compute the limit. We have seen that the limit measure is the Kesten-McKay measure $\rho_{\mathbb{T}_d}$, see Theorem 5.3.5. Since we have

$$\lim_{i \to \infty} \int \frac{1}{2} \ln(1 + tz^2)\, d\rho_{G_i}^{(m)}(z) = \int \frac{1}{2} \ln(1 + tz^2)\, d\rho_{\mathbb{T}_d}(z)$$

we can use Lemma 5.3.6 to get that

$$p_t(\mathbb{T}_d) = \int \frac{1}{2} \ln(1 + tz^2)\, d\rho_{\mathbb{T}_d}(z) = \frac{1}{2} \ln S_d(t),$$

where

$$S_d(t) = \frac{1}{\eta_t^2} \left(\frac{d-1}{d - \eta_t} \right)^{d-2} \quad \text{with} \quad \eta_t = \frac{\sqrt{1 + 4(d-1)t} - 1}{2(d-1)t}.$$

It is worth introducing the following substitution: $t = \frac{\frac{p}{d}\left(1 - \frac{p}{d}\right)}{(1-p)^2}$. As p runs through the interval $[0,1)$, t runs through the interval $[0,\infty)$ and we have

$$\eta_t = \frac{1-p}{1 - \frac{p}{d}} \quad \text{and} \quad S_d(t) = \frac{\left(1 - \frac{p}{d}\right)^d}{(1-p)^2}.$$

In summary, by a sequence of well-chosen 2-lifts we know that

$$p_t(G_0) \geq p_t(G_1) \geq p_t(G_2) \geq p_t(G_3) \geq \dots$$

and the sequence converges to $p_t(\mathbb{T}_d)$, hence $p_t(G) \geq p_t(\mathbb{T}_d)$ for any d–regular bipartite graph G. In other words, $\frac{1}{2n}\ln M(G,t) \geq \frac{1}{2}\ln S_d(t)$. With the substitution $t = \frac{\frac{p}{d}\left(1-\frac{p}{d}\right)}{(1-p)^2}$ we arrive to the inequality

$$M\left(G, \frac{\frac{p}{d}\left(1-\frac{p}{d}\right)}{(1-p)^2}\right) \geq \frac{1}{(1-p)^{2n}}\left(1-\frac{p}{d}\right)^n.$$

After multiplying by $(1-p)^{2n}$, we get that

$$\sum_{k=0}^{n} m_k(G)\left(\frac{p}{d}\left(1-\frac{p}{d}\right)\right)^k (1-p)^{2(n-k)} \geq \left(1-\frac{p}{d}\right)^{nd}.$$

This is true for all $p \in [0,1)$ and so by continuity it is also true for $p=1$, where it directly reduces to Schrijver's theorem since all, but the last term vanishes on the left hand side.

□

8. Bethe Approximation

In this chapter we introduce the Bethe partition function $Z_B(\mathcal{G})$ for a factor graph $\mathcal{G} = (F, V, E, (g_a)_{a\in F})$. First we need to introduce the pseudo-marginal polytope.

For each variable node v we introduce a probability distribution b_v on $\mathcal{X}$, and for each function node a we also introduce a probability distribution b_a on $\mathcal{X}^{\partial a}$:

$$\sum_{x\in\mathcal{X}} b_v(x) = 1 \quad \forall v \in V, \quad b_v(x) \geq 0 \quad \forall x \in \mathcal{X},$$

and

$$\sum_{\underline{x}\in\mathcal{X}^{\partial a}} b_a(\underline{x}) = 1 \quad \forall a \in F, \quad b_a(\underline{x}) \geq 0 \quad \forall \underline{x} \in \mathcal{X}^{\partial a}.$$

Furthermore, b_v and b_a have to be consistent in the following sense: for all $c \in \mathcal{X},\ a \in F, v \in \partial a$ we have

$$\sum_{\underline{x}\in\mathcal{X}^{\partial a\setminus v}} b_a(\underline{x}, c) = b_v(c).$$

We will call a $\underline{b} = ((b_v)_{v\in V}, (b_a)_{a\in F})$ a locally consistent set of marginals. The set of such $\underline{b}$ will be denoted by $\mathrm{Mar}(\mathcal{G})$.

Let $\mathbb{F}$ be the following function evaluated on a $\underline{b} \in \mathrm{Mar}(\mathcal{G})$:

$$\mathbb{F}(\underline{b}) = \sum_{a\in F} \sum_{\underline{x}\in\mathcal{X}^{\partial a}} b_a(\underline{x}) \ln \frac{g_a(\underline{x})}{b_a(\underline{x})} - \sum_{v\in V} (1 - |\partial v|) \sum_{x\in\mathcal{X}} b_v(x) \ln b_v(x).$$

Finally, let

$$H_B(\mathcal{G}) = \sup_{\underline{b}\in\mathrm{Mar}(\mathcal{G})} \mathbb{F}(\underline{b}),$$

and

$$Z_B(\mathcal{G}) = \exp(H_B(\mathcal{G})).$$

Here $H_B(\mathcal{G})$ is the Bethe free entropy, and $Z_B(\mathcal{G})$ is the Bethe partition functions.

As we have seen the motivation partly comes from studying covers. In fact, Bethe approximation was defined much before any connection with graph covers was revealed. Bethe approximation has at least three ways to approach. We have seen two: the variational problem and graph covers. A third one is coming from message passing algorithms.

It will turn out that $Z_B(\mathcal{G})$ is often a very good approximation of $Z(\mathcal{G})$. Furthermore, it is an important problem to find sufficient conditions under which

$$Z(\mathcal{G}) \geq Z_B(\mathcal{G}).$$

This question is very strongly related to both Schrijver's theorem on the number of perfect matchings, and to Sidorenko's conjecture on the number of homomorphisms. We will see two different sufficient condition ensuring the above inequality: a theorem of D. Straszak and N. Vishnoi [34], and a theorem of N. Ruozzi (Theorem 8.1.2) [29].

Example 8.0.1. As a continuation of Example 1.2.2 let us compute the Bethe partition function for this factor graph. For every edge $(u, v) \in E(G)$ there is a probability distribution $b_{(u,v)}$ on $[q] \times [q]$ such that the first marginal is b_u for every $v \in N_G(u)$. Then

$$\mathbb{F}(\underline{b}) = \sum_{(u,v)\in E(H)} \sum_{(s,t)\in[q]^2} b_{(u,v)}(s,t) \ln \frac{A_{s,t}}{b_{(u,v)}(s,t)} + \sum_{u\in V(H)} (d(u)-1) \sum_{s\in[q]} b_u(s) \ln b_u(s).$$

It is very tempting to assume that for all edges $b_{(u,v)}$ is the same b_e, and if the graph is non-bipartite then to assume that it is a symmetric distribution: $b_{(u,v)}(s,t) = b_{(u,v)}(t,s)$ for all $t, s \in [q]^2$, this way we can ensure that the marginals are the same for both u and v. If the graph is bipartite with bipartite classes $V_1(H), V_2(H)$, then this assumption on symmetry is unnecessary. In this case, b_u is equal to some b_1 for all $u \in V_1(H)$, and equal to b_2 for all $u \in V_2(H)$. Hence for a bipartite graph we get that

$$\mathbb{F}(b) = e(H) \sum_{(s,t)\in[q]^2} b_e(s,t) \ln \frac{A_{s,t}}{b_e(s,t)} + \sum_{j=1}^{2} (e(H) - |V_j(H)|) \sum_{s\in[q]} b_j(s) \ln b_j(s).$$

Now let us further assume that b_e is the same distribution τ for every edge that satisfies $\tau(s,t) = \frac{A_{s,t}}{A}$, where $A = \sum_{s,t} A_{s,t}$. Furthermore, set $A_{s,\cdot} = \sum_t A_{s,t}$ and $A_{\cdot,t} = \sum_s A_{s,t}$ and let τ_1 and τ_2 be the marginals of τ to the first and second coordinates. Then we get that

$$\begin{aligned} Z_B(\mathcal{G}) &\geq \exp\left(e(G) \sum_{s,t} \frac{A_{s,t}}{A} \ln A + \sum_{j=1}^{2} (e(H) - |V_j(H)|) \sum_{s\in[q]} \tau_j(s) \ln \tau_j(s) \right) \\ &= \exp\left(e(G) \ln A + \sum_{j=1}^{2} (e(G) - |V_j(G)|)(-H(\tau_j)) \right) \\ &\geq \exp\left(e(G) \ln A + \sum_{j=1}^{2} (e(G) - |V_j(G)|)(-\ln q) \right) \\ &= q^{v(G)} \left(\frac{A}{q^2} \right)^{e(G)} \end{aligned}$$

In the fourth step we used Jensen's inequality to the concave function $\ln x$, and $p_i = \frac{1}{q}$ and $a_i = A_{s,\cdot}$, or equivalently that the entropy is maximized by the uniform distribution.

Hence the inequality $Z(\mathcal{G}) \geq Z_B(\mathcal{G})$ would imply that the Sidorenko's conjecture is true.

Example 8.0.2. As a continuation of Example 1.2.3 let us compute the Bethe free entropy in case of permanents. Note that $\mathbb{F}(\underline{b}) = -\infty$ if there is some $\underline{x}$ for which $b_a(x) > 0$ and $g_a(\underline{x}) = 0$. So we only need to consider those locally consistent marginals that are supported on $\{\underline{x} \mid g_a(\underline{x}) > 0\}$. Let e_k be the vector that takes value 1 at k. coordinate and 0 everywhere else. In case of permanents the support $\{\underline{x} \mid g_a(\underline{x}) > 0\} = \{e_1, \dots, e_n\}$. This means that for all rows r_i and columns c_j there will be a probability distribution $(r^i_1, r^i_2, \dots r^i_n)$ and $(c^j_1, c^j_2, \dots c^j_n)$ such that $b_{r_i}(e_k) = r^i_k$ and $b_{c_j}(e_k) = c^j_k$. The local consistence means that

$$r^i_k = b_{(i,j)}(1) = c^j_k.$$

So if we introduce the matrix B for which $B_{ij} = b_{(i,j)}(1)$ then the rows will correspond to b_{r_i} and the columns will correspond to b_{c_j}. In particular, this is a doubly stochastic matrix. Since $|\partial v| = 2$ for all $v = (i,j)$, we get that

$$\mathbb{F}(\underline{b}) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} B_{ij} \ln \frac{A_{ij}^{1/2}}{B_{ij}}\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n} B_{ij} \ln \frac{A_{ij}^{1/2}}{B_{ij}}\right) + \sum_{i,j}(B_{ij} \ln B_{ij} + (1-B_{ij}) \ln(1-B_{ij})).$$

In other words,

$$\mathbb{F}(\underline{b}) = \sum_{i,j} B_{ij} \ln \frac{A_{ij}}{B_{ij}^2} + \sum_{i,j}(B_{ij} \ln B_{ij} + (1-B_{ij}) \ln(1-B_{ij})).$$

If we introduce the function $H(x) = -x\ln(x) - (1-x)\ln(1-x)$ then we can rewrite it as

$$\mathbb{F}(\underline{b}) = \sum_{i,j} B_{ij} \ln \frac{A_{ij}}{B_{ij}^2} - \sum_{i,j} H(B_{ij}).$$

Alternatively, we cancel some terms:

$$\mathbb{F}(\underline{b}) = \sum_{i,j} B_{ij} \ln \frac{A_{ij}}{B_{ij}} + \sum_{i,j}(1-B_{ij}) \ln(1-B_{ij}).$$

Hence

$$Z_B(\mathcal{G}) = \sup_{B \in \mathrm{DS}_n} \exp\left(\sum_{i,j} B_{ij} \ln \frac{A_{ij}}{B_{ij}} + \sum_{i,j}(1-B_{ij}) \ln(1-B_{ij})\right),$$

where DS_n is the set of doubly stochastic matrices. Later we will prove that for this factor graph we have $Z(\mathcal{G}) \geq Z_B(\mathcal{G})$, i.e,

$$\mathrm{per}(A) \geq \sup_{B \in \mathrm{DS}_n} \exp\left(\sum_{i,j} B_{ij} \ln \frac{A_{ij}}{B_{ij}} + \sum_{i,j}(1-B_{ij}) \ln(1-B_{ij})\right).$$

In particular, if A is itself doubly stochastic we can choose $B = A$ and get that

$$\ln \mathrm{per}(A) \geq \sum_{ij}(1-A_{ij}) \ln(1-A_{ij}).$$

Now suppose that A is the biadjacency matrix of a d–regular bipartite graph. Then $\mathrm{per}(A) = \mathrm{pm}(G)$ and $B = \frac{1}{d}A$ is a doubly stochastic matrix. Hence

$$\begin{aligned} Z_B(\mathcal{G}) &\geq \exp\left(\sum_{i,j} B_{ij} \ln \frac{A_{ij}}{B_{ij}} + \sum_{i,j} (1 - B_{ij}) \ln(1 - B_{ij})\right) \\ &= \exp\left(nd \cdot \frac{1}{d} \ln(d) + nd \cdot \frac{d-1}{d} \ln \frac{d-1}{d}\right) \\ &= \left(\frac{(d-1)^{d-1}}{d^{d-2}}\right)^n. \end{aligned}$$

Hence

$$\mathrm{pm}(G) \geq Z_B(\mathcal{G}) \geq \left(\frac{(d-1)^{d-1}}{d^{d-2}}\right)^n.$$

Hence we obtained a new proof of Schrijver's theorem!

Example 8.0.3. Let us consider the factor graph described in Example 1.2.4. The situation is very similar to the case of permanents. The slight difference is that the support $\{\underline{x} \mid g_a(\underline{x}) > 0\} = \{\underline{0}, e_1, \dots, e_n\}$. The local consistence still gives that

$$b_{r_i}(e_j) = b_{(i,j)}(1) = b_{c_j}(e_i).$$

So if we introduce the matrix B for which $B_{ij} = b_{(i,j)}(1)$ then the rows will correspond to b_{r_i} and the columns will correspond to b_{c_j}, but the sum in the rows and columns will be at most 1. It is worth introducing

$$B_{i,0} = 1 - \sum_{j=1}^{n} B_{i,j} = b_{r_i}(\underline{0}),$$

and

$$B_{0,j} = 1 - \sum_{i=1}^{n} B_{i,j} = b_{c_j}(\underline{0}).$$

It is also worth introducing $A_{i,0} = A_{0,j} = 1$. Then

$$\mathbb{F}(\underline{b}) = \sum_{1 \leq i,j \leq n} B_{ij} \ln \frac{A_{ij}}{B_{ij}^2} + \sum_{i=1}^{n} B_{i0} \ln \frac{1}{B_{i0}} + \sum_{j=1}^{n} B_{0j} \ln \frac{1}{B_{0j}} + \sum_{1 \leq i,j \leq n} (B_{ij} \ln B_{ij} + (1 - B_{ij}) \ln(1 - B_{ij})).$$

After canceling some terms we get that

$$\mathbb{F}(\underline{b}) = \sum_{1 \leq i,j \leq n} B_{ij} \ln \frac{A_{ij}}{B_{ij}} + \sum_{i=1}^{n} B_{i0} \ln \frac{1}{B_{i0}} + \sum_{j=1}^{n} B_{0j} \ln \frac{1}{B_{0j}} + \sum_{1 \leq i,j \leq n} (1 - B_{ij}) \ln(1 - B_{ij}).$$

Later we will prove that for this factor graph we have $Z(\mathcal{G}) \geq Z_B(\mathcal{G})$, i.e,

$$\mathrm{allper}(A) \geq \sup_{B \in \mathrm{SDS}_n} \exp(\mathbb{F}(\underline{b})),$$

where SDS_n is the set of non-negative matrices such that in each row and each column the sum of the elements is at most 1.

8.1 Covers of factor graphs

Since a factor graph is just a bipartite graph equipped with certain functions it is straightforward to generalize the concept of k-covers to factor graphs. The only extra thing we have to add to the definition is that for each lift of the same function node a has to get the same function $g_a(\underline{x})$. The following theorem of Vontobel gives a combinatorial characterization of Bethe partition function in terms of k-covers.

Theorem 8.1.1 (Vontobel [35]). *Let $\mathcal{G}$ be a factor graph, and let $C_N(\mathcal{G})$ be the set of all N-covers of $\mathcal{G}$. Then*

$$Z_B(\mathcal{G}) = \lim_{N\to\infty} \left(\mathbb{E}Z(\mathcal{H})\right)^{1/N}$$

exists, where $\mathbb{E}Z(\mathcal{H})$ is the average of $Z(\mathcal{H})$ over all elements of $C_N(\mathcal{G})$.

Proof. Let $|\mathcal{X}| = q$. First of all, the number of N-covers is $(N!)^{e(\mathcal{G})}$, because at each edge we have $N!$ choices to choose the perfect matching. For a function node a and a vector $\underline{x} \in \mathcal{X}^{d_a}$ let $s_a(\underline{x})$ denote the number of lifts of the vertex a who sees exactly the vector $\underline{x}$ at the adjacent variable nodes. So $\sum_{\underline{x}\in\mathcal{X}^{d_a}} s_a(\underline{x}) = N$. If we know the values of $s_a(\underline{x})$ for each function node a, then the contribution of this configuration to $Z(\mathcal{H})$ is

$$\prod_{a\in F} \prod_{\underline{x}\in\mathcal{X}_a^d} g_a(\underline{x})^{s_a(\underline{x})}.$$

Suppose now that we are given the numbers $s_a(\underline{x})$ for all $a \in F$ and $\underline{x} \in \mathcal{X}^{d_a}$, we would like to count the number of N-lifts that has a configuration with these numbers. First of all, observe that these numbers determine the distributions of the value of the variable at each variable node. So for $v \in V$ and $x \in \mathcal{X}$ let $s_v(x)$ denote the number of lifts of a with value x. Clearly, $\sum_{x\in\mathcal{X}} s_v(x) = N$. Now we can count the number of covers with all these numbers via the half-edge trick: we first decide for each function node a what they will see, then we make half-edges from a and write the corresponding variables to the end of the half-edges; then we decide for each variable node v their values and again create half-edges coming out of them; finally we pair up the half-edges in a compatible way, that is, having the same variable at their ends.

At a function node a we can decide what the lifts of a sees in

$$\binom{N}{s_a(\underline{x}_1), \dots, s_a(\underline{x}_m)},$$

ways, where $m = q^{d_a}$. Similarly, for a variable node v we can decide what the lifts of v sees in

$$\binom{N}{s_v(x_1), \dots, s_v(x_q)}$$

ways. Once the half-edges created we can match them up in

$$s_v(x_1)! \cdots s_v(x_q)!$$

ways. So the total contribution is

$$\prod_{a\in F}\binom{N}{s_a(\underline{x}_1),\dots,s_a(\underline{x}_m)}\prod_{v\in V}\binom{N}{s_v(x_1),\dots,s_v(x_q)}\prod_{(a,v)\in E(G)}(s_v(x_1)!\cdots s_v(x_q)!)\prod_{a\in F}\prod_{\underline{x}\in\mathcal{X}_a^d} g_a(\underline{x})^{s_a(\underline{x})}.$$

Now recall that we take the average over all covers, so we dived by $(N!)^{e(\mathcal{G})}$. After division, the contribution is

$$\prod_{a\in F}\binom{N}{s_a(\underline{x}_1),\dots,s_a(\underline{x}_m)}\prod_{v\in V}\binom{N}{s_v(x_1),\dots,s_v(x_q)}^{1-|\partial v|}\prod_{a\in F}\prod_{\underline{x}\in\mathcal{X}_a^d} g_a(\underline{x})^{s_a(\underline{x})}.$$

In general, if we have non-negative numbers $k_1,\dots,k_r$ such that $k_1+\dots+k_r=N$, then with the notation $p_i=k_i/N$ we have

$$\binom{N}{k_1,\dots,k_r}=\exp\left((1+o(1))N\sum_{i=1}^r p_i\ln\left(\frac{1}{p_i}\right)\right).$$

This is a quick consequence of Stirling's formula. So let us introduce the notation $b_a(\underline{x})=s_a(\underline{x})/N$, and $b_v(x)=s_v(x)/N$. Then the above expression can be written as

$$\exp\left((1+o(1))N\left(\sum_{a\in F}\sum_{\underline{x}\in\mathcal{X}^{\partial a}} b_a(\underline{x})\ln\frac{g_a(\underline{x})}{b_a(\underline{x})}-\sum_{v\in V}(1-|\partial v|)\sum_{x\in\mathcal{X}} b_v(x)\ln b_v(x)\right)\right).$$

Next observe that the number of terms is polynomial in N: for instance, the number N can be decomposed as a sum of q numbers in $\binom{N+q-1}{q-1}$ ways. In fact, the trivial upper bound gives $(N+1)^q$: any $s_v(x)$ can take $N+1$ values from 0 to N. So the number of terms is at most

$$\prod_{v\in V}(N+1)^q\cdot\prod_{a\in F}(N+1)^{q^{d_a}}=(N+1)^{q|V|+q^{e(\mathcal{G})}}=\exp(o(N)).$$

This means that the sum and the maximum value differ in a subexponential term. Let

$$\mathbb{F}(\underline{b})=\sum_{a\in F}\sum_{\underline{x}\in\mathcal{X}^{\partial a}} b_a(\underline{x})\ln\frac{g_a(\underline{x})}{b_a(\underline{x})}-\sum_{v\in V}(1-|\partial v|)\sum_{x\in\mathcal{X}} b_v(x)\ln b_v(x).$$

Then

$$(\mathbb{E}Z(\mathcal{H}))^{1/N}=\exp((1+o(1))\max_{\underline{b}}\mathbb{F}(\underline{b})),$$

where the maximum is taken over all probability distributions b_a, b_v that are consistent in the sense that the marginal of b_a to the coordinate corresponding to the vertex v is exactly b_v, and the probabilities are of the form s/N, where s is an integer. If we drop this last condition, then we clearly get a maximum at least as large as this one. On the other hand, any probability distributions b_a, b_v can be approximated arbitrarily well with this extra condition if N is large enough. Hence

$$\lim_{N\to\infty}(\mathbb{E}Z(\mathcal{H}))^{1/N}=\exp(\max_{\underline{b}}\mathbb{F}(\underline{b})),$$

where the maximum is taken over all probability distributions b_a, b_v that are consistent in the sense that the marginal of b_a to the coordinate corresponding to the vertex v is exactly b_v.

Later we will call this quantity the Bethe approximation of the partition function. □

The simplest k-cover of $\mathcal{G}$ is of course just the union of k disjoint copies of $\mathcal{G}$. Let us denote it by $\mathcal{G}^k$. Clearly, $Z(\mathcal{G})^k = Z(\mathcal{G}^k)$. If we could prove that $Z(\mathcal{G}^k) \geq Z(\mathcal{H})$ for any $\mathcal{H} \in C_k(\mathcal{G})$ then Vontobel's theorem, Theorem 8.1.1, would immediately imply that $Z(\mathcal{G}) \geq Z_B(\mathcal{G})$. We have seen that Ruozzi's theorem (Theorem 7.4.1) provides a simple sufficient condition implying that $Z(\mathcal{G}^k) \geq Z(\mathcal{H})$ for binary factor graphs. Hence it implies the following theorem.

Theorem 8.1.2 (Ruozzi [29]). *Suppose that for the factor graph $\mathcal{G} = (F, V, E, (g_a)_{a \in F})$ all functions g_a are log-supermodular. Then $Z(\mathcal{G}) \geq Z_B(\mathcal{G})$.*

Proof. This theorem immediately follows from Theorem 7.4.1 and Theorem 8.1.1.

□

9. Stable Polynomials

9.1 Multivariate polynomials

The goal of this section is the study of the so-called stable polynomials.

Definition 9.1.1. Let $\Omega \subseteq \mathbb{C}^n$. A multivariate polynomial $p(z_1, \dots, z_n) \in \mathbb{C}[z_1, \dots, z_n]$ is called Ω-stable if $p(z_1, \dots, z_n) \neq 0$ if $z_i \in \Omega$ for all $1 \leq i \leq n$. When

$$\Omega = \mathbb{H} = \{(z_1, \dots, z_n) \mid \operatorname{Im}(z_i) > 0 \text{ for all } \ 1 \leq i \leq n\},$$

then we simply say that $p(z_1, \dots, z_n)$ is a stable polynomial (instead of $\mathbb{H}$-stable). We say that p is real stable if it is stable and its coefficients are real.

The motivation behind stability is the following. Suppose that $P(z) = \sum_{k=0}^n a_k z^k$ is a univariate polynomial with only real coefficients such that $P(z) \neq 0$ if $\operatorname{Im}(z) > 0$. Then all zeros of $P(z)$ are real. So one can think of an stability as a multivariate version of real-rootedness.

Stable polynomials show up in various combinatorial problems. In this chapter we will use them to give lower bounds on permanents and the number of perfect matchings. Some form of negative correlation can be grasped by the use of stable polynomials.

9.1.1 Univariate real-rooted polynomials

Theorem 9.1.2 (Lucas). *Let $f(z) = \sum_{i=1}^n a_i z^i$ be non-constant polynomial. Then the zeros of $f'(z)$ is in the convex hull of the zeros of $f(z)$. In particular, if $f(z)$ is real-rooted, then $f'(z)$ is also real-rooted.*

Proof. Let the zeros of $f(z)$ be $\rho_1, \dots, \rho_n$, and let γ be a zero of $f'(z)$. Then

$$f'(z) = f(z) \sum_{i=1}^n \frac{1}{z - \rho_i}.$$

If γ is among the zeros of $f(z)$ then the statement is trivially true. So suppose that $f(\gamma) \neq 0$. Then from

$$0 = f'(\gamma) = f(\gamma) \sum_{i=1}^n \frac{1}{\gamma - \rho_i}$$

we get that

$$0 = \sum_{i=1}^{n} \frac{1}{\gamma - \rho_i}.$$

Note that

$$\sum_{i=1}^{n} \frac{1}{\gamma - \rho_i} = \sum_{i=1}^{n} \frac{\overline{\gamma} - \overline{\rho_i}}{|\gamma - \rho_i|^2},$$

and so

$$\overline{\gamma} \sum_{i=1}^{n} \frac{1}{|\gamma - \rho_i|^2} = \sum_{i=1}^{n} \frac{\overline{\rho_i}}{|\gamma - \rho_i|^2}.$$

Next let us conjugate both sides:

$$\gamma \sum_{i=1}^{n} \frac{1}{|\gamma - \rho_i|^2} = \sum_{i=1}^{n} \frac{\rho_i}{|\gamma - \rho_i|^2}.$$

Now set

$$c_i = \frac{\frac{1}{|\gamma - \rho_i|^2}}{\sum_{i=1}^{n} \frac{1}{|\gamma - \rho_i|^2}}.$$

Then

$$\gamma = \sum_{i=1}^{n} c_i \rho_i$$

Here $c_i \geq 0$ and $\sum_{i=1}^{n} c_i = 1$, so γ is a convex combination of ρ_i's. □

Theorem 9.1.3 (Newton)**.** *Let $f(z) = \sum_{k=0}^{n} a_k z^k$ be a real-rooted polynomial of degree n with $a_k \geq 0$. Then for all $1 \leq k \leq n-1$ we have*

$$\frac{a_{k-1}}{\binom{n}{k-1}} \frac{a_{k+1}}{\binom{n}{k+1}} \leq \left(\frac{a_k}{\binom{n}{k}} \right)^2.$$

Proof. First, let us differentiate the polynomial $f(z)$ $k-1$ times. Then the obtained polynomial is

$$g(z) = \frac{d^{k-1}}{dz^{k-1}} f(z) = \sum_{j=k-1}^{n} j(j-1)\dots(j-(k-2)) a_k z^{j-(k-1)}.$$

Then $g(z)$ is a real-rooted polynomial by applying Lucas's theorem (Theorem 9.1.2) successively. The polynomial $h(z) = z^{n-(k-1)} g\left(\frac{1}{z}\right)$ is also real rooted:

$$h(z) = \sum_{j=k-1}^{n} j(j-1)\dots(j-(k-2)) a_j z^{n-j}.$$

Now let us differentiate $h(z)$ $n-(k+1)$ times, the obtained polynomial is a degree 2 polynomial:

$$a_{k-1} \frac{(k-1)!(n-(k-1))!}{2} z^2 + a_k k!(n-k)! z + a_{k+1} \frac{(k+1)!(n-(k+1))!}{2}.$$

This polynomial is again real-rooted by Lucas's theorem. Since a quadratic polynomial $az^2 + bz + c$ is real-rooted if and only if $b^2 \geq 4ac$ we get that

$$(a_k k!(n-k)!)^2 \geq (a_{k-1}(k-1)!(n-(k-1))!)(a_{k+1}(k+1)!(n-(k+1))!).$$

After dividing by $n!^2$ we get that

$$\frac{a_{k-1}}{\binom{n}{k-1}}\frac{a_{k+1}}{\binom{n}{k-1}} \leq \left(\frac{a_k}{\binom{n}{k}}\right)^2.$$

□

The following criterion gives an easy way to check whether a polynomial is real-rooted.

Theorem 9.1.4 (Hermite-Sylvester criterion)**.** *Let $P(x) = \sum_{k=0}^{n} a_k x^k = \prod_{j=1}^{n}(x - \alpha_i)$ be a polynomial with real coefficients. Let*

$$m_k = \sum_{j=1}^{n} \alpha_j^k.$$

Then $P(x)$ is real-rooted if and only the $n \times n$ matrix $M = (m_{i+j-2})_{1 \leq i,j \leq n}$ is positive semi-definite.

We follow the argument of the paper [26].

Proof. First suppose that $P(x)$ is real-rooted and consider the column vector $v_j = (1, \alpha_j, \alpha_j^2, \dots, \alpha_j^{n-1}) \in \mathbb{R}^n$. Then $M = \sum_j v_j v_j^T$ and so M is positive semidefinite.

Next suppose that $P(x)$ has a non-real root, say α_1 and $\alpha_2 = \overline{\alpha_1}$. If $v = (c_0, c_1, \dots, c_{n-1})$ and $Q(z) = \sum_{j=0}^{n-1} c_j z^j$. Then

$$v^T M v = \sum_j \left(\sum_r c_r \alpha_j^r\right)^2 = \sum_j Q(\alpha_j)^2.$$

The idea is that we create a polynomial Q such that $Q(\alpha_1) = i, Q(\alpha_2) = -i$ and if $\alpha_k \neq \alpha_1, \alpha_2$, then $Q(\alpha_k) = 0$. Then $v^T M v = -2m_1$, where m_1 is the multiplicity of α_1. If r is the number of different roots of P, then we know that there is such a polynomial Q of degree at most $r-1$ by Lagrange interpolation. What we do not know a priori –but true– that it has real coefficients. Indeed, if $Q(z) = \sum_{j=0}^{n-1} c_j z^j$ and $Q'(z) = \sum_{j=0}^{n-1} \overline{c_j} z^j$, then for $\lambda \in \{\alpha_1, \dots, \alpha_n\}$ we have

$$Q(\lambda) = Q\left(\overline{\overline{\lambda}}\right) \stackrel{*}{=} \overline{Q(\overline{\lambda})} = \overline{\sum_{j=0}^{n-1} c_j \overline{\lambda}^j} = \sum_{j=0}^{n-1} \overline{c_j} \lambda^j = Q'(\lambda).$$

In the second step we used that $\lambda \in \{\alpha_1, \dots, \alpha_n\}$. So Q and Q' are polynomials of degree at most $r-1$ that are equal on r different places. Hence $Q = Q'$, that is, Q has real coefficients. Since $v^T M v < 0$ we get that M is not positive semidefinite. □

Theorem 9.1.5. *Let P be a polynomial and suppose that $(P_n)_n$ is a sequence of real-rooted polynomials such that $P_n \to P$ coefficientwise. Then P is real–rooted.*

Proof. Suppose that P has degree d. Let $M_n = (m^{(n)}_{i+j-2})_{1\le i,j\le d}$ be the matrix consisting of the entries $m(n)_k = \sum \alpha^k_{n,j}$, where $\alpha_{n,j}$ are the roots of P_n. Then M_n is positive semidefinite. Note that the elements of M_n can be expressed by the coefficients of P_n by the Newton-Girard formulas. Hence $M_n \to M$ elementwise. Then M is positive semidefinite. Hence P is real-rooted. □

9.1.2 Real stable polynomials

Theorem 9.1.6. *A multivariate polynomial $f(z_1, \ldots, z_n) \in \mathbb{R}[z_1, \ldots, z_n]$ is stable if and only if for all $\underline{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ and $\underline{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n_{>0}$, the univariate polynomial $g(t) = f(v_1 + tu_1, \ldots, v_n + tu_n)$ is real-rooted.*

Proof. First suppose that $g(t)$ is not real rooted, then it has a root $a+bi \in C$, where $b \neq 0$. Since the coefficients of g are real, $a \pm bi$ are both zero of g. So we can assume that $b > 0$. But then $\mathrm{Im}(v_j + u_j(a+bi)) = bu_j > 0$ shows that the numbers $z_j = v_j + u_j(a+bi) \in \mathbb{H}$ for $j = 1, \ldots, n$ and $f(z_1, \ldots, z_n) = 0$, thus f is not stable.

Next suppose that f is not stable, it has some zero $(z_1, \ldots, z_n) \in \mathbb{H}^n$. Let $z_j = a_j + b_j i$. Then $b_j > 0$. Then the polynomial $g(t) = f(a_1 + tb_1, \ldots, a_n + tb_n)$ satisfies that $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $(b_1, \ldots, b_n) \in \mathbb{R}^n_{>0}$ and has a non-real zero, namely i. □

Theorem 9.1.7 (Hurwitz)**.** *Let Ω be an open connected subset of $\mathbb{C}^n$. Suppose that the sequence of analytic functions $(f_k)_k$ converge to some function f such a way that the convergence is uniform on all compact subsets of Ω. If f_k is Ω-stable for all k then f is Ω-stable or the constant 0 function.*

Theorem 9.1.8. *Suppose that $f(z_1, \ldots, z_n)$ and $g(z_1, \ldots, z_n)$ are stable polynomial.*

(a) Then $f \cdot g$ is also stable.

(b) If $\sigma \in S_n$ is a permutation then $f(z_{\sigma(1)}, \ldots, z_{\sigma(n)})$ is a stable polynomial.

(c) If $c \in \mathbb{C}$ and $a_1, \ldots, a_n \in \mathbb{R}_+$ then $cf(a_1z_1, \ldots, a_nz_n)$ is a stable polynomial.

(d) If $i, j \in [n]$ then $f(z_1, \ldots, z_n)\big|_{z_i=z_j}$ is a stable polynomial.

(e) If $a \in H$ then $f(a, z_2, \ldots, z_n)$ is a stable polynomial.

(f) If $\deg_{z_1}(f) = d$ then $z_1^d f(-z_1^{-1}, z_2, \ldots, z_n)$ is a stable polynomial.

Proof. All statements are trivial. In part (f) observe that the map $z \mapsto -z^{-1}$ is a map from $\mathbb{H}$ to $\mathbb{H}$.

□

Theorem 9.1.9. *Suppose that $f(z_1,\dots,z_n)$ is a stable polynomial.*

(a) If $a \in \mathbb{R}$ then $f(a,z_2,\dots,z_n)$ is either a stable polynomial or the constant 0 polynomial.

(b) $\partial_{z_1} f$ is also stable polynomial or the constant 0 polynomial. In general, $\partial^\alpha f$ is stable or the constant 0 polynomial.

Proof. To prove part (a) we can use Hurwitz's theorem together with part (e) of Theorem 9.1.8 applied to $f_n(z_2,\dots z_n) = f(a+\frac{i}{n}, z_2,\dots,z_n)$. Alternatively, one can avoid the use of Hurwitz's theorem as follows.

By Lemma 9.1.6 the polynomial $P(a+\varepsilon t, v_2+tu_2,\dots,v_n+tu_n)$ is real-rooted for every $\varepsilon, u_2,\dots u_n \in \mathbb{R}_{>0}$ and $a, v_2,\dots,v_n \in \mathbb{R}$. By letting $\varepsilon \to 0$ we get that $P(a, v_2+tu_2,\dots,v_n+tu_n)$ is real-rooted or the constant 0 function for every $u_2,\dots u_n \in \mathbb{R}_{>0}$ and $v_2,\dots,v_n$. Now using the the other direction of Lemma 9.1.6 we get that $P(a,x_2,\dots,x_n)$ is a real stable polynomial.

Next let us prove part (b). Let

$$P(x_1,\dots,x_n) = \sum_{k=0}^{d} P_k(x_2,\dots,x_n)x_1^k.$$

First we show that $P_d(x_2,\dots,x_n)$ is a real stable polynomial (and non-zero because ww cann assume that d is the degree of x_1). By part (a)

$$R(x_1,\dots,x_n) := x_1^d P(-1/x_1, x_2,\dots x_n) = \sum_{k=0}^{d} P_k(x_2,\dots,x_n)(-1)^k x_1^{d-k}$$

is real stable. Then $R(0,x_2,\dots,x_n) = (-1)^d P_d(x_2,\dots,x_n)$ is real stable, and so $P_d(x_2,\dots,x_n)$ is real stable.

Now let $Q = \frac{\partial}{\partial x_1}P$, and let $\underline{a} = (a_1,a_2,\dots,a_n) \in \mathbb{H}^n$. We show that if $Q \not\equiv 0$, then $\mathrm{Im}\left(\frac{Q(\underline{a})}{P(\underline{a})}\right) < 0$. Note that here we use that P is stable, so $P(\underline{a}) \neq 0$, and we can divide with it. If $d=0$ then $Q \equiv 0$. We can assume that $d \geq 1$. Let

$$g(x) := P(x,a_2,\dots,a_n) = \sum_{k=0}^{d} P_k(a_2,\dots,a_n)x^k.$$

Note that $P_d(a_2,\dots,a_n) \neq 0$ since P_d is real-stable. So $g(x)$ has degree $d \geq 1$. Then $g(x) = c\prod_{i=1}^{t}(x-\rho_i)$, and we have

$$\frac{g'(x)}{g(x)} = \sum_{i=1}^{t} \frac{1}{x-\rho_i}.$$

Note that $\mathrm{Im}(\rho_i) \leq 0$, otherwise $P(\rho_i,a_2,\dots,a_n) = 0$ would yield a zero in $\mathbb{H}^n$. Hence

$$\mathrm{Im}\left(\frac{Q(\underline{a})}{P(\underline{a})}\right) = \mathrm{Im}\left(\frac{g'(a_1)}{g(a_1)}\right) = \mathrm{Im}\left(\sum_{i=1}^{t}\frac{1}{a_1-\rho_i}\right) < 0.$$

In particular, this shows that $Q(\underline{a}) \neq 0$. Hence Q is stable. (Remark: we essentially repeated the proof of Gauss–Lucas theorem that asserts that zeros of the derivative of a polynomial lie in the convex hull of the zeros of the polynomial.) $\square$

The following theorem is a quick application of Theorem 9.1.9.

Theorem 9.1.10. *Suppose that the polynomial*

$$P(z_1, \dots, z_n, w_1, \dots w_k) = \sum_{\underline{m}} P_{\underline{m}}(w_1, \dots, w_k) \prod_{i=1}^{n} z_i^{m_i}$$

is a stable polynomial. Then for all $\underline{m}$ the polynomial $P_{\underline{m}}(w_1, \dots, w_k)$ is stable or 0.

Proof. The polynomial

$$\frac{\partial^{m_1+\dots+m_n}}{\partial^{m_1} z_1 \dots \partial^{m_n} z_n} P(z_1, \dots, z_n, w_1, \dots w_k))|_{z_1=\dots=z_n=0}$$

is stable or 0 by part (d) and (f) of Theorem 9.1.9. On the other hand, this is $P_{\underline{m}}(w_1, \dots, w_k) \prod_{i=1}^{n} m_i!$. Hence $P_{\underline{m}}(w_1, \dots, w_k)$ is stable or 0. $\square$

Theorem 9.1.11. *The elementary symmetric polynomial $E_k(x_1, \dots, x_n) = \sum_{S \subseteq [n], |S|=k} \prod_{i \in S} x_i$ is stable.*

Proof. The polynomial $P(z, x_1, \dots, x_n) = \prod_{i=1}^{n}(z + x_i)$ is trivially stable. Since

$$P(z, x_1, \dots, x_n) = \sum_{k=0}^{n} E_k(x_1, \dots x_n) z^k$$

the claim follows from Theorem 9.1.10. $\square$

Theorem 9.1.12. *(a) Let $a, b, c \geq 0$. Then the polynomial $ax^2 + bxy + cy^2$ is stable if and only if $b^2 \geq 4ac$.*

(b) Let $a, b, c, d \geq 0$. Then the polynomial $a + bx + cy + dxy$ is stable if and only if $bc \geq ad$.

9.1.3 Capacity of polynomials

Definition 9.1.13. The capacity $\operatorname{cap}(p)$ of a polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ is defined as

$$\operatorname{cap}(p) := \inf \frac{p(x_1, x_2, \dots, x_n)}{\prod_{j=1}^{n} x_j}$$

where the infimum ranges over all $(x_1, \dots, x_n) \in \mathbb{R}_+^n$.

Theorem 9.1.14. *Let $p \in \mathbb{R}_+[x_1, x_2, \dots, x_n]$ be stable and homogeneous of degree n. Then for*

$$q(x_1, \dots, x_{n-1}) = \frac{\partial}{\partial x_n} p(x_1, \dots, x_n)|_{x_n=0},$$

we have $q \equiv 0$ or q is stable. Moreover,

$$\operatorname{cap}(q) \geq \operatorname{cap}(p) r(k),$$

where $k = \deg_{x_n}(p)$ denotes the degree of x_n in p, and

$$r(0) := 1 \quad \text{and} \quad r(t) := \left(\frac{t-1}{t}\right)^{t-1}$$

for $t \geq 1$.

We first prove a theorem for univariate polynomials.

Theorem 9.1.15 (Gurvits)**.** *Let $f(z) = \sum_{k=0}^{n} a_k z^k$ be a real-rooted polynomial with $a_k \geq 0$. Then*

$$f'(0) \geq \left(\frac{n-1}{n}\right)^{n-1} \inf_{t>0} \frac{f(t)}{t}.$$

Proof. Since $f(z)$ is real-rooted and $a_i \geq 0$, the zeros must be non-positive. If there is root equal to 0, i. e., $a_0 = 0$ then

$$\inf_{t>0} \frac{f(t)}{t} = a_1 = f'(0).$$

So the claim is trivially true. Thus we can assume that the zeros are all negative. In this case,

$$f(z) = a_n(z + \rho_1) \ldots (z + \rho_n).$$

Then using $a_0 = a_n \prod_{i=1}^{n} \rho_i$ we have

$$f(t) = a_n \prod_{i=1}^{n} (t + \rho_i) = a_0 \prod_{i=1}^{n} \left(1 + \frac{t}{\rho_i}\right).$$

By the arithmetic-geometric mean inequality

$$f(t) = a_0 \prod_{i=1}^{n} \left(1 + \frac{t}{\rho_i}\right) \leq a_0 \left(1 + \frac{t}{n} \sum_{i=1}^{n} \frac{1}{\rho_i}\right)^n = a_0 \left(1 + \frac{t f'(0)}{n a_0}\right)^n.$$

Now let us choose

$$t_0 = \frac{n a_0}{(n-1) f'(0)}.$$

Then

$$f(t_0) \leq a_0 \left(1 + \frac{t_0 f'(0)}{n a_0}\right)^n = a_0 \left(1 + \frac{1}{n-1}\right)^n.$$

Then

$$\left(\frac{n-1}{n}\right)^{n-1} f(t_0) \leq a_0 \frac{n}{n-1} = a_0 \frac{t_0 f'(0)}{a_0} = t_0 f'(0).$$

Hence

$$\left(\frac{n-1}{n}\right)^{n-1} \frac{f(t_0)}{t_0} \leq f'(0).$$

Whence

$$f'(0) \geq \left(\frac{n-1}{n}\right)^{n-1} \frac{f(t_0)}{t_0} \geq \left(\frac{n-1}{n}\right)^{n-1} \inf_{t>0} \frac{f(t)}{t}.$$

□

We mention the following theorem of Hoeffding without proof.

Theorem 9.1.16 (Hoeffding). *Let $f(z) = \sum_{k=0}^{n} p_k z^k$ be a real-rooted polynomial with $p_k \geq 0$, and $f(1) = 1$, i. e., $\sum_{k=0}^{n} p_k = 1$. Let p be defined by the equation $\sum_{k=0}^{n} kp_k = np$. Suppose that for non-negative integers b and c we have $b \leq np \leq c$. Then*

$$\sum_{k=b}^{c} p_k \geq \sum_{k=b}^{c} \binom{n}{k} p^k (1-p)^{n-k}.$$

Let us show that Theorem 9.1.16 implies Theorem 9.1.15. In fact, it implies that for a real-rooted polynomial $f(z) = \sum_{i=0}^{n} a_i z^i$ with non-negative coefficients it gives that

$$a_k \geq \binom{n}{k} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} \inf_{t>0} \frac{f(t)}{t^k}.$$

For $k = 1$ this is exactly the content of Theorem 9.1.15. For a $t > 0$ let us consider the probability distribution $p_j = \frac{a_j t^j}{f(t)}$. Then $\sum_{j=0}^{n} p_j z^j$ is still real-rooted polynomial. Choose t_k in such a way that $\sum_{j=0}^{n} jp_j = k = np$, i.e. $p = \frac{k}{n}$. Next let us apply Theorem 9.1.16 with $b = c = k$. Then

$$\frac{a_k t_k^k}{f(t_k)} = p_k \geq \binom{n}{k} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k}.$$

In other words,

$$a_k \geq \binom{n}{k} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} \frac{f(t_k)}{t_k^k} \geq \binom{n}{k} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} \inf_{t>0} \frac{f(t)}{t^k}.$$

Proof of Theorem 9.1.14. The fact that $q(x_1, \dots, x_n)$ is stable follows from the fact that we first differentiate a stable polynomial, and then substitute 0 into it. According to part (d) and (f) of Theorem 9.1.9 both operations preserve stability.

Let us start to prove the second part. Let $\varepsilon > 0$ be arbitrary. By the definition of $\mathrm{cap}(q)$ we can choose $a_1, \dots, a_{n-1} > 0$ such that

$$\frac{p'(a_1, \dots, a_{n-1})}{\prod_{i=1}^{n-1} a_i} < \mathrm{cap}(q) + \varepsilon.$$

Next consider the polynomial

$$g(x) = p(a_1, \dots, a_{n-1}, x).$$

Clearly, $g'(0) = p'(a_1, \dots, a_{n-1})$. Since p is stable, and we substituted $a_i \in \mathbb{R}$ into it, $g(x)$ is stable. In other words, it is real-rooted. Hence we can use Theorem 9.1.15:

$$\begin{aligned}\frac{q(a_1, \dots, a_{n-1})}{\prod_{i=1}^{n-1} a_i} = \frac{g'(0)}{\prod_{i=1}^{n-1} a_i} &\geq \frac{1}{\prod_{i=1}^{n-1} a_i} \left(\frac{k-1}{k}\right)^{k-1} \inf_{x>0} \frac{g(x)}{x}. \\ &= \left(\frac{k-1}{k}\right)^{k-1} \inf \frac{p(a_1, \dots, a_{n-1}, x)}{\prod_{i=1}^{n-1} a_i \cdot x} \\ &\geq \left(\frac{k-1}{k}\right)^{k-1} \operatorname{cap}(p).\end{aligned}$$

Whence $\operatorname{cap}(q) + \varepsilon \geq \operatorname{cap}(p) r(k)$. Since $\varepsilon > 0$ was arbitrary we get that

$$\operatorname{cap}(q) \geq \operatorname{cap}(p) r(k).$$

□

9.1.4 Schrijver's and Gurvits's theorem

In this section we prove Schrijver's theorem, Theorem 9.1.17, on the number of perfect matchings of regular graphs together with van der Waerden's conjecture, Theorem 9.1.18 that was originally proved by Falikman and Egorychev. We derive these statements from a theorem of Gurvits, Theorem 9.1.19.

Theorem 9.1.17 (Schrijver [31])**.** *Let* $\operatorname{pm}(G)$ *denote the number of perfect matchings. Then for a d–regular bipartite graph G we have*

$$\operatorname{pm}(G)^{1/v(G)} \geq \left(\frac{(d-1)^{d-1}}{d^{d-2}}\right)^{1/2}.$$

Theorem 9.1.18 (Van der Waerden's conjecture, Falikman and Egorychev's theorem [13, 12])**.** *Let $A = (a_{ij})$ be a doubly stochastic $n \times n$ matrix. Then*

$$\operatorname{per}(A) \geq \frac{n!}{n^n}.$$

Equality holds if and only if $a_{ij} = 1/n$ for all i and j.

Theorem 9.1.19 (Gurvits [17])**.** *Let $A = (a_{ij})$ be a doubly stochastic $n \times n$ matrix. Then*

$$\operatorname{per}(A) \geq \prod_{i=1}^{n} r(\min(i, \lambda_A(i))),$$

where $\lambda_A(i)$ is the number of nonzeros in the i-th column of A, and where

$$r(0) := 1 \quad \textit{and} \quad r(k) := \left(\frac{k-1}{k}\right)^{k-1}$$

for $k \geq 1$.

Proof of Theorem 9.1.18. Let A be a doubly stochastic matrix. Then

$$\operatorname{per}(A) \geq r(1)r(2)r(3)\dots r(n) =$$

$$= 1 \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{2}{3}\right)^2 \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{4}{5}\right)^4 \cdot \dots \cdot \left(\frac{n-1}{n}\right)^{n-1} = \frac{n!}{n^n}.$$

This proves van der Waerden's conjecture. □

Proof of Theorem 9.1.17. Let G be a d-regular bipartite graph on $2n$ vertices. Let A be its $0-1$ biadjacency matrix. Then

$$\operatorname{per}(A) = d^n \operatorname{Per}\left(\frac{1}{d}A\right) \geq d^n r(d)^n = \left(\frac{(d-1)^{d-1}}{d^{d-2}}\right)^n.$$

This proves Schrijver's theorem. □

From now on let us start to prove Theorem 9.1.19.

Definition 9.1.20. For a non-negative matrix A let

$$P_A(x_1, \dots, x_n) = \prod_{i=1}^{n} (a_{i,1}x_1 + a_{i,2}x_2 + \dots + a_{i,n}x_n).$$

Lemma 9.1.21. *(i) If A is non-negative matrix (without all-zero row) then* $P_A(x_1, \dots, x_n)$ *is stable.*
(ii) If A is a doubly stochastic matrix then $\operatorname{cap}(P_A) = 1$.

Proof. The fist part is trivial since if $\operatorname{Im}(z_i) > 0$ for all $1 \leq i \leq n$ then $\operatorname{Im}\left(\sum_i a_{j,i}z_i\right) > 0$, in particular it is not 0.

To prove the second claim we will use the following well-known inequality called the weighted arithmetic-geometric mean inequality: let $\lambda_1, \lambda_2, \dots, \lambda_n, x_1, \dots, x_n \in \mathbb{R}_+$ with $\sum_{i=1}^n \lambda_i = 1$ then

$$\sum_{i=1}^{n} \lambda_i x_i \geq \prod_{i=1}^{n} x_i^{\lambda_i}.$$

If A is a doubly stochastic matrix then

$$P_A(x_1, \dots, x_n) = \prod_{i=1}^{n} (a_{i,1}x_1 + a_{i,2}x_2 + \dots + a_{i,n}x_n) \geq \prod_{i=1}^{n} x_1^{a_{i,1}} x_2^{a_{i,2}} \dots x_n^{a_{i,n}} = x_1x_2\dots x_n.$$

On the other hand, $P_A(1, 1, \dots, 1) = 1$. This means that

$$\operatorname{cap}(P_A) = \inf \frac{P_A(x_1, x_2, \dots, x_n)}{\prod_{i=1}^n x_i} = 1.$$

□

As a next step we define a series of multivariate polynomials that connects P_A with per(A).

$$Q_n(x_1,\dots,x_n) = P_A(x_1,\dots,x_n),$$
$$Q_{n-1}(x_1,\dots,x_{n-1}) = \frac{\partial}{\partial x_n}P_A(x_1,\dots,x_n)|_{x_n=0},$$
$$Q_{n-1}(x_1,\dots,x_{n-2}) = \frac{\partial^2}{\partial x_{n-1}\partial x_n}P_A(x_1,\dots,x_n)|_{x_{n-1}=x_n=0}$$
$$\vdots$$
$$Q_i(x_1,\dots x_i) = \frac{\partial^{n-i}}{\partial x_{i+1}\dots\partial x_n}P_A(x_1,\dots,x_n)|_{x_{i+1}=\dots=x_{n-1}=x_n=0}$$
$$\vdots$$
$$Q_0 = \frac{\partial^n}{\partial x_1\dots\partial x_n}P_A(x_1,\dots,x_n)|_{x_1=\dots=x_{n-1}=x_n=0}.$$

Alternatively, this sequence can be defined as follows.

$$Q_i(x_1,\dots x_i) = \frac{\partial}{\partial x_{i+1}}Q_{i+1}(x_1,x_2,\dots,x_{i+1})|_{x_{i+1}=0},$$

Assume that for some $R_i = R_i(x_1,\dots,x_i)$ we have

$$Q_{i+1}(x_1,x_2,\dots,x_{i+1}) = R_0 + R_1x_{i+1} + R_2x_{i+1}^2 + \dots + R_kx_{i+1}^t.$$

Then

$$Q_i(x_1,\dots x_i) = R_1(x_1,x_2,\dots,x_i).$$

In other words, Q_i is the "coefficient" of $x_{i+1}x_{i+2}\dots x_n$ in $Q_n = P_A$. In particular, Q_0 is the coefficient of $x_1x_2\dots x_n$ in $Q_n = P_A$. Of course, in

$$P_A(x_1,\dots,x_n) = \prod_{i=1}^{n}(a_{i,1}x_1 + a_{i,2}x_2 + \dots + a_{i,n}x_n).$$

the coefficient of $x_1x_2\dots x_n$ is per(A).

From this we can prove Theorem 9.1.19.

Proof of Theorem 9.1.19. We have seen that $Q_n = P_A$ is stable. Then the previous theorem asserts that $Q_n, Q_{n-1},\dots,Q_0$ are all stable polynomial. We have $\mathrm{cap}(Q_n) = 1$ and $\mathrm{cap}(Q_0) = \mathrm{per}(A)$. By theorem 9.1.14 we know that

$$\mathrm{cap}(Q_i) \geq \mathrm{cap}(Q_{i+1})r(t_i),$$

where $t_i = \deg_{x_{i+1}}(Q_{i+1})$ denotes the degree of x_{i+1} in Q_i. Is is easy to see that $t_i \leq i+1$ and $t_i \leq \lambda_A(i+1)$, the number of non-zeros in the $i+1$-th column of A. Hence

$$\mathrm{per}(A) = \mathrm{cap}(Q_0) \geq \prod_{i=1}^{n} r(\min(i,\lambda_A(i))).$$

□

We end this section with a classical inequality on permanents.

Theorem 9.1.22. *Let A be $n \times n$ non-negative matrix with column vectors $\underline{a}_1, \ldots, \underline{a}_n$. Then*

$$\operatorname{per}(\underline{a}_1, \underline{a}_2, \underline{a}_3 \ldots, \underline{a}_n)^2 \geq \operatorname{per}(\underline{a}_1, \underline{a}_1, \underline{a}_3 \ldots, \underline{a}_n)\operatorname{per}(\underline{a}_2, \underline{a}_2, \underline{a}_3 \ldots, \underline{a}_n).$$

Proof. Consider the polynomial

$$Q_2(x_1, x_2) = \frac{\partial^{n-2}}{\partial x_{i+1} \ldots \partial x_n} P_A(x_1, \ldots, x_n)|_{x_3 = \cdots = x_{n-1} = x_n = 0}.$$

This can be written in the form $Ax_1^2 + Bx_1x_2 + Cx_2^2$, where

$$A = \frac{1}{2}\operatorname{per}(\underline{a}_1, \underline{a}_1, \underline{a}_3 \ldots, \underline{a}_n), \quad B = \operatorname{per}(\underline{a}_1, \underline{a}_2, \underline{a}_3 \ldots, \underline{a}_n), \quad C = \frac{1}{2}\operatorname{per}(\underline{a}_2, \underline{a}_2, \underline{a}_3 \ldots, \underline{a}_n).$$

Note that $Q_2(x_1, x_2)$ is stable, and so $B^2 \geq 4AC$. Hence

$$\operatorname{per}(\underline{a}_1, \underline{a}_2, \underline{a}_3 \ldots, \underline{a}_n)^2 \geq \operatorname{per}(\underline{a}_1, \underline{a}_1, \underline{a}_3 \ldots, \underline{a}_n)\operatorname{per}(\underline{a}_2, \underline{a}_2, \underline{a}_3 \ldots, \underline{a}_n).$$

□

10. Entropy

10.1 Information and counting

The entropy of a probability distribution of $\underline{p} = (p_1, \dots, p_n)$ is

$$H(\underline{p}) = \sum_{i=1}^{n} p_i \ln \frac{1}{p_i}.$$

The intuition behind entropy that it encodes certain information contained in the probability distribution. This intuition can be formalized by various inequalities, see Proposition 10.2.1. For instance,

$$H(\underline{p}) \leq \ln n,$$

and equality holds true if and only if $\underline{p}$ is the uniform distribution, i. e., $\underline{p} = (\frac{1}{n}, \dots, \frac{1}{n})$. Based on this inequality one can prove lower bounds in various counting problem. Suppose that we would like to give a lower bound to the cardinality of some set S. If we can give a probability distribution $\underline{p}$ on S and compute $H(\underline{p})$, then we know that $|S| \geq \exp(H(\underline{p}))$. Another idea provides upper bound on $|S|$. Here we start from the uniform distribution on the set S, and uses entropy inequalities such as Shearer's inequality (Theorem 10.2.2) to give an upper bound on the entropy of this uniform distribution that is $\ln |S|$. Such strategy will be carried out in the case of matchings, see Brégman's theorem (Theorem 10.3.1), and in the case of homomorphisms, see the theorem of Galvin and Tetali (Theorem 10.4.1).

10.2 Basic properties of entropy

In this section we give a brief account into the theory of entropy. For a thorough treatment, see for example [8].

Let X be a discrete random variable taking its values in a finite set. The range of X will be denoted by $R(X)$. The entropy of X is defined as

$$H(X) = \sum_{x \in R(X)} \mathbb{P}(X = x) \ln \frac{1}{\mathbb{P}(X = x)}.$$

In case of an event Q we write

$$H(X|Q) = \sum_{x \in R(X)} \mathbb{P}(X = x|Q) \ln \frac{1}{\mathbb{P}(X = x|Q)}.$$

If X and Y discrete random variables then the conditional entropy is defined as

$$\begin{aligned} H(X|Y) &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) H(X|\{Y = y\}) \\ &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) \sum_{x \in R(X)} \mathbb{P}(X = x|Y = y) \ln \frac{1}{\mathbb{P}(X = x|Y = y)}. \end{aligned}$$

Next we collect some basic facts about entropy

Proposition 10.2.1. *We have*
(a) $0 \leq H(X) \leq \ln |R(X)|$. *Furthermore, if* X *has the uniform distribution on* $R(X)$ *then* $H(X) = \ln |R(X)|$.
(b) $H(X|Y) = H(X,Y) - H(Y)$.
(c) $H(X,Y) \leq H(X) + H(Y)$.
(d) $H(X,Z|Y) \leq H(X|Y) + H(Z|Y)$.
(e) $H(X) \leq H(X,Y)$.
(f) $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$.
(g) $H(X|Y,Z) \leq H(X|Y)$.
(h) $H(X|Y) \leq H(X|f(Y))$.
(j) $H(f(X)|X) = 0$ *, alternatively,* $H(f(X),X) = H(X)$.

Proof. (a) The lower bound is clear from the definition since each term is non-negative. The upper bound follows from Jensen's inequality since $f(x) = \ln(x)$ is a concave function (indeed, $f''(x) = \frac{-1}{x^2} < 0$). So let $p_i = \mathbb{P}(X = x)$ and $a_i = \frac{1}{\mathbb{P}(X=x)}$ then

$$H(X) = \sum_{x \in R(X)} \mathbb{P}(X = x) \ln \frac{1}{\mathbb{P}(X = x)} \leq \ln \left(\sum_{x \in R(X)} \mathbb{P}(X = x) \cdot \frac{1}{\mathbb{P}(X = x)} \right) = \ln |R(X)|.$$

Clearly, we have equality if and only if X has the uniform distribution on $R(X)$, and then $H(X) = \ln |R(X)|$.
(b)

$$\begin{aligned} H(X|Y) &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) H(X|\{Y = y\}) \\ &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) \sum_{x \in R(X)} \mathbb{P}(X = x|Y = y) \ln \frac{1}{\mathbb{P}(X = x|Y = y)} \\ &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) \sum_{x \in R(X)} \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \ln \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{x\in R(X), y\in R(Y)} \mathbb{P}(X=x, Y=y) \ln \frac{\mathbb{P}(Y=y)}{\mathbb{P}(X=x,Y=y)} \\
&= H(X,Y) - H(Y).
\end{aligned}
$$

(c) Note that

$$
H(X) + H(Y) - H(X,Y) = \sum_{x\in R(X), y\in R(Y)} \mathbb{P}(X=x, Y=y) \ln \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)}.
$$

Now let us apply Jensen's inequality to the function $f(x) = x \ln x$ with

$$
p_i = \mathbb{P}(X=x)\mathbb{P}(Y=y) \quad \text{and} \quad a_i = \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)}.
$$

Here $f(x)$ is convex as $f''(x) = \frac{1}{x} > 0$, and naturally the sum of p_i's is 1. Let us introduce the notation $I(X,Y) = H(X) + H(Y) - H(X,Y)$. This is called the mutual information.

$$
\begin{aligned}
I(X,Y) &= H(X) + H(Y) - H(X,Y) \\
&= \sum_{x\in R(X), y\in R(Y)} \mathbb{P}(X=x, Y=y) \ln \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)} \\
&= \sum_{x\in R(X), y\in R(Y)} \mathbb{P}(X=x)\mathbb{P}(Y=y) \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)} \ln \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)} \\
&= \sum_i p_i f(a_i) \geq f\left(\sum_i p_i a_i\right) \\
&= f\left(\sum_{x\in R(X), y\in R(Y)} \mathbb{P}(X=x)\mathbb{P}(Y=y) \cdot \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)}\right) \\
&= f(1) = 0
\end{aligned}
$$

(d) By the previous part we have

$$
H(X|Y=y) + H(Z|Y=y) \geq H(X,Z|Y=y)
$$

for all $y \in R(Y)$. Hence

$$
\begin{aligned}
H(X|Y) + H(Z|Y) &= \sum_{y\in R(Y)} \mathbb{P}(Y=y)(H(X|Y=y) + H(Z|Y=y)) \\
&\geq \sum_{y\in R(Y)} \mathbb{P}(Y=y) H(X,Z|Y=y) \\
&= H(X,Z|Y)
\end{aligned}
$$

(e) We have

$$
H(X,Y) - H(X) = H(Y|X) = \sum_{x\in X} \mathbb{P}(X=x) H(Y|X=x) \geq 0
$$

termwise.
(f) This is a direct consequence of $H(X|Y)+H(Z|Y) \geq H(X,Z|Y)$ using that $H(X|Y) = H(X,Y) - H(Y)$, $H(Z|Y) = H(Z,Y) - H(Y)$, and $H(X,Z|Y) = H(X,Z,Y) - H(Y)$.
(g) This is again a direct consequence of $H(X|Y)+H(Z|Y) \geq H(X,Z|Y)$ or equivalently $H(X,Y,Z)+H(Y) \leq H(X,Y)+H(Y,Z)$ using that $H(X|Y,Z) = H(X,Y,Z)-H(Y,Z)$ and $H(X|Y) = H(X,Y) - H(Y)$.
(h) We have seen that

$$H(X|Y) = \sum_{x\in R(X),y\in R(Y)} \mathbb{P}(X = x, Y = y) \ln \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)}.$$

Similarly,

$$H(X|f(Y)) = \sum_{x\in R(X),z\in R(f(Y))} \mathbb{P}(X = x, f(Y) = z) \ln \frac{\mathbb{P}(f(Y) = z)}{\mathbb{P}(X = x, f(Y) = z)}.$$

Now fix an $x \in R(X)$ and a $z \in R(f(Y))$, and observe that

$$\begin{aligned}
T_{x,z} &:= \sum_{y:f(y)=z} \mathbb{P}(X = x, Y = y) \ln \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)} \\
&= \mathbb{P}(X = x, f(Y) = z) \sum_{y:f(y)=z} \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x, f(Y) = z)} \ln \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)} \\
&\leq \mathbb{P}(X = x, f(Y) = z) \ln \left(\sum_{y:f(y)=z} \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, f(Y) = z)} \right) \\
&\leq \mathbb{P}(X = x, f(Y) = z) \ln \frac{\mathbb{P}(f(Y) = z)}{\mathbb{P}(X = x, f(Y) = z)}.
\end{aligned}$$

In the second step we used Jensen's inequality to the function $f(x) = \ln x$ with

$$p_i = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x, f(Y) = z)} \quad \text{and} \quad a_i = \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)}.$$

Now by summing these inequalities for all $x \in R(X)$ and $z \in R(f(Y))$ we get that $H(X|Y) \leq H(X|f(Y))$.
(j) $H(f(X)|X) = \sum_{x\in R(X)} \mathbb{P}(X = x)H(f(X)|\{X = x\}) = 0$ since the inner sum $H(f(X)|\{X = x\}) = 0$ for each $x \in R(X)$. By part (b) we have $0 = H(f(X)|X) = H(f(X), X) - H(X)$. □

For a random vector $X = (X_1, \ldots, X_m)$ and an $A \subseteq [m]$ let $X_A = (X_i \mid i \in A)$.

Theorem 10.2.2 (Shearer [7])**.** *Let $X = (X_1, \ldots, X_m)$ be a random vector, and $\mathcal{A}$ be a collection of subsets of $[m]$ possibly with repeats such that each element of $[m]$ is contained in at least t members of $\mathcal{A}$. Then*

$$H(X) \leq \frac{1}{t} \sum_{A\in\mathcal{A}} H(X_A).$$

Proof. By part (f) of Proposition 10.2.1 we know that

$$H(X_A) + H(X_B) \geq H(X_{A\cap B}) + H(X_{A\cup B}).$$

So if we see two sets $A, B \in \mathcal{A}$ such that neither $A \subseteq B$ nor $B \subseteq A$ then we can replace them by $A \cap B$ and $A \cup B$. We can do this step even if A and B are disjoint. This way we cannot increase $\sum H(X_A)$ and every element will be contained in exactly the same number of sets. Moreover we cannot do this step infinitely many times because

$$|A|^2 + |B|^2 \leq |A \cap B|^2 + |A \cup B|^2$$

with strict inequality if $A \setminus B$ and $B \setminus A$ are non-empty. So in each step the sum $\sum |A|^2$ will increase by at least 1 and it is at most $|\mathcal{A}|m^2$. This means that the process will halt. For the final set system $\mathcal{B}$ it will be true that for any $A, B \in \mathcal{B}$ we have $A \subseteq B$ or $B \subseteq A$. This means that we get a set system $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n$ where $n = |\mathcal{A}| = |\mathcal{B}|$. Since every element is still contained in at least t sets we get that that the last t sets must be the whole set $[m]$. Hence

$$H(X) \leq \frac{1}{t}\sum_{A\in\mathcal{B}} H(X_A) \leq \frac{1}{t}\sum_{A\in\mathcal{A}} H(X_A).$$

□

10.3 Matchings: Brégman's theorem

Theorem 10.3.1 (Brégman [5])**.** *Let $G = (A, B, E)$ be a bipartite graph with $|A| = |B| = n$. Assume that the degrees of the vertices of A are $d_1, \dots, d_n$. Let $\mathrm{pm}(G)$ denote the number of perfect matchings of G. Then*

$$\mathrm{pm}(G) \leq \prod_{i=1}^{n} (d_i!)^{1/d_i}.$$

The following theorem on regular bipartite graphs is an immediate corollary of Brégman's theorem. We will see later that the condition on bipartiteness can be dropped.

Theorem 10.3.2. *Let $\mathrm{pm}(G)$ denote the number of perfect matchings. Then for a d–regular bipartite graph G we have*

$$\mathrm{pm}(G)^{1/v(G)} \leq \mathrm{pm}(K_{d,d})^{1/v(K_{d,d})}$$

Proof of Theorem 10.3.1. We will consider a perfect matching as an $f : [n] \to [n]$, $f(i) = j$ if (a_i, b_j) is en edge of the perfect matching. Let X be the random vector $(f(1), \dots, f(n))$, where we choose a perfect matching f uniformly among all perfect matchings. Then the entropy of X is $H(X) = \ln \mathrm{pm}(G)$. Next we will give an upper bound on $H(X)$. In general, we have

$$H(X) = H(X_1) + \sum_{i=2}^{n} H(X_i | X_{i-1}, \dots, X_1).$$

We can think of this process as follows: we reveal one by one the neighbors of the vertices in A in the random matching f. When we arrive to some fixed vertex $a \in A$ it might occur that some of its neighbors are already covered by the perfect matching so we can be sure that the conditional entropy is definitely not $\ln d_a$, but something smaller. Unfortunately, it is not clear how much smaller it is since we have no control on how many neighbors of a are already occupied. We overcome this difficulty with a little trick: choose a random permutation $\pi \in S_n$ and apply the chain rule for this random order.

$$H(X) = H(X_{\pi(1)}) + \sum_{i=2}^{n} H(X_{\pi(i)} | X_{\pi(i-1)}, \dots, X_{\pi(1)}).$$

It will be more convenient to rewrite it as follows:

$$H(X) = \sum_{v \in A} H(X_v | X_{\{v' : \pi(v') < \pi(v)\}}).$$

Let us average it over all $n!$ permutations of S_n:

$$H(X) = \sum_{v \in A} \frac{1}{n!} \sum_{\pi \in S_n} H(X_v | X_{\{v' : \pi(v') < \pi(v)\}}).$$

For a fixed vertex $v \in A$ let us study the quantity

$$\frac{1}{n!} \sum_{\pi \in S_n} H(X_v | X_{\{v' : \pi(v') < \pi(v)\}}).$$

For a moment let us stop to examine a general conditional entropy:

$$\begin{aligned} H(X|Y) &= \sum_{y \in R(Y)} \mathbb{P}(y) \sum_{x \in R(X)} \mathbb{P}(X = x | Y = y) \ln \frac{1}{\mathbb{P}(X = x | Y = y)} \\ &\leq \sum_{y \in R(Y)} \mathbb{P}(y) \ln |R(X | Y = y)|. \end{aligned}$$

The point is that the range of X conditioned on $Y = y$ might be smaller than the range of X. In particular, this happens if some neighbor of the vertex v is already occupied. So let $N_v(\pi, f)$ be the number of choices remaining for v if we already know $f(v')$ for all v' for which $\pi(v') < \pi(v)$. Then

$$\frac{1}{n!} \sum_{\pi \in S_n} H(X_v | X_{\{v' : \pi(v') < \pi(v)\}}) \leq \sum_{j=1}^{d_v} \mathbb{P}(N_v(\pi, f) = j) \ln j$$

$$= \sum_{j=1}^{d_v} \ln j \frac{|\{(\pi, f) \mid N_v(\pi, f) = j\}|}{n!\mathrm{pm}(G)}.$$

Now the crucial observation is that

$$\frac{|\{(\pi, f) \mid N_v(\pi, f) = j\}|}{n!\mathrm{pm}(G)} = \frac{1}{d_v}$$

independently of j. In fact, it is independent of f: once we have fixed f the chance that $N_v(\pi, f) = j$ is $\frac{1}{d_v}$. The reason is simple: let us consider the d_v vertices in A whose f-neighbors are exactly the neighors of v. If we keep only the ordering of these vertices from π then with probability $\frac{1}{d_v}$ the vertex v will be the first, with probability $\frac{1}{d_v}$ the vertex v will be the second, etc. Hence

$$\frac{1}{n!} \sum_{\pi \in S_n} H(X_v | X_{\{v' : \pi(v') < \pi(v)\}}) \leq \sum_{j=1}^{d_v} \frac{\ln j}{d_v} = \frac{\ln(d_v!)}{d_v}.$$

Hence

$$H(X) \leq \sum_{v \in V} \frac{\ln(d_v!)}{d_v}.$$

Since $H(X) = \ln \mathrm{pm}(G)$ we get that

$$\mathrm{pm}(G) \leq \prod_{v \in A} (d_v!)^{1/d_v}.$$

□

10.4 Homomorphisms

Theorem 10.4.1 (Galvin and Tetali [15]). *Let G be a d–regular bipartite graph, and H be a fixed graph. Then*

$$\hom(G, H)^{1/v(G)} \leq \hom(K_{d,d}, H)^{1/v(K_{d,d})}.$$

Equivalently,

$$t(G, H)^{1/v(G)} \leq t(K_{d,d}, H)^{1/v(K_{d,d})},$$

or

$$Z(G, A)^{1/v(G)} \leq Z(K_{d,d}, A)^{1/v(K_{d,d})}.$$

Proof. Let $G = (A, B, E)$ and $K_{d,d} = (A_d, B_d, E_d)$. First, we suppose that H is a bipartite graph, and $V(H) = U \cup L$ is the partition. (Later we remove this condition on H.) Let

$$\mathrm{Hom}^{L,U}(G, H) = \{f \in \mathrm{Hom}(G, H) : f(A) \subseteq L, f(B) \subseteq U\}.$$

First we consider $|\mathrm{Hom}^{L,U}(K_{d,d},H)|$. For any set $S\subseteq L$ let

$$\mathcal{H}(S)=\{f\in \mathrm{Hom}^{L,U}(K_{d,d},H)\mid f(A_d)=S\},$$

$$T(S)=\{g:[d]\to S: g \text{ surjective}\},$$

and

$$C^U(S)=\{j\in U:(j,i)\in E(H)\ \forall i\in S\}.$$

Then

$$|\mathrm{Hom}^{L,U}(K_{d,d},H)|=\sum_{S\subseteq L}|T(S)||C^U(S)|^d.$$

Next we show that

$$|\mathrm{Hom}^{L,U}(G,H)|\le |\mathrm{Hom}^{L,U}(K_{d,d},H)|^{v(G)/(2d)}.$$

Let f be chosen uniformly at random from $\mathrm{Hom}^{L,U}(G,H)$. We think of f as a vector $(f(v))_{v\in V}$, and f_S denotes the random vector $(f(v))_{v\in S}$. Let $M_v=\{f(w)\ |w\in N(v)\}$. Note that M_v is a set, while $f_{N(v)}$ is a vector. Clearly, $f_{N(v)}$ carries more information than M_v, or in other words, M_v is a function of $f_{N(v)}$. For $v\in B$ and $S\subseteq L$ let $m_v(S)$ denote the probability $\mathbb{P}(M_v=S)$. Clearly, $\sum_S m_v(S)=1$. Then

$$\begin{aligned}
\ln|\mathrm{Hom}^{L,U}(G,H)| &\overset{(a)}{=} H(f_V)\\
&\overset{(b)}{=} H(f_A)+H(f_B|f_A)\\
&\overset{(d)}{\le} H(f_A)+\sum_{v\in B}H(f(v)|f_A)\\
&\overset{(g)}{\le} H(f_A)+\sum_{v\in B}H(f(v)|f_{N(v)})\\
&\overset{(S)}{\le} \frac{1}{d}\sum_{v\in B}H(f_{N(v)})+\sum_{v\in B}H(f(v)|f_{N(v)})\\
&\overset{(j)}{=} \frac{1}{d}\sum_{v\in B}H(f_{N(v)},M_v)+\sum_{v\in B}H(f(v)|f_{N(v)})\\
&\overset{(b)}{=} \frac{1}{d}\sum_{v\in B}(H(M_v)+H(f_{N(v)}|M_v))+\sum_{v\in B}H(f(v)|f_{N(v)})\\
&= \frac{1}{d}\sum_{v\in B}(H(M_v)+H(f_{N(v)}|M_v)+dH(f(v)|f_{N(v)}))\\
&\overset{(h)}{\le} \frac{1}{d}\sum_{v\in B}(H(M_v)+H(f_{N(v)}|M_v)+dH(f(v)|M_v))\\
&\overset{\text{def}}{=} \frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L}\left(m_v(S)\ln\frac{1}{m_v(S)}+m_v(S)H(f_{N(v)}|\{M_v=S\})\right)+\\
&+\frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L}(dm_v(S)H(f(v)|\{M_v=S\}))
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L}\left(m_v(S)\ln\frac{1}{m_v(S)} + m_v(S)\ln|T(S)| + dm_v(S)\ln|C^U(S)|\right)\\
&= \frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L} m_v(S)\ln\frac{|T(S)||C^U(S)|^d}{m_v(S)}\\
&\overset{(J)}{\leq} \frac{1}{d}\sum_{v\in B}\ln\left(\sum_{S\subseteq L}|T(S)||C^U(S)|^d\right)\\
&= \frac{v(G)}{2d}\ln|\mathrm{Hom}^{L,U}(K_{d,d},H)|.
\end{aligned}
$$

On the top of the signs $=$ or $\leq$ one can see which part of Proposition 10.2.1 we have used. The sign S refers to Shearer's inequality, Theorem 10.2.2. The sign J refers to Jensen's inequality applied to $\ln x$. The def simply means that we use the definition of the (conditional) entropy. Finally, we use our previously found formula for $|\mathrm{Hom}^{L,U}(K_{d,d},H)|$.

Now to finish the proof of the theorem we remove the condition that H is bipartite. Let $H' = H \times K_2$, so H' is a bipartite graph with vertex set $V(H') = V(H) \times \{0,1\}$ and $((v,0),(w,1)) \in E(H')$ if $(v,w) \in E(H)$. Let $U = \{(v,0) \mid v \in V(H)\}$ and $L = \{(v,1) \mid v \in V(H)\}$. Then

$$|\mathrm{Hom}^{L,U}(G,H')| = |\mathrm{Hom}(G,H)|.$$

Then we are done. $\square$

Bibliography

[1] M. Abért and T. Hubai, *Benjamini–Schramm convergence and the distribution of chromatic roots for sparse graphs*, Combinatorica, **35** (2015), pp. 127–151.

[2] R. Ahlswede and D. E. Daykin, *An inequality for the weights of two families of sets, their unions and intersections*, Probability Theory and Related Fields, **43** (1978), pp. 183–185.

[3] N. Alon and S. Friedland, *The maximum number of perfect matchings in graphs with a given degree sequence*, Electron. J. Combin, **15** (2008).

[4] C. Borgs, J. Chayes, J. Kahn, and L. Lovász, *Left and right convergence of graphs with bounded degree*, Random Structures & Algorithms, **42** (2013), pp. 1–28.

[5] L. M. Brégman, *Some properties of nonnegative matrices and their permanents*, in Soviet Math. Dokl, vol. 14, 1973, pp. 945–949.

[6] F. Chung and S.-T. Yau, *Coverings, heat kernels and spanning trees*, Journal of Combinatorics, **6** (1998), pp. 163–184.

[7] F. R. K. Chung, P. Frankl, R. L. Graham, and J. B. Shearer, *Some intersection theorems for ordered sets and graphs*, Journal of Combinatorial Theory, Series A, **43** (1986), pp. 23–37.

[8] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[9] P. Csikvári, *Lower matching conjecture, and a new proof of Schrijver's and Gurvits's theorems*, Journal of European Mathematical Society, **19** (2017), pp. 1811–1844.

[10] P. Csikvári and Z. Lin, *Sidorenko's conjecture, colorings and independent sets*, The Electronic Journal of Combinatorics, **24** (2017), p. P 1.2.

[11] J. Cutler and J. A. Radcliffe, *Extremal graphs for homomorphisms*, Journal of Graph Theory, **67** (2011), pp. 261–284.

[12] G. P. Egorychev, *Proof of the van der waerden conjecture for permanents*, Siberian Mathematical Journal, **22** (1981), pp. 854–859.

[13] D. I. Falikman, *Proof of the van der waerden conjecture regarding the permanent of a doubly stochastic matrix*, Mathematical notes of the Academy of Sciences of the USSR, **29** (1981), pp. 475–479.

[14] H. Finner, *A generalization of Hölder's inequality and some probability inequalities*, The Annals of probability, **20** (1992), pp. 1893–1901.

[15] D. Galvin and P. Tetali, *On weighted graph homomorphisms*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **63** (2004), pp. 97–104.

[16] C. D. Godsil, *Algebraic combinatorics*, vol. 6, CRC Press, 1993.

[17] L. Gurvits, *Van der Waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all*, The Electronic Journal of Combinatorics, **15** (2008).

[18] H. Hatami, *Graph norms and Sidorenko's conjecture*, Israel Journal of Mathematics, **175** (2010), pp. 125–150.

[19] O. J. Heilmann and E. H. Lieb, *Theory of monomer-dimer systems*, Communications in Mathematical Physics, (1972), pp. 190–232.

[20] J. Kahn, *An entropy approach to the hard-core model on bipartite graphs*, Combinatorics, Probability and Computing, **10** (2001), pp. 219–237.

[21] J. H. Kim, C. Lee, and J. Lee, *Two approaches to Sidorenko's conjecture*, Transactions of the American Mathematical Society, **368** (2016), pp. 5057–5074.

[22] N. Linial, *Lifts of graphs (talk slides)*, `http://www.cs.huji.ac.il/~nati/PAPERS/lifts_talk.pdf`.

[23] E. Lubetzky and Y. Zhao, *On replica symmetry of large deviations in random graphs*, Random Structures & Algorithms, **47** (2015), pp. 109–146.

[24] B. D. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra and its Applications, **40** (1981), pp. 203–216.

[25] B. D. McKay, *Spanning trees in regular graphs*, European Journal of Combinatorics, **4** (1983), pp. 149–160.

[26] M. B. Nathanson, *The hermite-sylvester criterion for real-rooted polynomials*, arXiv preprint arXiv:1911.01745, (2019).

[27] R. C. Read, *An introduction to chromatic polynomials*, Journal of Combinatorial Theory, **4** (1968), pp. 52–71.

[28] G. Regts, *Zero-free regions of partition functions with applications to algorithms and graph limits*, arXiv preprint arXiv:1507.02089, (2015).

[29] N. Ruozzi, *The Bethe partition function of log-supermodular graphical models*, in Advances in Neural Information Processing Systems, 2012, pp. 117–125.

[30] A. Sah, M. Sawhney, D. Stoner, and Y. Zhao, *The number of independent sets in an irregular graph*, Journal of Combinatorial Theory, Series B, **138** (2019), pp. 172–195.

[31] A. Schrijver, *Counting 1-factors in regular bipartite graphs*, Journal of Combinatorial Theory, Series B, **72** (1998), pp. 122–135.

[32] A. Sidorenko, *A correlation inequality for bipartite graphs*, Graphs and Combinatorics, **9** (1993), pp. 201–204.

[33] A. D. Sokal, *Bounds on the complex zeros of (di)chromatic polynomials and Potts-model partition functions*, Combinatorics, Probability and Computing, **10** (2001), pp. 41–77.

[34] D. Straszak and N. K. Vishnoi, *Belief propagation, bethe approximation and polynomials*, arXiv preprint arXiv:1708.02581, (2017).

[35] P. O. Vontobel, *Counting in graph covers: A combinatorial characterization of the Bethe entropy function*, IEEE Transactions on Information Theory, **59** (2013), pp. 6018–6048.

[36] Y. Zhao, *The number of independent sets in a regular graph*, Combinatorics, Probability and Computing, **19** (2010), pp. 315–320.

[37] Y. Zhao, *Extremal regular graphs: independent sets and graph homomorphisms*, arXiv preprint arXiv:1610.09210, (2016).