



Random Cluster Model on Regular Graphs

Ferenc Bencs¹, Márton Borbényi^{2,3}, Péter Csikvári^{2,3} 

¹ Korteweg de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands. E-mail: ferenc.bencs@gmail.com

² Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest 1117, Hungary.
E-mail: marton.borbenyi@gmail.com

³ Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13-15, Budapest 1053, Hungary.
E-mail: peter.csikvari@gmail.com

Received: 24 May 2022 / Accepted: 14 October 2022
Published online: 14 November 2022 – © The Author(s) 2022

Abstract: For a graph $G = (V, E)$ with $v(G)$ vertices the partition function of the random cluster model is defined by

$$Z_G(q, w) = \sum_{A \subseteq E(G)} q^{k(A)} w^{|A|},$$

where $k(A)$ denotes the number of connected components of the graph (V, A) . Furthermore, let $g(G)$ denote the girth of the graph G , that is, the length of the shortest cycle. In this paper we show that if $(G_n)_n$ is a sequence of d -regular graphs such that the girth $g(G_n) \rightarrow \infty$, then the limit

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w) = \ln \Phi_{d,q,w}$$

exists if $q \geq 2$ and $w \geq 0$. The quantity $\Phi_{d,q,w}$ can be computed as follows. Let

$$\begin{aligned} \Phi_{d,q,w}(t) := & \left(\sqrt{1 + \frac{w}{q}} \cos(t) + \sqrt{\frac{(q-1)w}{q}} \sin(t) \right)^d \\ & + (q-1) \left(\sqrt{1 + \frac{w}{q}} \cos(t) - \sqrt{\frac{w}{q(q-1)}} \sin(t) \right)^d, \end{aligned}$$

then

$$\Phi_{d,q,w} := \max_{t \in [-\pi, \pi]} \Phi_{d,q,w}(t),$$

The first author is supported by the NKFIH (National Research, Development and Innovation Office, Hungary) grant KKP-133921. The second author is supported by the ÚNKP-21-2 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund. The third author is supported by the Counting in Sparse Graphs Lendület Research Group.

The same conclusion holds true for a sequence of random d -regular graphs with probability one. Our result extends the work of Dembo, Montanari, Sly and Sun for the Potts model (integer q), and we prove a conjecture of Helmuth, Jenssen and Perkins about the phase transition of the random cluster model with fixed q .

1. Introduction

For a graph $G = (V, E)$ the partition function of the random cluster model is defined by

$$Z_G(q, w) = \sum_{A \subseteq E(G)} q^{k(A)} w^{|A|},$$

where $k(A)$ denotes the number of connected components of the graph (V, A) . In many papers, one uses the parametrization $w = e^\beta - 1$.

When q is a positive integer, then $Z_G(q, w)$ is also the partition function of the Potts-model with q spins, moreover there is a natural coupling between the two models, see for example [16]. In this paper we call a model a spin model with r spins if there is an $r \times r$ symmetric matrix N and a vector $\underline{\mu} \in \mathbb{R}^r$ such that for a graph $G = (V, E)$ the probability of a $\sigma : V \rightarrow \{1, 2, \dots, r\}$ is

$$\mathbb{P}(\sigma) = \frac{1}{Z_G(N, \underline{\mu})} \prod_{v \in V} \mu_{\sigma(v)} \prod_{(u,v) \in E(G)} N_{\sigma(u), \sigma(v)},$$

where with the notation $[r] = \{1, 2, \dots, r\}$ we have

$$Z_G(N, \underline{\mu}) = \sum_{\sigma: V \rightarrow [r]} \prod_{v \in V} \mu_{\sigma(v)} \prod_{(u,v) \in E(G)} N_{\sigma(u), \sigma(v)}.$$

In both expressions the second product is over the edge set $E(G)$, the symmetricity of N ensures that the expression is well-defined. The quantity $Z_G(N, \underline{\mu})$ is the partition function of the model. In case of the Potts-model we have $r = q$ and $N = J_q + wI_q$, where J_q is the $q \times q$ matrix consisting of 1's and I_q is the $q \times q$ identity matrix. The vector $\underline{\mu}$ is the constant 1 vector in this case. In general, if $\underline{\mu}$ is the constant 1 vector, then we will simply write $Z_G(N)$ instead of $Z_G(N, \underline{\mu})$.

Let $v(G)$ denote the number of vertices of a graph G . In this paper we study the quantity

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w)$$

when $(G_n)_n$ is an essentially large girth sequence of d -regular graphs. A graph sequence $(G_n)_n$ is called essentially large girth if for all g we have $\lim_{n \rightarrow \infty} \frac{L(G_n, g)}{v(G_n)} = 0$, where $L(G, g)$ denotes the number of cycles of length at most $g - 1$. It is known that a sequence of random d -regular graphs is essentially large girth graph sequence with probability one (see for instance [19]). So the problems of determining $\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \mathbb{E} \ln Z_{G_n}(q, w)$ or $\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln \mathbb{E} Z_{G_n}(q, w)$ for random d -regular graph sequence $(G_n)_n$ are very strongly related to this question. In fact, it will turn out that all these limits are the same. The main theorem of this paper is the following.

Theorem 1.1. *If $(G_n)_n$ is an essentially large girth sequence of d -regular graphs, then the limit*

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w) = \ln \Phi_{d,q,w}$$

exists for $q \geq 2$ and $w \geq 0$. The quantity $\Phi_{d,q,w}$ can be computed as follows. Let

$$\begin{aligned} \Phi_{d,q,w}(t) := & \left(\sqrt{1 + \frac{w}{q}} \cos(t) + \sqrt{\frac{(q-1)w}{q}} \sin(t) \right)^d \\ & + (q-1) \left(\sqrt{1 + \frac{w}{q}} \cos(t) - \sqrt{\frac{w}{q(q-1)}} \sin(t) \right)^d, \end{aligned}$$

then

$$\Phi_{d,q,w} := \max_{t \in [-\pi, \pi]} \Phi_{d,q,w}(t).$$

The same conclusion holds true with probability one for a sequence of random d -regular graphs.

The quantity $\Phi_{d,q,w}$ has various alternative descriptions. As far as we know the description in the theorem is new even for the Potts model. There is a critical value $w_c(d, q)$ such that if $0 \leq w \leq w_c(d, q)$, then $\Phi_{d,q,w}(t)$ is maximized at $t = 0$, and so $\Phi_{d,q,w} = q \left(1 + \frac{w}{q}\right)^{d/2}$, and for $w > w_c(d, q)$ we have $\Phi_{d,q,w} > q \left(1 + \frac{w}{q}\right)^{d/2}$. Moreover, if $q > 2$, then $\frac{\partial}{\partial w} \Phi_{d,q,w}$ is discontinuous at $w_c(q)$, that is, there is a first order phase transition at $w_c(q)$. We will see that

$$w_c(d, q) = \frac{q-2}{(q-1)^{1-2/d} - 1} - 1.$$

For random d -regular graphs and the Potts model this was established by Galanis et al. [15]. For not necessarily integer $q \geq 2$ this was conjectured by Helmuth et al. [17]. We will also prove that for any d -regular graph G we have

$$Z_G(q, w) \geq \Phi_{d,q,w}^{v(G)},$$

and we also show that if G contains $\varepsilon v(G)$ cycles of length at most g for some fixed ε and g , then $Z_G(q, w)$ is exponentially larger than that bound. Related results were obtained by Ruozzi [22].

1.1. Related works. There has been a lot of work on Potts model and random cluster model both on random regular graphs and essentially large girth graph sequences. The papers mentioned below treat various related problems, but we only mention the results that are directly related to Theorem 1.1.

The case $q = 2$ and $w \geq 0$ is the so-called ferromagnetic Ising-model. In this case, Dembo and Montanari [12] proved that $Z_{G_n}(2, w)^{1/v(G_n)}$ converges if $(G_n)_n$ is an essentially large girth sequence of d -regular graphs. In fact, they proved a significantly more general theorem about essentially large girth graphs that are not necessarily regular. Note that they use the terminology locally tree-like for what we use essentially large girth.

When q is a positive integer and $w \geq 0$, that is, in the case of the ferromagnetic Potts model, Dembo et al. [14] proved the convergence of $Z_{G_n}(q, w)^{1/v(G_n)}$ for essentially large girth sequence of d -regular graphs for every d except when w belongs to a certain interval (w_0, w_1) . Later Dembo et al. [13] proved the convergence of $Z_{G_n}(q, w)^{1/v(G_n)}$ for essentially large girth sequence of d -regular graphs, when d is even, q is a positive integer and $w \geq 0$ even if $w \in (w_0, w_1)$. Very recently, Helmuth et al. [17] proved the convergence of $Z_{G_n}(q, w)^{1/v(G_n)}$ for essentially large girth sequence of d -regular graphs for large (not necessarily integer) q and $w \geq 0$ with the additional hypothesis that $(G_n)_n$ satisfies some expansion condition. This line of research using cluster expansion and an expansion property of $(G_n)_n$ was extended by Carlson et al. [7] and by Carlson et al. [8] for the Potts model. Ferromagnetic Potts models on random regular graphs were also studied by Galanis et al. [15].

Our theorem does not cover the case when $w < 0$. When q is a positive integer and $w = -1$, then $Z_G(q, -1)$ counts the number of proper colorings of the graph G . This case was treated by Bandyopadhyay and Gamarnik [2]. They showed that if $q \geq d + 1$, then for an essentially large girth graph sequence of d -regular graphs $(G_n)_n$ we have

$$\lim_{n \rightarrow \infty} Z_{G_n}(q, -1)^{1/v(G_n)} = q \left(1 - \frac{1}{q}\right)^{d/2}.$$

Their result was extended for integer $q \geq 2\Delta$ and $w \geq -1$ by Borgs et al. [6] for general Benjamini–Schramm convergent graph sequences, where Δ is a bound on the degrees of all G_n in the sequence. The result of Bandyopadhyay and Gamarnik [2] was extended for not necessarily integer $q \geq 8\Delta$ and $w = -1$ by Abért and Hubai [1] also for arbitrary Benjamini–Schramm convergent graph sequences. Csikvári and Frenkel [11] showed that the same conclusion holds true for every fixed $w \geq 0$ and q sufficiently large in terms of w and Δ . The partition function of the random cluster model is strongly related to the so-called Tutte polynomial [25]. For a graph $G = (V, E)$ the Tutte polynomial $T_G(x, y)$ is defined by

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

The connection between the Tutte polynomial and the random cluster model is

$$T_G(x, y) = (x - 1)^{-k(E)} (y - 1)^{-v(G)} Z_G((x - 1)(y - 1), y - 1).$$

Bencs and Csikvári [4] proved that for an essentially large girth sequence of d -regular graphs $(G_n)_n$ we have

$$\lim_{n \rightarrow \infty} T_{G_n}(x, y)^{1/v(G_n)} = \begin{cases} (d - 1) \left(\frac{(d-1)^2}{(d-1)^2 - x}\right)^{d/2-1} & \text{if } x \leq d - 1, \\ x \left(1 + \frac{1}{x-1}\right)^{d/2-1} & \text{if } x > d - 1 \end{cases}$$

for $x \geq 1$ and $0 \leq y \leq 1$. The next theorem summarizes the known results for the Tutte polynomial for non-negative x, y .

Theorem 1.2. *Let $(G_n)_n$ be an essentially large girth sequence of d -regular graphs. Then the limit*

$$\lim_{n \rightarrow \infty} T_{G_n}(x, y)^{1/v(G_n)} = t_d(x, y)$$

exists if x, y satisfy one the following conditions

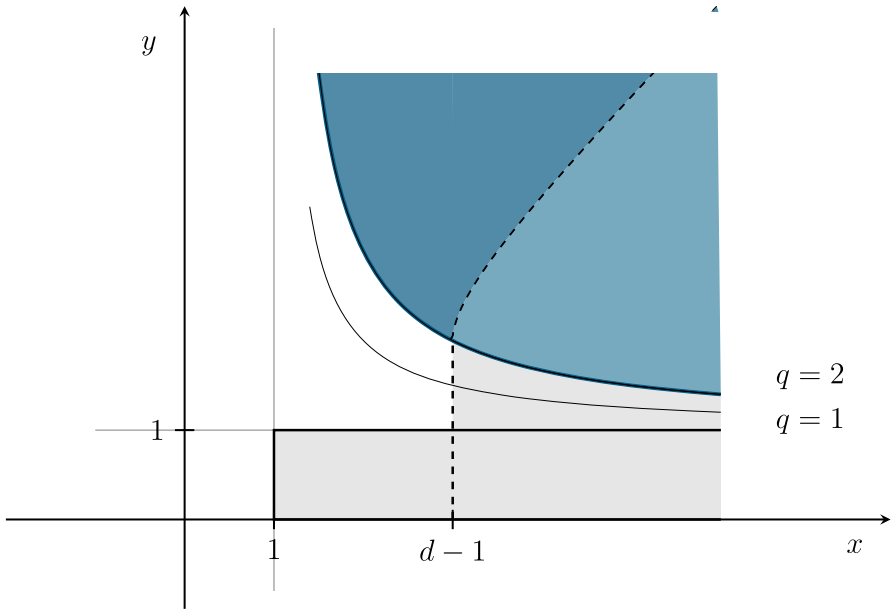


Fig. 1. The investigated parameters of the article (in blue). For description of the related regions (in gray) see Sect. 4. The dashed lines are $x = d - 1$ and the phase transition parametrized in x, y

- (i) (Theorem 1.1) $(x - 1)(y - 1) \geq 2$ and $y > 1$,
- (ii) (see Sect. 4) $x \geq d - 1$ and $y \geq 0$,
- (iii) (Bencs and Csikvári [4]) $x \geq 1$ and $0 \leq y \leq 1$.

Figure 1 depicts the regions described in Theorem 1.2.

1.2. Plan of the paper. This paper has essentially two parts. In the first part we show that $Z_G(q, w)$ can be approximated by the partition function of a 2-spin model for essentially large girth graphs, namely $Z_G(q, w) \approx Z_G(M'_2, \nu_2)$, where

$$M'_2 = \begin{pmatrix} 1 + w & 1 \\ 1 & 1 + \frac{w}{q-1} \end{pmatrix} \quad \text{and} \quad \nu_2 = \begin{pmatrix} 1 \\ q - 1 \end{pmatrix}.$$

The precise statement is Theorem 2.4. In this theorem we do not use that G is regular so we believe that this statement is very useful for studying random cluster model on other essentially large girth graphs like Erdős–Rényi random graphs $G(n, \frac{c}{n})$. This statement implies that for an essentially large girth sequence of d -regular graphs $(G_n)_n$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w) = \lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(M'_2, \nu_2).$$

At that point one can simply cite a theorem of Sly and Sun [23,24] (relying on a theorem of Dembo and Montanari [12]) that shows that the aforementioned limit exists since $\det(M'_2) > 0$. Indeed, Sly and Sun [23,24] proved that for regular graphs any

2-spin model $(N, \underline{\mu})$ having a positive determinant is equivalent with a ferromagnetic Ising model. Dembo and Montanari [12] showed that for the ferromagnetic Ising-model the limit indeed exists. In these papers the main technique is an abstract interpolation method. Nevertheless, in this paper we do not rely on these papers, instead we build out a little theory for “ferromagnetic” 2-spin models that builds on Lee–Yang theory [18, 26, 27] and the gauge theory of Chertkov and Chernyak [9, 10]. This approach has some additional gains. First of all, it shows that the limit exists not only for essentially large girth sequence of d -regular graphs but for Benjamini–Schramm convergent sequence of d -regular graphs (for the definition of a Benjamini–Schramm convergent graph sequence see Definition 3.27, for the precise statement see Theorem 3.29). Secondly, we prove a stability theorem that shows that if a d -regular graph contains a linear number of short cycles, that is, it contains at least $\varepsilon v(G)$ cycles of length at most g for some fixed ε and g , then both $Z_G(q, w)$ and $Z_G(N, \underline{\mu})$ are exponentially larger than the number obtained for an essentially large girth sequence of d -regular graphs (for the precise statement see Theorem 3.41 and the remark after the theorem).

So the second part of the paper is an elaborate analysis of the 2-spin model $(N, \underline{\mu})$, where N is a positive definite 2×2 matrix with positive entries and $\underline{\mu}$ is a vector in \mathbb{R}^2 with positive entries. We will show that there exists a quantity $\Phi_d(N, \underline{\mu})$ such that for every essentially large girth sequence of d -regular graphs $(G_n)_n$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu}) = \ln \Phi_d(N, \underline{\mu}).$$

To analyse these models we use a strategy that can be of independent interest. The idea is that we can associate many different polynomials to the same computational problem and the zeros of one of these polynomials satisfy Lee–Yang theorem, that is, they are on a circle. Indeed, for a d -regular graph G let us introduce the following polynomials [5, 26]

$$F_G(x_0, \dots, x_d) = \sum_{A \subseteq E} \left(\prod_{v \in V} x_{d_A(v)} \right),$$

and a bit more generally,

$$F_G(x_0, \dots, x_d | z) = \sum_{A \subseteq E} \left(\prod_{v \in V} x_{d_A(v)} \right) z^{2|A|} = F_G(x_0, x_1 z, x_2 z, \dots, x_d z^d).$$

We call $F_G(x_0, \dots, x_d)$ and $F_G(x_0, \dots, x_d | z)$ the subgraph counting polynomial.

We show that if N is a 2×2 positive definite matrix, then there are vectors $\underline{v}(t) \in \mathbb{R}^{d+1}$ for each $t \in [0, 2\pi]$ such that $F_G(\underline{v}(t)) = Z_G(N, \underline{\mu})$. We will use these polynomials for two different things. First we show that there exists a t_1 such that the zeros of $F_G(\underline{v}(t_1) | z)$ lie on a circle for all d -regular graphs G . This will enable us to use a standard technique about limits of root measures. More details about this plan can be found in the introduction of Sect. 3. The second application is that there is a t_0 such that the first coordinate of $\underline{v}(t_0)$ is exactly $\Phi_d(N, \underline{\mu})$ and all other coordinates have a nice sign structure. This will enable us to prove the aforementioned stability theorem for graphs containing a linear number of short cycles.

Notations. Given a graph $G = (V, E)$ we use the notation $v(G)$ and $e(G)$ for the number of vertices and edges, respectively. Given a set $S \subseteq V$ let $E(S)$ denote the edges induced by S , that is, $\{(u, v) \in E(G) \mid u, v \in S\}$ and let $e(S) = |E(S)|$. Let $G[S]$ denote the

induced subgraph with vertex set S and edge set $E(S)$. Similarly, $e(V - S)$ is the number of edges induced by $V \setminus S$, and $G - S$ denotes the subgraph induced by $V \setminus S$.

For an $A \subseteq E(G)$ and a $v \in V$ let $d_A(v)$ denote the degree of the vertex v in the graph (V, A) , that is, the number of edges of A incident to v .

For an $A \subseteq E(G)$ and an $S \subseteq V(G)$ let $A[S] = \{(u, v) \in A \mid u, v \in S\}$, so these are the edges of A that are induced by S .

Given graphs H and G let $\text{hom}(H, G)$ denote the number of homomorphisms from H to G , that is, the number of maps $\varphi : V(H) \rightarrow V(G)$ such that $(\varphi(u), \varphi(v)) \in E(G)$ whenever $(u, v) \in E(H)$.

The notation $[q]$ stands for the set $\{1, 2, \dots, q\}$. We denote the scalar products of vectors \underline{x} and \underline{y} by $\langle \underline{x}, \underline{y} \rangle$.

This paper is organized as follows. In the next section we introduce the rank 1 and rank 2 approximations of $Z_G(q, w)$ and study its basic properties. In Sect. 3 we study the rank 2 approximation or more generally, $Z_G(N, \underline{\mu})$. We end the paper with some remarks about the case $1 < q < 2$.

2. Approximations

In this section we introduce various approximations of the partition function of the random cluster model. In the sequel the rank 2 approximation will be especially important for us.

2.1. Rank 1 approximation. For motivational purposes let us assume for a moment that q is a positive integer. Then it is known that

$$Z_G(q, w) = Z_G(M),$$

where M is the $q \times q$ matrix with entries $1 + w$ in the diagonal and 1's as off-diagonal elements. It is a natural idea to approximate M with the rank 1 matrix M_1 such that the sum of all entries of M and M_1 are equal. In other words, let M_1 be the $q \times q$ matrix with entries $1 + \frac{w}{q}$ everywhere. Note that by the definition of $Z_G(M_1)$ we have

$$Z_G(M_1) = q^{v(G)} \left(1 + \frac{w}{q} \right)^{e(G)}.$$

Let us call the quantity

$$Z_G^{(1)}(q, w) = q^{v(G)} \left(1 + \frac{w}{q} \right)^{e(G)}$$

the rank 1 approximation of $Z_G(q, w)$. This quantity makes sense even if q is positive, but not necessarily integer and we will refer to it as the rank 1 approximation of $Z_G(q, w)$ even in this case.

Lemma 2.1. *If $q \geq 1$, then*

$$Z_G(q, w) \geq Z_G^{(1)}(q, w).$$

If $0 < q \leq 1$, then

$$Z_G(q, w) \leq Z_G^{(1)}(q, w).$$

Proof. Using the fact that $k(A) \geq v(G) - |A|$ for an $A \subseteq E(G)$ we get that for $q \geq 1$ we have

$$Z_G(q, w) = \sum_{A \subseteq E(G)} q^{k(A)} w^{|A|} \geq \sum_{A \subseteq E(G)} q^{v(G)-|A|} w^{|A|} = q^{v(G)} \left(1 + \frac{w}{q}\right)^{e(G)}.$$

For $q \leq 1$ we have the opposite inequality in the above computation. □

Lemma 2.1 implies that for any d -regular graph G and $q > 1$ we have

$$Z_G(q, w)^{1/v(G)} \geq q \left(1 + \frac{w}{q}\right)^{d/2}.$$

For $q < 1$ the same quantity is an upper bound. Note that the very same quantity appears as $\Phi_{d,q,w}(0)$ in Theorem 1.1.

2.2. Rank 2 approximation. What is better than a rank 1 approximation? Naturally, a rank 2 approximation.

Again for motivational purposes let us assume for a moment that $q \geq 2$ is an integer. This time let us approximate the matrix M with the following rank 2 matrix M_2 .

$$M_2 = \begin{pmatrix} 1+w & 1 & \dots & 1 \\ 1 & 1 + \frac{w}{q-1} & \dots & 1 + \frac{w}{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 + \frac{w}{q-1} & \dots & 1 + \frac{w}{q-1} \end{pmatrix}.$$

Then

$$Z_G(M_2) = \sum_{S \subseteq V} (1+w)^{e(S)} (q-1)^{v(G)-|S|} \left(1 + \frac{w}{q-1}\right)^{e(G-S)}.$$

Indeed, let $S = \varphi^{-1}(1)$ in the definition of $Z_G(M_2)$. Let us introduce the quantity

$$Z_G^{(2)}(q, w) = \sum_{S \subseteq V} (1+w)^{e(S)} (q-1)^{v(G)-|S|} \left(1 + \frac{w}{q-1}\right)^{e(G-S)}.$$

The definition of $Z_G^{(2)}(q, w)$ makes perfect sense if $q > 1$, but not necessarily integer and we will refer to it as the rank 2 approximation of $Z_G(q, w)$. Recall that

$$M'_2 = \begin{pmatrix} 1+w & 1 \\ 1 & 1 + \frac{w}{q-1} \end{pmatrix} \quad \text{and} \quad \underline{v}_2 = \begin{pmatrix} 1 \\ q-1 \end{pmatrix},$$

and note that

$$Z_G^{(2)}(q, w) = Z_G(M'_2, \underline{v}_2)$$

even if q is not an integer.

This time it is less clear that it is a natural approximation, but as it will turn out this is an asymptotically precise approximation for essentially large girth graphs if $q \geq 2$ and $w \geq 0$. We can prove it through a series of lemmas.

Lemma 2.2. *We have*

$$Z_G(q, w) = \sum_{S \subseteq V} (1+w)^{e(S)} Z_{G-S}(q-1, w).$$

Proof. This identity is trivially true for positive integer q using the interpretation of $Z_G(q, w)$ as the partition function of the Potts-model. Since we have polynomials on both sides we get that it is true for all q and w . \square

Lemma 2.3. *For $q \geq 2$ we have*

$$Z_G(q, w) \geq Z_G^{(2)}(q, w).$$

For $1 < q \leq 2$ we have

$$Z_G(q, w) \leq Z_G^{(2)}(q, w).$$

Proof. By Lemma 2.2 we have

$$Z_G(q, w) = \sum_{S \subseteq V} (1+w)^{e(S)} Z_{G-S}(q-1, w).$$

By the definitions of $Z_G^{(2)}(q, w)$ and $Z_G^{(1)}(q, w)$ we have

$$Z_G^{(2)}(q, w) = \sum_{S \subseteq V} (1+w)^{e(S)} Z_{G-S}^{(1)}(q-1, w).$$

Now the claim follows by Lemma 2.1 \square

Now we are ready to prove that the rank 2 approximation is asymptotically precise for essentially large girth graphs if $q \geq 2$ and $w \geq 0$.

Theorem 2.4. *Let G be a graph on n vertices with $L = L(G, g)$ cycles of length at most $g - 1$. Let $q \geq 2$. Then*

$$Z_G^{(2)}(q, w) \leq Z_G(q, w) \leq q^{n/g+L} Z_G^{(2)}(q, w).$$

Proof. The lower bound was already proven in Lemma 2.3. So we only need to prove the upper bound.

Given $A \subseteq E(G)$ we can decompose A as follows. Let V_1, \dots, V_r be the vertex sets of the connected components of the graph $H = (V, A)$, and let A_1, \dots, A_r be the corresponding subsets of A . If V_i is an isolated vertex, then $A_i = \emptyset$.

Let us say that V_i is small if the induced graph $G[V_i]$ does not contain a cycle. In particular, A_i does not contain a cycle either. Note that it is possible that A_i does not contain a cycle, but the induced graph $G[V_i]$ contains a cycle, and so V_i is not small. Let \mathcal{S}_A denote the set of small V_i 's. We say that V_i is large if it is not small, and we denote by \mathcal{L}_A the set of large V_i 's. Note that $|\mathcal{L}_A| \leq n/g + L$ since each large connected component has size at least g or it contains a cycle of length at most $g - 1$.

Finally, let us say that a vertex set R is compatible with A if R is the union of some small V_i 's. Note that R may be the empty set. We denote this relation by $R \sim A$. Furthermore, let $A[R]$ be the edges of A induced by the vertex set R . Note that if $R \sim A$, then $A[R]$ is a forest. On the other hand, there is no restriction on $A[V \setminus R]$. Figure 2 depicts an example for the introduced concepts.

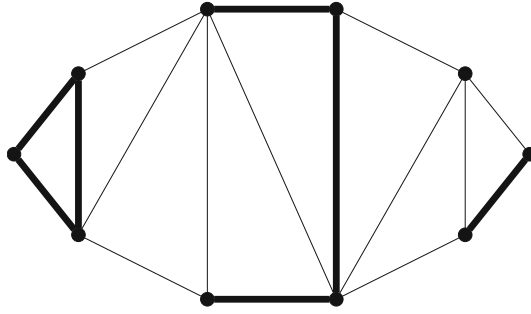


Fig. 2. A subgraph A is depicted with thick edges. There are 4 components. The edge sets with connected components of size 3 and 4 belong to A_ℓ . The edge sets with connected components of size 1 and 2 belong to A_s . A compatible set R is either the vertex set of the latter components or the empty set or the union of these two connected components

Let $k(R, A[[R]])$ denote the number of connected components of the graph $(R, A[[R]])$. By the binomial identity we have

$$q^{|\mathcal{S}_A|} = ((q - 1) + 1)^{|\mathcal{S}_A|} = \sum_{R \sim A} (q - 1)^{k(R, A[[R]])}.$$

Then

$$\begin{aligned} Z_G(q, w) &= \sum_{A \subseteq E(G)} q^{k(A)} w^{|A|} \\ &= \sum_{A \subseteq E(G)} q^{|\mathcal{S}_A| + |\mathcal{L}_A|} w^{|A|} \\ &\leq q^{n/g+L} \sum_{A \subseteq E(G)} q^{|\mathcal{S}_A|} w^{|A|} \\ &= q^{n/g+L} \sum_{A \subseteq E(G)} \sum_{R: R \sim A} (q - 1)^{k(R, A[[R]])} w^{|A|} \\ &= q^{n/g+L} \sum_{R \subseteq V(G)} \sum_{A: R \sim A} (q - 1)^{k(R, A[[R]])} w^{|A[[R]]| + |A[V \setminus R]|} \\ &= q^{n/g+L} \sum_{R \subseteq V(G)} (1 + w)^{e(V \setminus R)} \sum_D (q - 1)^{k(R, D)} w^{|D|}, \end{aligned}$$

where in the last sum, $D = A[[R]]$ is a subset of the edges induced by R such that none of the induced connected components contains a cycle. Then

$$\sum_D (q - 1)^{k(R, D)} w^{|D|} = \sum_D (q - 1)^{|R| - |D|} w^{|D|} \leq (q - 1)^{|R|} \left(1 + \frac{w}{q - 1}\right)^{e(R)}.$$

Hence

$$Z_G(q, w) \leq q^{n/g+L} \sum_{R \subseteq V(G)} (1 + w)^{e(V \setminus R)} Z_{G[R]}^{(1)}(q - 1, w),$$

that is

$$Z_G(q, w) \leq q^{n/s+L} Z_G^{(2)}(q, w).$$

□

The following theorem is an immediate consequence of Theorem 2.4.

Theorem 2.5. *Let $q \geq 2$ and $w \geq 0$. Let $(G_n)_n$ be an essentially large girth sequence of d -regular graphs. If the limit*

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}^{(2)}(q, w)$$

exists, then the limit

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w)$$

exists too, and they have the same value.

3. Ferromagnetic 2-Spin Models

In this section we analyze the rank 2 approximation of the random cluster model. Since $Z_G^{(2)}(q, w) = Z_G(M'_2, \underline{v}_2)$ for a 2×2 matrix M'_2 we will actually prove that if N is a 2×2 positive definite matrix with positive entries and $\underline{\mu} = (\mu_1, \mu_2)$ is a positive vector, then

$$\lim_{n \rightarrow \infty} Z_{G_n}(N, \underline{\mu})^{1/v(G_n)} = \Phi_d(N, \underline{\mu})$$

exists for every essentially large girth sequence of d -regular graphs $(G_n)_n$. In fact, we will prove a much stronger theorem about Benjamini–Schramm convergent graph sequences.

The plan is to connect the quantity $Z_G(N, \underline{\mu})$ with Lee–Yang theory. This connection is built out through the so-called subgraph counting polynomial (see [26]).

3.1. Subgraph counting polynomial. From now on we always assume that G is a d -regular graph.

Let us introduce the so-called subgraph counting polynomial

$$F_G(x_0, \dots, x_d) = \sum_{A \subseteq E} \left(\prod_{v \in V} x_{d_A(v)} \right),$$

and a bit more generally,

$$F_G(x_0, \dots, x_d | z) = \sum_{A \subseteq E} \left(\prod_{v \in V} x_{d_A(v)} \right) z^{2|A|} = F_G(x_0, x_1 z, x_2 z, \dots, x_d z^d)$$

As an example we give the subgraph counting polynomial $F_{K_5}(x_0, x_1, x_2, x_3, x_4)$ of the complete graph K_5 on 5 vertices. The first term corresponds to the empty subgraph, the last term corresponds to the graph itself.

$$\begin{aligned} & x_0^5 + 10x_0^3x_1^2 + 15x_0x_1^4 + 30x_0^2x_1^2x_2 + 30x_1^4x_2 + 60x_0x_1^2x_2^2 + 10x_0^2x_2^3 + 70x_1^2x_2^3 + 15x_0x_2^4 \\ & + 12x_2^5 + 20x_0x_1^3x_3 + 60x_1^3x_2x_3 + 60x_0x_1x_2^2x_3 + 120x_1x_2^3x_3 + 60x_1^2x_2x_3^2 + 30x_0x_2^2x_3^2 \\ & + 70x_2^3x_3^2 + 60x_1x_2x_3^3 + 5x_0x_3^4 + 30x_2x_3^4 + 5x_1^4x_4 + 30x_1^2x_2^2x_4 + 15x_2^4x_4 + 60x_1x_2^2x_3x_4 \\ & + 60x_2^2x_3^2x_4 + 20x_1x_3^3x_4 + 15x_3^4x_4 + 10x_2^3x_4^2 + 30x_2x_3^2x_4^2 + 10x_3^2x_4^3 + x_4^5. \end{aligned}$$

The general plan is the following. In the next section we show that there are vectors $\underline{v}(t)$ for each $t \in [0, 2\pi]$ such that

$$F_G(\underline{v}(t)) = Z_G(N, \underline{\mu}).$$

We will show that there exists a t_1 such that all zeros of $F_G(\underline{v}(t_1)|z)$ lie on a circle for all d -regular graph G . This will imply the convergence of the sequence $\frac{1}{v(G_n)} \ln Z_G(N, \underline{\mu})$ for not only essentially large girth d -regular graphs but for all Benjamini–Schramm graph convergences.

We will also show that there exists a t_0 such that the first coordinate of $\underline{v}(t_0)$ is exactly $\Phi_d(N, \underline{\mu})$, and all other coordinates have a nice sign structure. This will enable us to show that $Z_G(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)}$ for all d -regular graph G , and if G contains a linear number of short cycles, then $Z_G(N, \underline{\mu}) \geq ((1+\delta)\Phi_d(N, \underline{\mu}))^{v(G)}$ for some $\delta > 0$.

3.2. Rank 2 matrices. Suppose that we can write an $r \times r$ matrix N into the form $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ and let $\underline{\mu} \in \mathbb{R}^r$. Then

$$\begin{aligned} Z_G(N, \underline{\mu}) &= \sum_{\varphi: V \rightarrow [r]} \prod_{v \in V} \mu_{\varphi(v)} \prod_{(u,v) \in E} N_{\varphi(u)\varphi(v)} \\ &= \sum_{\varphi: V \rightarrow [r]} \prod_{v \in V} \mu_{\varphi(v)} \prod_{(u,v) \in E} (\underline{a}\underline{a}^T + \underline{b}\underline{b}^T)_{\varphi(u)\varphi(v)} \\ &= \sum_{A \subseteq E} \sum_{\varphi: V \rightarrow [r]} \prod_{v \in V} \mu_{\varphi(v)} \prod_{(u,v) \in E \setminus A} (\underline{a}\underline{a}^T)_{\varphi(u)\varphi(v)} \prod_{(u,v) \in A} (\underline{b}\underline{b}^T)_{\varphi(u)\varphi(v)} \\ &= \sum_{A \subseteq E} \sum_{\varphi: V \rightarrow [r]} \prod_{v \in V} \mu_{\varphi(v)} \prod_{(u,v) \in E \setminus A} (\underline{a}_{\varphi(u)}\underline{a}_{\varphi(v)}) \prod_{(u,v) \in A} (\underline{b}_{\varphi(u)}\underline{b}_{\varphi(v)}) \\ &= \sum_{A \subseteq E} \prod_{v \in V} \left(\sum_{k=1}^r \mu_k a_k^{d-d_S(v)} b_k^{d_S(v)} \right) \\ &= F_G(r_0, \dots, r_d), \end{aligned}$$

where $r_j = \sum_{k=1}^r \mu_k a_k^{d-j} b_k^j$. On the other hand, \underline{a} and \underline{b} are not the only vectors satisfying $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$. Indeed, let us define the vectors $\underline{a}(t)$ and $\underline{b}(t)$ as follows:

$$\underline{a}(t)_j = a_j \cos(t) + b_j \sin(t),$$

and

$$\underline{b}(t)_j = -a_j \sin(t) + b_j \cos(t).$$

Then $N = \underline{a}(t)\underline{a}(t)^T + \underline{b}(t)\underline{b}(t)^T$. So each pair $\underline{a}(t), \underline{b}(t)$ gives rise to a vector $\underline{v}(t) = (r_0(t), \dots, r_d(t))$ such that

$$F_G(\underline{v}(t)) = Z_G(N, \underline{\mu}).$$

Remark 3.1. We can apply our argument to $N = M'_2, \underline{\mu} = \underline{v}_2$ with the following vectors.

$$\underline{a} = \begin{pmatrix} \sqrt{1 + \frac{w}{q}} \\ \sqrt{1 + \frac{w}{q}} \end{pmatrix} \quad \text{and} \quad \underline{b} = \begin{pmatrix} \sqrt{\frac{(q-1)w}{q}} \\ -\sqrt{\frac{w}{q(q-1)}} \end{pmatrix}.$$

One can check that $M'_2 = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ indeed holds true. We can again introduce the vectors $\underline{a}(t), \underline{b}(t)$ giving rise to a vector $\underline{v}(t) = (r_0(t), \dots, r_d(t))$ such that

$$F_G(\underline{v}(t)) = Z_G(M'_2, \underline{v}_2) = Z_G^{(2)}(q, w).$$

In this case

$$\begin{aligned} r_j(t) &= \sum_{k=1}^2 \mu_k a(t)_k^{d-j} b(t)_k^j \\ &= \left(\sqrt{1 + \frac{w}{q}} \cos(t) + \sqrt{\frac{(q-1)w}{q}} \sin(t) \right)^{d-j} \\ &\quad \times \left(-\sqrt{1 + \frac{w}{q}} \sin(t) + \sqrt{\frac{(q-1)w}{q}} \cos(t) \right)^j \\ &\quad + (q-1) \left(\sqrt{1 + \frac{w}{q}} \cos(t) - \sqrt{\frac{w}{q(q-1)}} \sin(t) \right)^{d-j} \\ &\quad \times \left(-\sqrt{1 + \frac{w}{q}} \sin(t) - \sqrt{\frac{w}{q(q-1)}} \cos(t) \right)^j. \end{aligned}$$

In particular,

$$\begin{aligned} r_0(t) &= \left(\sqrt{1 + \frac{w}{q}} \cos(t) + \sqrt{\frac{(q-1)w}{q}} \sin(t) \right)^d \\ &\quad + (q-1) \left(\sqrt{1 + \frac{w}{q}} \cos(t) - \sqrt{\frac{w}{q(q-1)}} \sin(t) \right)^d. \end{aligned}$$

In other words, $r_0(t) = \Phi_{d,q,w}(t)$.

3.2.1. Decompositions of 2×2 positive definite matrices Sometimes it will be convenient to require extra conditions about the vectors \underline{a} and \underline{b} in the decomposition of $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$.

Lemma 3.2. *Let N be a 2×2 positive definite matrix with positive entries.*

(i) *Then there exists a decomposition $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ such that $a_1, a_2, b_1 > 0$ and $b_2 < 0$.*

(ii) There is also a decomposition $N = \underline{a}'\underline{a}'^T + \underline{b}'\underline{b}'^T$ such that $a'_1, a'_2, b'_1, b'_2 > 0$.

Proof. First we prove (i). Let $\underline{v}_1, \underline{v}_2$ be the orthonormal set of eigenvectors of N corresponding to eigenvalues $\lambda_1, \lambda_2 > 0$. If $\lambda_1 \geq \lambda_2$, then by the Perron–Frobenius theory we can assume that \underline{v}_1 has positive entries. Since $\underline{v}_1, \underline{v}_2$ are orthogonal, one of the entries of \underline{v}_2 is positive, the other is negative. By considering $-\underline{v}_2$ if necessary we can assume that the first entry is positive, the second is negative. Hence $\underline{a} = \sqrt{\lambda_1}\underline{v}_1$ and $\underline{b} = \sqrt{\lambda_2}\underline{v}_2$ satisfies the conditions.

Next let us prove (ii). We can assume that we have already found an \underline{a} and \underline{b} such that $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ and $a_1, a_2, b_1 > 0$ and $b_2 < 0$. Let

$$a'_1 = a_1 \cos(\alpha) + b_1 \sin(\alpha) \quad \text{and} \quad b'_1 = -a_1 \sin(\alpha) + b_1 \cos(\alpha),$$

and

$$a'_2 = a_2 \cos(\alpha) + b_2 \sin(\alpha) \quad \text{and} \quad b'_2 = -a_2 \sin(\alpha) + b_2 \cos(\alpha),$$

If we choose α such a way that $\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$, that is, $\cos(\alpha) > 0$ and

$$-\frac{a_2}{b_2} > \frac{b_1}{a_1} > 0 > \frac{b_2}{a_2} > \tan(\alpha) > -\frac{a_1}{b_1},$$

then $a'_1, a'_2, b'_1, b'_2 > 0$. Note that $\frac{b'_2}{a'_2} > -\frac{a'_1}{b'_1}$ since $N_{12} = a_1 a_2 + b_1 b_2 > 0$. \square

3.2.2. The functions $a_1(t), a_2(t), b_1(t), b_2(t)$ In this section we introduce some functions that will appear many times in this paper.

Definition 3.3. For $a_1, a_2, b_1, b_2 \in \mathbb{R}$ let

$$\begin{aligned} a_1(t) &= a_1 \cos(t) + b_1 \sin(t) \quad \text{and} \quad b_1(t) = b_1 \cos(t) - a_1 \sin(t), \\ a_2(t) &= a_2 \cos(t) + b_2 \sin(t) \quad \text{and} \quad b_2(t) = b_2 \cos(t) - a_2 \sin(t). \end{aligned}$$

Lemma 3.4. Suppose that for the 2×2 positive definite matrix N we have $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T = \hat{\underline{a}}\hat{\underline{a}}^T + \hat{\underline{b}}\hat{\underline{b}}^T$. Then there exists a t such that $\underline{a}(t) = \hat{\underline{a}}$ and $\underline{b}(t) = \hat{\underline{b}}$ or there exists a t such that $\underline{a}(t) = \hat{\underline{a}}$ and $\underline{b}(t) = -\hat{\underline{b}}$, and all vectors of those forms are solutions.

Proof. Consider the vectors $\underline{x} = (a_1, b_1)$ and $\underline{y} = (a_2, b_2)$. Our goal is to prove that $U(2)$ act transitively on the pairs $\underline{x}, \underline{y}$. The equation $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ is equivalent to $N_{11} = \langle \underline{x}, \underline{x} \rangle$, $N_{12} = \langle \underline{x}, \underline{y} \rangle$, $N_{22} = \langle \underline{y}, \underline{y} \rangle$. Thus we know the length of $\underline{x}, \underline{y}$, and from these the angle between them. Thus with unitary operation we can transform any solution to any other solution, and by a unitary action applied to a solution we always get a solution. \square

Remark 3.5. For the specific choice $a_1 = a_2 = \sqrt{1 + \frac{w}{q}}$, $b_1 = \sqrt{\frac{(q-1)w}{q}}$ and $b_2 = -\sqrt{\frac{w}{q(q-1)}}$ we use the notation

$$\begin{aligned} a_{q,w,1}(t) &= \sqrt{1 + \frac{w}{q}} \cos(t) + \sqrt{\frac{(q-1)w}{q}} \sin(t) \quad \text{and} \\ b_{q,w,1}(t) &= -\sqrt{1 + \frac{w}{q}} \sin(t) + \sqrt{\frac{(q-1)w}{q}} \cos(t), \end{aligned}$$

$$a_{q,w,2}(t) = \sqrt{1 + \frac{w}{q}} \cos(t) - \sqrt{\frac{w}{q(q-1)}} \sin(t) \quad \text{and}$$

$$b_{q,w,2}(t) = -\sqrt{1 + \frac{w}{q}} \sin(t) - \sqrt{\frac{w}{q(q-1)}} \cos(t).$$

We collected some claims about $a_1(t)$, $a_2(t)$, $b_1(t)$, $b_2(t)$ whose proof is just a straightforward computation. First we describe the sign structure of the functions $a_1(t)$, $a_2(t)$, $b_1(t)$, $b_2(t)$ on the interval $[0, \frac{\pi}{2})$.

Lemma 3.6. *Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ such that $a_1, a_2, b_1 > 0$ and $b_2 < 0$ and $a_1 a_2 + b_1 b_2 > 0$. Let $t \in [0, \frac{\pi}{2})$. Then*

- (a) if $0 \leq \tan(t) \leq \frac{b_1}{a_1}$ we have $a_1(t), a_2(t), b_1(t) \geq 0$ and $b_2(t) < 0$,
- (b) if $\frac{b_1}{a_1} \leq \tan(t) \leq \frac{a_2}{-b_2}$ we have $a_1(t), a_2(t) \geq 0$ and $b_1(t), b_2(t) \leq 0$,
- (c) if $\frac{a_2}{-b_2} \leq \tan(t)$, then $a_1(t) > 0$ and $a_2(t), b_1(t), b_2(t) \leq 0$

Lemma 3.7. *Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$. Then*

$$\frac{\partial}{\partial t} \left(\frac{a_1(t)b_1(t)}{a_2(t)b_2(t)} \right) = \frac{(a_1 a_2 + b_1 b_2)(a_2 b_1 - a_1 b_2)}{a_2(t)^2 b_2(t)^2}.$$

Lemma 3.8. *Let Q be a 2×2 real matrix with non-zero determinant. Then for every $c \in \mathbb{R}$, there is a unique $t \in [0, \pi)$ such that $\frac{Q_{11} \cos(t) + Q_{12} \sin(t)}{Q_{21} \cos(t) + Q_{22} \sin(t)} = c$.*

Proof. Let $F_Q(t) = \frac{Q_{11} \cos(t) + Q_{12} \sin(t)}{Q_{21} \cos(t) + Q_{22} \sin(t)}$. We have $\frac{\partial}{\partial t} F_Q(t) = \frac{Q_{12} Q_{21} - Q_{11} Q_{22}}{(Q_{21} \cos(t) + Q_{22} \sin(t))^2}$. Hence $F_Q(t)$ is either strictly monotone decreasing or strictly monotone increasing on $[0, \pi)$ depending on the sign of $\det(Q)$ with a discontinuity at t_0 , where $\tan(t_0) = -\frac{Q_{21}}{Q_{22}}$. Since $F_Q(0) = F_Q(\pi) = \frac{Q_{11}}{Q_{21}}$, and $\lim_{t \searrow t_0} F_Q(t) = \pm\infty$ and $\lim_{t \nearrow t_0} F_Q(t) = \mp\infty$, the claim follows.

We will also use the following identities.

Lemma 3.9. *For arbitrary $a_1, a_2, b_1, b_2 \in \mathbb{R}$ we have*

$$a_1(t)^2 + b_1(t)^2 = a_1^2 + b_1^2, \quad a_2(t)^2 + b_2(t)^2 = a_2^2 + b_2^2,$$

$$a_1(t)a_2(t) + b_1(t)b_2(t) = a_1 a_2 + b_1 b_2, \quad a_2(t)b_1(t) - a_1(t)b_2(t) = a_2 b_1 - a_1 b_2.$$

Remark 3.10. In case of $Z_G^{(2)}(q, w)$ we get

$$a_{q,w,1}(t)a_{q,w,2}(t) + b_{q,w,1}(t)b_{q,w,2}(t) = 1.$$

It is also true that

$$a_{q,w,1}(t)b_{q,w,1}(t) + (q-1)a_{q,w,2}(t)b_{q,w,2}(t) = -q \cos(t) \sin(t).$$

3.3. The vector $\underline{v}(t_1)$. In this section we show that there exists a t_1 such that for all d -regular graph G the zeros of $F_G(\underline{v}(t_1)|z)$ lie on a circle.

3.3.1. *Wagner’s subgraph counting technique.* In this section we will recall some theorem of Wagner (Theorem 3.2 of [26]) about the location of zeros of $F_G(x_0, \dots, x_d|z)$. For any fixed x_0, \dots, x_d let us define the following *key-polynomial*

$$K(x_0, \dots, x_d|z) = \sum_{k=0}^d \binom{d}{k} x_k z^k.$$

Theorem 3.11 (Wagner [26]). *If $K(x_0, \dots, x_d|z)$ has no complex zero in the open disk of radius κ around 0, then $F_G(x_0, \dots, x_d|z)$ has no complex zero in the open disk of radius κ around 0 for any d -regular graph G .*

If $K(x_0, \dots, x_d|z)$ has no complex zero in the complement of a closed disk of radius κ around 0, then $F_G(x_0, \dots, x_d|z)$ has no complex zero in the complement of a closed disk of radius κ around 0 for any d -regular graph G .

In particular, if $K(x_0, \dots, x_d|z)$ has only zeros on the circle of radius κ around 0, then $F_G(x_0, \dots, x_d|z)$ has complex zeros only on the circle of radius κ for any d -regular graph G .

3.3.2. *Key polynomials for rank 2 matrices* Suppose that we have a rank 2 matrix N of the form $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T \in \mathbb{R}$ and a $\underline{\mu} \in \mathbb{R}^2$. Then we know that $F_G(\underline{v}(t)) = Z_G(N, \underline{\mu})$, where $\underline{v}(t) = (r_0(t), \dots, r_d(t))$ for any $t \in [0, 2\pi)$.

Lemma 3.12. *Let $\underline{a}, \underline{b}, \underline{\mu} \in \mathbb{R}^r$. For $k = 1, \dots, r$ let*

$$a_k(t) = a_k \cos(t) + b_k \sin(t) \quad \text{and} \quad b_k(t) = b_k \cos(t) - a_k \sin(t),$$

and for $j = 0, \dots, d$ let $r_j(t) = \sum_{k=1}^r \mu_k a_k(t)^{d-j} b_k(t)^j$. Finally, let $\underline{v}(t) = (r_0(t), \dots, r_d(t))$ and

$$K(\underline{v}(t)|z) = \sum_{j=0}^d \binom{d}{j} r_j(t) z^j.$$

Then

$$K(\underline{v}(t)|z) = \sum_{k=1}^r \mu_k (b_k(t)z + a_k(t))^d$$

Proof. By definition we have

$$\begin{aligned} K(\underline{v}(t)|z) &= \sum_{j=0}^d \binom{d}{j} r_j(t) z^j \\ &= \sum_{j=0}^d \binom{d}{j} \left(\sum_{k=1}^r \mu_k a_k(t)^{d-j} b_k(t)^j \right) z^j \\ &= \sum_{k=1}^r \sum_{j=0}^d \binom{d}{j} \mu_k a_k(t)^{d-j} b_k(t)^j z^j \\ &= \sum_{k=1}^r \mu_k (a_k(t) + b_k(t)z)^d. \end{aligned}$$

□

Lemma 3.13. *Let $\mu_1, \mu_2 \in \mathbb{R}$ and $a_1, a_2, b_1, b_2 \in \mathbb{R}$ such that $\mu_1, \mu_2 > 0$, then all the complex zeros of $K(\underline{v}(t)|z)$ are on a circle or on a line.*

Moreover, if t_1 satisfies

$$\frac{a_1(t_1)b_1(t_1)}{a_2(t_1)b_2(t_1)} = \left(\frac{\mu_2}{\mu_1}\right)^{2/d},$$

then the circle has center at 0. Furthermore, the radius of this circle is

$$R_c = \left(\frac{\mu_2}{\mu_1}\right)^{1/d} \left| \frac{a_2(t_1)}{b_1(t_1)} \right| = \left(\frac{\mu_2}{\mu_1}\right)^{-1/d} \left| \frac{a_1(t_1)}{b_2(t_1)} \right| = \left| \frac{a_1(t_1)a_2(t_1)}{b_1(t_1)b_2(t_1)} \right|^{1/2}.$$

Proof. From Lemma 3.12 we have that

$$K(\underline{v}(t)|z) = \mu_1(a_1(t) + b_1(t)z)^d + \mu_2(a_2(t) + b_2(t)z)^d.$$

Let us assume that $K(\underline{v}(t)|\zeta) = 0$.

If $a_1(t) + b_1(t)\zeta = a_2(t) + b_2(t)\zeta = 0$, then ζ is the only zero of $K(\underline{v}(t)|z)$ with multiplicity d , thus all the complex zeros are on a circle of radius $|\zeta| = \left| \frac{a_1(t)a_2(t)}{b_1(t)b_2(t)} \right|^{1/2}$ with center at 0.

If $a_1(t) + b_1(t)\zeta$ or $a_2(t) + b_2(t)\zeta$ is not 0, then by symmetry we can assume that $a_2(t) + b_2(t)\zeta \neq 0$, and we get that

$$\begin{aligned} \mu_1(a_1(t) + b_1(t)\zeta)^d + \mu_2(a_2(t) + b_2(t)\zeta)^d &= 0 \\ \left(\frac{a_1(t) + b_1(t)\zeta}{a_2(t) + b_2(t)\zeta}\right)^d &= -\frac{\mu_2}{\mu_1} \\ M_t(\zeta)^d &= -\frac{\mu_2}{\mu_1}, \end{aligned}$$

where $M_t(z) = \frac{a_1(t)+b_1(t)z}{a_2(t)+b_2(t)z}$ is a Möbius transformation with real coefficients. Let us introduce the notation $T = \left(\frac{\mu_2}{\mu_1}\right)^{2/d}$. Thus we obtained that for any ζ zero of $K(\underline{v}(t)|z)$ we have

$$|M_t(\zeta)| = \sqrt{T}.$$

Since $M_t(z)$ is Möbius transformation, therefore $M_t^{(-1)}(z)$ maps cycles into cycles and lines, i.e. $\zeta \in M_t^{(-1)}(S_{\sqrt{T}})$, where S_c is a circle of radius c around 0.

In order to prove the second part of the statement we have to investigate when does the circle $M_{t_1}^{(-1)}(S_{\sqrt{T}})$ have a center at 0. Since $M_t(z)$ is Möbius transformation with real coefficients, thus $M_t^{(-1)}(z)$ is also a Möbius transformation with real coefficients. This means that the image of a circle that is perpendicular to the real line is also perpendicular to the real line. We claim that $M_{t_1}^{(-1)}(S_{\sqrt{T}})$ is not a line. To see it is enough to show that $M_{t_1}^{(-1)}(\pm\sqrt{T})$ is not ∞ , or equivalently $M_{t_1}(\infty) \neq \pm\sqrt{T}$. If this would be the case, then $M_{t_1}(\infty) = \frac{b_1(t_1)}{b_2(t_1)} = \pm\sqrt{T}$ would imply that $a_1(t_1) + b_1(t_1)z$ and $a_2(t_1) + b_2(t_1)z$ have a common zero, which lead us to a contradiction.

Thus the center of $M_{t_1}^{(-1)}(S_{\sqrt{T}})$ is at

$$\frac{1}{2} \left(M_{t_1}^{(-1)}(\sqrt{T}) + M_{t_1}^{(-1)}(-\sqrt{T}) \right).$$

This is 0 if and only if

$$\begin{aligned} M_{t_1}^{(-1)}(\sqrt{T}) &= -M_{t_1}^{(-1)}(-\sqrt{T}) \\ \frac{a_2(t_1)\sqrt{T} - a_1(t_1)}{-b_2(t_1)\sqrt{T} + b_1(t_1)} &= \frac{a_2(t_1)\sqrt{T} + a_1(t_1)}{b_2(t_1)\sqrt{T} + b_1(t_1)} \\ a_2(t_1)b_2(t_1)T &= a_1(t_1)b_1(t_1) \end{aligned}$$

This is equivalent to

$$T = \frac{a_1(t_1)b_1(t_1)}{a_2(t_1)b_2(t_1)}.$$

To find the corresponding radius we have to calculate $|M_{t_1}^{(-1)}(\sqrt{T})|$.

$$\begin{aligned} M_{t_1}^{(-1)}(\sqrt{T}) &= \frac{a_2(t_1)\sqrt{T} - a_1(t_1)}{-b_2(t_1)\sqrt{T} + b_1(t_1)} \\ &= \frac{a_2(t_1)}{b_1(t_1)} \sqrt{T} \left(\frac{a_2(t_1)b_1(t_1)\sqrt{T} - a_1(t_1)b_1(t_1)}{-a_2(t_1)b_2(t_1)T + b_1(t_1)a_2(t_1)\sqrt{T}} \right) \\ &= \frac{a_2(t_1)}{b_1(t_1)} \sqrt{T} \end{aligned}$$

This implies that $R_c = \sqrt{T} \left| \frac{a_2(t_1)}{b_1(t_1)} \right|$. Thus by equation $T = \frac{a_1(t_1)b_1(t_1)}{a_2(t_1)b_2(t_1)}$ we also have $R_c = T^{-1/2} \left| \frac{a_1(t_1)}{b_2(t_1)} \right|$, and by multiplying the two equations we get that $R_c^2 = \left| \frac{a_1(t_1)a_2(t_1)}{b_1(t_1)b_2(t_1)} \right|$. \square

Lemma 3.14. *Let $a_1, a_2, b_1, b_2, \mu_1, \mu_2 \in \mathbb{R}$ such that $a_1, a_2, b_1, \mu_1, \mu_2 > 0$ and $b_2 < 0$ and $a_1a_2 + b_1b_2 > 0$, then there is a unique $t_1 \in [0, \frac{\pi}{2}]$ such that $\frac{b_1}{a_1} < \tan(t_1) < \frac{a_2}{-b_2}$ and*

$$\frac{a_1(t_1)b_1(t_1)}{a_2(t_1)b_2(t_1)} = \left(\frac{\mu_2}{\mu_1} \right)^{2/d}.$$

For such a t_1 we have $a_1(t_1), a_2(t_1) > 0$ and $b_1(t_1), b_2(t_1) < 0$ implying that $\frac{a_1(t_1)a_2(t_1)}{b_1(t_1)b_2(t_1)} > 0$.

Proof. Note that the function $\frac{a_1(t)b_1(t)}{a_2(t)b_2(t)}$ is only positive at $t \in [0, \frac{\pi}{2}]$ if $a_1(t), a_2(t) > 0$ and $b_1(t), b_2(t) < 0$, that is, $\frac{b_1}{a_1} < \tan(t) < \frac{a_2}{-b_2}$. When $t \rightarrow \arctan\left(\frac{b_1}{a_1}\right)$, then $b_1(t) \rightarrow 0$, and so $\frac{a_1(t)b_1(t)}{a_2(t)b_2(t)} \rightarrow 0$. If $t \rightarrow \arctan\left(\frac{a_2}{-b_2}\right)$, then $a_2(t) \rightarrow 0$, and so $\frac{a_1(t)b_1(t)}{a_2(t)b_2(t)} \rightarrow \infty$. Since

$$\frac{\partial}{\partial t} \left(\frac{a_1(t)b_1(t)}{a_2(t)b_2(t)} \right) = \frac{(a_1a_2 + b_1b_2)(a_2b_1 - a_1b_2)}{a_2(t)^2b_2(t)^2} > 0$$

the function is strictly monotone increasing, hence there is a unique t_1 satisfying $\frac{a_1(t_1)b_1(t_1)}{a_2(t_1)b_2(t_1)} = \left(\frac{\mu_2}{\mu_1}\right)^{2/d}$. \square

Theorem 3.15. *Let N be a 2×2 positive definite matrix with positive entries and let $\mu \in \mathbb{R}_{>0}^2$. Then there exists a $\underline{v}_c \in \mathbb{R}^{d+1}$ and an $R_c(N, \underline{\mu}) \in \mathbb{R}_{>0}$ such that for any \bar{d} -regular graph G we have $Z_G(N, \underline{\mu}) = F_G(\underline{v}_c)$ and all complex zeros of $F_G(\underline{v}_c|z)$ lie on a circle around 0 of radius $R_c(N, \underline{\mu})$. Moreover, $R = R_c(N, \underline{\mu})$ is a positive real solution of*

$$(N_{11}N_{22} - N_{12}^2)R^4 + (-N_{22}^2T + 2N_{12}^2 - N_{11}^2T^{-1})R^2 + (N_{11}N_{22} - N_{12}^2) = 0,$$

where $T = \left(\frac{\mu_2}{\mu_1}\right)^{2/d}$.

Proof. The first part of the claim follows from combining Lemma 3.2, 3.14, 3.13 and Theorem 3.11. Indeed, by Lemma 3.2 we know that there exists a decomposition $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ such that $a_1, a_2, b_1 > 0$ and $b_2 < 0$. Then Lemma 3.14 implies that there exists a t_1 such that $\frac{a_1(t_1)b_1(t_1)}{a_2(t_1)b_2(t_1)} = \left(\frac{\mu_2}{\mu_1}\right)^{2/d}$. Then Lemma 3.13 shows that the zeros of $K(\underline{v}(t_1)|z)$ lie on a circle that has center at 0. Then Theorem 3.11 implies that all complex zeros of $F_G(\underline{v}(t_1)|z)$ lie on a circle around 0 for any d -regular graph G . Thus $\underline{v}_c = \underline{v}(t_1)$ satisfies the conditions of the theorem.

To prove the statement concerning the radius of the circle note that by Lemma 3.9 and 3.13 we have $N_{11} = a_1(t_1)^2 + b_1(t_1)^2$, $N_{22} = a_2(t_1)^2 + b_2(t_1)^2$, $N_{12} = a_1(t_1)a_2(t_1) + b_1(t_1)b_2(t_1)$, $T = \frac{a_1(t_1)b_1(t_1)}{a_2(t_1)b_2(t_1)}$ and $R^2 = \frac{a_1(t_1)a_2(t_1)}{b_1(t_1)b_2(t_1)}$. Let us introduce the notations $\bar{a}_1 = a_1(t_1)$, $\bar{a}_2 = a_2(t_1)$, $\bar{b}_1 = b_1(t_1)$ and $\bar{b}_2 = b_2(t_1)$. Then we get that

$$\begin{aligned} & (N_{11}N_{22} - N_{12}^2)R^4 + (-N_{22}^2T + 2N_{12}^2 - N_{11}^2T^{-1})R^2 + (N_{11}N_{22} - N_{12}^2) \\ &= ((\bar{a}_1^2 + \bar{b}_1^2)(\bar{a}_2^2 + \bar{b}_2^2) - (\bar{a}_1\bar{a}_2 + \bar{b}_1\bar{b}_2)^2) \left(\frac{\bar{a}_1^2\bar{a}_2^2}{\bar{b}_1^2\bar{b}_2^2} + 1 \right) \\ & \quad - (\bar{a}_2^2 + \bar{b}_2^2)^2 \frac{\bar{a}_1\bar{b}_1}{\bar{a}_2\bar{b}_2} \cdot \frac{\bar{a}_1\bar{a}_2}{\bar{b}_1\bar{b}_2} + 2(\bar{a}_1\bar{a}_2 + \bar{b}_1\bar{b}_2)^2 \cdot \frac{\bar{a}_1\bar{a}_2}{\bar{b}_1\bar{b}_2} - (\bar{a}_1^2 + \bar{b}_1^2)^2 \frac{\bar{a}_2\bar{b}_2}{\bar{a}_1\bar{b}_1} \cdot \frac{\bar{a}_1\bar{a}_2}{\bar{b}_1\bar{b}_2} \\ &= \frac{(\bar{a}_1^2\bar{b}_2^2 - 2\bar{a}_1\bar{a}_2\bar{b}_1\bar{b}_2 + \bar{a}_2^2\bar{b}_1^2)(\bar{a}_1^2\bar{a}_2^2 + \bar{b}_1^2\bar{b}_2^2)}{\bar{b}_1^2\bar{b}_2^2} - \frac{(\bar{a}_1^4 + 2\bar{a}_2^2\bar{b}_2^2 + \bar{b}_2^4)\bar{a}_1^2\bar{b}_2^2}{\bar{b}_1^2\bar{b}_2^2} \\ & \quad + \frac{2(\bar{a}_1^2\bar{a}_2^2 + 2\bar{a}_1\bar{a}_2\bar{b}_1\bar{b}_2 + \bar{b}_1^2\bar{b}_2^2)\bar{a}_1\bar{a}_2\bar{b}_1\bar{b}_2}{\bar{b}_1^2\bar{b}_2^2} - \frac{(\bar{a}_1^4 + 2\bar{a}_1^2\bar{b}_1^2 + \bar{b}_1^4)\bar{a}_2^2\bar{b}_2^2}{\bar{b}_1^2\bar{b}_2^2} \end{aligned}$$

Now one can see that everything cancels, and this is indeed 0. \square

3.4. Random regular graphs and Bethe approximation. In this section we recall about some results of Dembo et al. [13] on Bethe approximation. We introduce a quantity $\Phi_d(N, \underline{\mu})$ for which it is true that if G is a random d -regular graph on n vertices, then we have $\mathbb{E}Z_G(N, \underline{\mu}) = n^{O(1)}\Phi_d(N, \underline{\mu})^n$. As a consequence of a theorem of Ruozzi we will also get that $Z_G(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)}$ for a d -regular graph G if N is a 2×2 positive definite matrix.

In general, let $N \in \mathbb{R}_{>0}^{r \times r}$ be a symmetric matrix and $\underline{\mu} \in \mathbb{R}_{>0}^r$. Let $B_{N, \underline{\mu}}$ be a symmetric distribution on $[r]^2$. Let $b_{N, \underline{\mu}}$ be the marginal of $B_{N, \underline{\mu}}$ to its first coordinate. Let us define

$$\mathbb{F}_{\text{hom}}(B_{N, \underline{\mu}}) := \frac{d}{2} H(B_{N, \underline{\mu}}) - (d - 1) H(b_{N, \underline{\mu}}) + [\ln N]_{B_{N, \underline{\mu}}} + \frac{d}{2} [\ln \underline{\mu}]_{b_{N, \underline{\mu}}},$$

where for a probability distribution $P = (p_1, \dots, p_n)$ and a vector $f = (f_1, \dots, f_n)$ we have

$$H(P) = \sum_{i=1}^n p_i \ln \frac{1}{p_i} \quad \text{and} \quad [f]_P = \sum_{i=1}^n p_i f_i$$

with the usual convention $0 \cdot \ln \frac{1}{0} = 0$. The subscript hom in \mathbb{F}_{hom} simply stands for homomorphism. Let

$$\Phi_d(N, \underline{\mu}) = \max_{B_{N, \underline{\mu}}} \exp\left(\mathbb{F}_{\text{hom}}(B_{N, \underline{\mu}})\right).$$

The quantity $\Phi_d(N, \underline{\mu})$ also has a description through belief propagation equation, see Proposition 14.6 of [20] or Section 1.2 of [13]. Let $h \in \mathbb{R}^r$ be a probability distribution. The belief propagation equation or Bethe recursion is

$$\text{BP}(h)_\sigma := \frac{1}{z_h} \mu_\sigma \left(\sum_{\sigma'} N_{\sigma, \sigma'} h_{\sigma'} \right)^{d-1}$$

for all $\sigma \in [r]$, where z_h is the normalizing constant ensuring that $\text{BP}(h)$ is a probability distribution too. Let \mathcal{H}^* be the set of probability distributions for which $\text{BP}(h) = h$. The Bethe functional is defined as

$$\tilde{\Phi}_{N, \underline{\mu}, d}(h) = \ln \left(\sum_{\sigma} \mu_\sigma \left(\sum_{\sigma'} N_{\sigma, \sigma'} h_{\sigma'} \right)^d \right) - \frac{d}{2} \ln \left(\sum_{\sigma, \sigma'} N_{\sigma, \sigma'} h_\sigma h_{\sigma'} \right).$$

Then

$$\Phi_d(N, \underline{\mu}) = \sup_{h \in \mathcal{H}^*} \exp(\tilde{\Phi}_{N, \underline{\mu}, d}(h)).$$

Dembo et al. [13] showed that the quantity $\Phi_d(N, \underline{\mu})$ is directly related to the expected value of $Z_G(N, \underline{\mu})$ for random d -regular graphs.

Theorem 3.16 (Dembo et al. [13]). *Let G be a random d -regular graph on n vertices. Then*

$$\mathbb{E} Z_G(N, \underline{\mu}) = n^{O(1)} \Phi_d(N, \underline{\mu})^n.$$

3.4.1. *A theorem of Ruzozzi.* In this section we show that if N is a 2×2 positive definite matrix with positive entries and $\underline{\mu}$ is a positive vector, then for any d -regular graph G we have $Z_G(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^n$. First we recall the setting of factor graphs.

Definition 3.17. 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 the alphabet \mathcal{X} . For each a there is an associated function $g_a : \mathcal{X}^{\partial a} \rightarrow \mathbb{R}_{\geq 0}$. 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 an example.

Example 3.18. Suppose that $G = (V, E)$ is an (ordinary) graph. We can associate a factor graph \mathcal{G} as follows. For each $v \in V$ we introduce a variable node v and function node v' , and for each edge $e = (u, v)$ we introduce a function node e . In \mathcal{G} let us connect v with v' and $e = (u, v)$ with u and v . Set $\mathcal{X} = [r]$. Let N be a $r \times r$ matrix and $\underline{\mu} \in \mathbb{R}^r$. For each function node v' we introduce the function $g_{v'}(x) = \mu_x$ and for each edge e we introduce the function $g_e(x, y) = N_{x,y}$. Then

$$Z(\mathcal{G}) = Z_G(N, \underline{\mu}).$$

The middle picture of Fig. 3 depicts the factor graph \mathcal{G} for the diamond graph G .

Example 3.19. Let $G = (V, E)$ be a graph. Recall that

$$F_G(x_0, \dots, x_d) = \sum_{A \subseteq E} \left(\prod_{v \in V} x_{d_A(v)} \right).$$

For $(a_0, \dots, a_d) \in \mathbb{R}^{d+1}$ let us consider the following factor graph $\mathcal{G}' = (F', V', E', \mathcal{X}', (g'_a)_{a \in F'})$. We subdivide each edge of E with one vertex. In the resulting bipartite graph one side corresponds to F' , the other side corresponds to V' . So with a slight abuse of notation we have $F' = V$ and $V' = E$. Let $\mathcal{X}' = \{0, 1\}$. For each $v \in V$ let us introduce the function

$$g'_v(x_{e_1}, \dots, x_{e_{d_v}}) = a_{|x|},$$

where $|x| = x_{e_1} + \dots + x_{e_{d_v}}$, and e_1, \dots, e_{d_v} are the edges incident to v . Then

$$Z(\mathcal{G}') = F_G(a_0, \dots, a_d).$$

As we can see \mathcal{G}' is in some sense the dual of \mathcal{G} . The picture on the right hand side of Fig. 3 depicts the factor graph \mathcal{G}' for the diamond graph G .

Next we need the concept of the Bethe approximation for factor graphs. First we need to introduce the pseudo-marginal polytope.

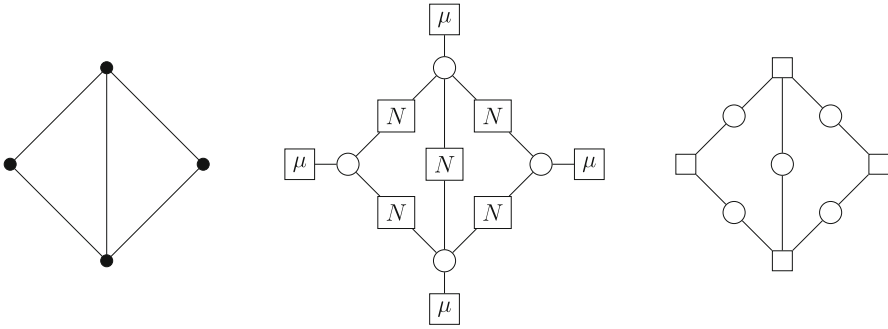


Fig. 3. A graph with two graphical models. Square shape nodes are function vertices, circle shape nodes are variable vertices. We can see the factor graph of Example 3.18 in the middle and the factor graph of Example 3.19 on the right. In this example the original graph is not regular, nevertheless we can see that in \mathcal{G}' the variable nodes correspond to the edges of the original graph while the function nodes correspond to the vertices of the original graph

Definition 3.20. For each variable node v let us introduce a probability distribution b_v on \mathcal{X} , and for each function node a let us 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 or simply pseudo-marginal. The set of such \underline{b} will be denoted by $\text{Mar}(\mathcal{G})$.

Definition 3.21. The Bethe partition function $Z_B(\mathcal{G})$ is defined as follows. Let \mathbb{F} be the following function evaluated on a $\underline{b} \in \text{Mar}(\mathcal{G})$:

$$\mathbb{F}(\underline{b}) = \sum_{a \in F} \sum_{\underline{x} \in \mathcal{X}^{\partial a}} b_a(\underline{x}) \ln \frac{g_a(\underline{x})}{b_a(\underline{x})} - \sum_{v \in V} (1 - |\partial v|) \sum_{x \in \mathcal{X}} b_v(x) \ln b_v(x).$$

The notation \mathbb{F} is consistent with our previous notation \mathbb{F}_{hom} as it will be explained later. Finally, let

$$H_B(\mathcal{G}) = \sup_{\underline{b} \in \text{Mar}(\mathcal{G})} \mathbb{F}(\underline{b}),$$

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 function. We note that if $g_a(x) = 0$, then we require $b_a(x) = 0$ and use the convention $0 \cdot \ln \frac{0}{0} = 0$.

Example 3.22. By continuing examples 3.18 and 3.19 we can consider the Bethe partition functions of \mathcal{G} and \mathcal{G}' . For \mathcal{G} we will denote it by $Z_G^B(N, \underline{\mu})$.

Recall that a function g is log-supermodular if for all $\underline{x}, \underline{y} \in \{0, 1\}^k$ we have

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

where $\underline{x} \wedge \underline{y}, \underline{x} \vee \underline{y} \in \{0, 1\}^k$ such that $(\underline{x} \wedge \underline{y})_i = \min(x_i, y_i)$ and $(\underline{x} \vee \underline{y})_i = \max(x_i, y_i)$ for $i \in [k]$.

Theorem 3.23 (Ruozzi [21]). *Let $\mathcal{G} = (F, V, E, \mathcal{X}, (g_a)_{a \in F})$ be a factor graph with $\mathcal{X} = \{0, 1\}$ such that for all $a \in F$ the functions g_a are log-supermodular. Then $Z(\mathcal{G}) \geq Z_B(\mathcal{G})$.*

Lemma 3.24. *For an $r \times r$ matrix N and $\underline{\mu} \in \mathbb{R}_{\geq 0}^r$, and for a d -regular graph G we have $Z_G^B(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)}$.*

Proof. By using the same probability distribution $B_{N, \underline{\mu}}$ everywhere in the definition of $Z_G^B(N, \underline{\mu})$ the consistency of marginals is immediately satisfied. Then the function \mathbb{F} on this pseudo-marginal simplifies to $\mathbb{F}_{\text{hom}}(B_{N, \underline{\mu}})$, and we get that $Z_G^B(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)}$. □

Theorem 3.25. *For a 2×2 positive definite matrix N with positive entries and $\underline{\mu} \in \mathbb{R}_{\geq 0}^2$, and a d -regular graph G we have $Z_G(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)}$.*

Proof. Note that the log-supermodularity of g_a in the case of the factor graph \mathcal{G} in Example 3.18 simply means that $N_{11}N_{22} \geq N_{12}N_{21}$ which is satisfied as N is positive definite. Hence by combining Ruozzi’s theorem with the previous lemma we get that

$$Z_G(N, \underline{\mu}) \geq Z_G^B(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)}.$$

Remark 3.26. For Theorem 3.25 we will give a new proof in Sect. 3.7 that implies a slightly stronger statement. Namely, if G is a d -regular graph such that for some g and $\varepsilon > 0$ the graph G contains at least $\varepsilon v(G)$ cycles of length g , then $Z_G(N, \underline{\mu}) > ((1 + \delta)\Phi_d(N, \underline{\mu}))^{v(G)}$ for some $\delta = \delta(d, N, \underline{\mu}, g, \varepsilon) > 0$.

3.5. Convergence of $Z_{G_n}(N, \underline{\mu})$. In this section we prove that if N is a 2×2 positive definite matrix with positive entries and $\underline{\mu}$ is a positive vector, then $\frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu})$ is convergent for an essentially large girth sequence of d -regular graphs $(G_n)_n$. In fact, we will prove a stronger statement. Namely, we prove the convergence of the sequence $\frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu})$ for any Benjamini–Schramm convergent graph sequence $(G_n)_n$ of regular graphs.

Definition 3.27. 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 $(G_n)_n$ is *Benjamini–Schramm convergent* if for all finite rooted graphs α and $r > 0$, the probabilities $\mathbb{P}(G_n, \alpha, r)$ converge.

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

Given a vector $\underline{a} \in \mathbb{R}^{d+1}$ and a d -regular graph G on n vertices let $\lambda_1(G), \dots, \lambda_{nd}(G)$ be the zeros of the polynomial $F_G(\underline{a}|z)$. Let us define the probability measure $\rho_{G,\underline{a}}$ on \mathbb{C} as follows:

$$\rho_{G,\underline{a}} := \frac{1}{nd} \sum_{k=1}^{nd} \delta_{\lambda_i(G)},$$

where δ_λ is the Dirac-measure on the number λ .

Lemma 3.28. (a) For any integer $k \geq 0$, a vector $\underline{a} \in \mathbb{R}^{d+1}$ and a Benjamini–Schramm convergent sequence of d -regular graphs $(G_n)_n$ the sequence

$$\int z^k d\rho_{G_n,\underline{a}}(z)$$

is convergent.

(b) Let $\underline{v}_c \in \mathbb{R}^{d+1}$ be such that the zeros of $F_G(\underline{v}_c|z)$ lie on a circle of radius R_c for all graph G . If $(G_n)_n$ is a Benjamini–Schramm convergent sequence of d -regular graphs, then the sequence of measures $\rho_{G_n,\underline{v}_c}$ converges weakly.

Proof. Part (a) is a special case of a much more general theorem claiming that

$$\int z^k d\rho_{G,\underline{a}}(z) = \frac{1}{dv(G)} \sum_{j=1}^{dv(G)} \lambda_j(G)^k$$

can be expressed as $\frac{1}{v(G)} \sum_H c_{H,k} \text{hom}(H, G)$ for a fixed finite set of **connected** graphs H , and the fact that a sequence of bounded degree graphs $(G_n)_n$ is Benjamini–Schramm convergent if and only if for all connected graphs H the sequence $\frac{\text{hom}(H,G_n)}{v(G_n)}$ is convergent. For details see the paper of Csikvári and Frenkel [11].

Part (a) implies part (b) for the following reasons. The weak convergence of measures ρ_n on \mathbb{C} is equivalent with the convergence of $\int z^k \bar{z}^\ell d\rho_n(z)$ for all integers $k, \ell \geq 0$. But if ρ_n are supported on a fixed circle and they are symmetric to the real line, then this is equivalent with the convergence of $\int z^m d\rho_n(z)$ for all positive integer m . \square

Theorem 3.29. For any Benjamini–Schramm convergent sequence of d -regular graphs $(G_n)_n$ the sequence

$$\frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu})$$

is convergent.

Proof. By Theorem 3.15 there exists a $\underline{v}_c \in \mathbb{R}^{d+1}$ such that for any d -regular graph G we have $Z_G(N, \underline{\mu}) = F_G(\underline{v}_c)$ and all zeros of $F_G(\underline{v}_c|z)$ lie on a circle of radius $R_c(N, \underline{\mu})$. First suppose that $R_c = R_c(N, \underline{\mu}) \neq 1$. We have

$$\frac{1}{v(G)} \ln Z_G(N, \underline{\mu}) = \frac{1}{v(G)} \ln F_G(\underline{v}_c|z) \Big|_{z=1}$$

$$\begin{aligned}
 &= \frac{1}{v(G)} \ln \left(\prod_{j=1}^{dv(G)} (1 - \lambda_j(G)) \right) \\
 &= d \int \ln |z - 1| d\rho_{G, \underline{v}_c}(z).
 \end{aligned}$$

The measures $\rho_{G_n, \underline{v}_c}$ are supported on a circle of radius $R_c \neq 1$, thus $\ln |z - 1|$ is a continuous function on a region containing the circle but avoid an open neighborhood of $z = 1$. Since the measures $\rho_{G_n, \underline{v}_c}$ are weakly convergent we get that the sequence $\frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu})$ is convergent.

Next we show that the limit exists even if $R_c(N, \underline{\mu}) = 1$. Let $\Phi_L(N, \underline{\mu})$ be defined by

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu}) = \ln \Phi_L(N, \underline{\mu})$$

if $R_c(N, \underline{\mu}) \neq 1$. Here L in $\Phi_L(N, \underline{\mu})$ simply stands for the word limit. We show that $\Phi_L(N, \underline{\mu})$ is a monotone increasing continuous function of μ_1 . Indeed, if $\mu'_1 < \mu_1$, then

$$Z_G(N, (\mu'_1, \mu_2)) \leq Z_G(N, (\mu_1, \mu_2)) \leq \left(\frac{\mu_1}{\mu'_1} \right)^{v(G)} Z_G(N, (\mu'_1, \mu_2)).$$

So if $R_c(N, (\mu'_1, \mu_2)) \neq 1$, then

$$\begin{aligned}
 \ln \Phi_L(N, (\mu'_1, \mu_2)) &= \lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, (\mu'_1, \mu_2)) \\
 &\leq \liminf_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, (\mu_1, \mu_2)) \\
 &\leq \limsup_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, (\mu_1, \mu_2)) \\
 &\leq \ln \left(\frac{\mu_1}{\mu'_1} \right) + \ln \Phi_L(N, (\mu'_1, \mu_2)).
 \end{aligned}$$

Note that if $R_c(N, (\mu_1, \mu_2)) = 1$, then

$$2(N_{11}N_{22} - N_{12}^2) + (-N_{22}^2T + 2N_{12}^2 - N_{11}^2T^{-1}) = 0,$$

where $T = \left(\frac{\mu_2}{\mu_1} \right)^{2/d}$. For fixed N and μ_2 there are at most two μ_1 such that this equation is satisfied. Thus for such a μ_1 we can define

$$\Phi_L(N, (\mu_1, \mu_2)) = \lim_{\mu'_1 \rightarrow \mu_1} \Phi_L(N, (\mu'_1, \mu_2)),$$

and we get that

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, (\mu_1, \mu_2)) = \Phi_L(N, (\mu_1, \mu_2)).$$

□

Next we need some result about the number of short cycles in random regular graphs. There are many such results in the literature, we chose one.

Lemma 3.30 (McKay et al. [19]). *Let $\{c_1, \dots, c_t\}$ be a non-empty subset of $\{3, \dots, g\}$. For a random regular graph G of order n and degree d , define $M_C(G) = (m_1, \dots, m_t)$, where m_i is the number of cycles of length c_i in G for $1 \leq i \leq t$. For $1 \leq i \leq t$ let $\mu_i = \frac{(d-1)^{c_i}}{2c_i}$. Let S be a set of non-negative integer t -tuples. Then as $n \rightarrow \infty$ the probability that $M_C(G) \in S$ is equal to*

$$(1 + o(1)) \left(\sum_{(m_1, \dots, m_t) \in S} \prod_{i=1}^t \frac{e^{-\mu_i} \mu_i^{m_i}}{m_i!} \right) + o(1).$$

Now we are ready to give a new proof of the fact that the limit of $\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu})$ is $\ln \Phi_d(N, \underline{\mu})$ for random regular graphs and essentially large girth sequence of regular graphs.

Theorem 3.31 (Sly and Sun [23,24] building on Dembo and Montanari [12]). *Let N be a 2×2 positive definite matrix with positive entries and let $\underline{\mu} \in \mathbb{R}_{>0}^2$. If $(G_n)_n$ is an essentially large girth sequence of d -regular graphs, then*

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu}) = \ln \Phi_d(N, \underline{\mu}).$$

The same statement holds true for a sequence of random d -regular graphs with probability one.

Proof. We know from Theorem 3.29 that $\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu})$ exists for an essentially large girth sequence of d -regular graphs. We only have to prove that this limit is $\ln \Phi_d(N, \underline{\mu})$. To prove this it is enough to show one essentially large girth sequence of d -regular graphs $(G_n)_n$ for which $\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu}) = \ln \Phi_d(N, \underline{\mu})$. Let G_n be a random d -regular graph on n vertices, then by Markov’s inequality

$$\mathbb{P} \left(Z_{G_n}(N, \underline{\mu}) \geq n^2 \mathbb{E} Z_{G_n}(N, \underline{\mu}) \right) \leq \frac{1}{n^2}.$$

Note that $\mathbb{E} Z_{G_n}(N, \underline{\mu}) = n^C \Phi_d(N, \underline{\mu})^n$ by Theorem 3.16, and for all graph G we have $Z_{G_n}(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G_n)}$ by Theorem 3.25. By Borel–Cantelli lemma we immediately get that $\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu}) = \ln \Phi_d(N, \underline{\mu})$ holds true with probability one. By Lemma 3.30 we can easily find an essentially large girth sequence of d -regular graphs $(G_n)_n$ such that $\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu}) = \ln \Phi_d(N, \underline{\mu})$. \square

3.6. Trigonometric Bethe approximation. In this subsection we will define some trigonometric polynomial and prove that its maximum is exactly the Bethe approximation.

Definition 3.32. By using the notations of Definition 3.3 for $\underline{a}, \underline{b}, \underline{\mu} \in \mathbb{R}^2$ then

$$\begin{aligned} \Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t) &= \mu_1 a_1(t)^d + \mu_2 a_2(t)^d = \mu_1 (a_1 \cos(t) + b_1 \sin(t))^d \\ &\quad + \mu_2 (a_2 \cos(t) + b_2 \sin(t))^d. \end{aligned}$$

The following lemma is an immediate consequence of Lemma 3.4.

Lemma 3.33. *If $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T = \underline{\hat{a}}\underline{\hat{a}}^T + \underline{\hat{b}}\underline{\hat{b}}^T$, then there exist an $s \in \{-1, 1\}$ and an $\alpha \in [0, 2\pi]$ such that*

$$\Phi_{\underline{\hat{a}}, \underline{\hat{b}}, \underline{\mu}}(t) = \Phi_{\underline{a}, \underline{b}, \underline{\mu}}(st + \alpha).$$

By Lemma 3.33 we can introduce the following concept.

Definition 3.34. Let N be a 2×2 positive definite matrix with positive entries, and let $\underline{\mu} \in \mathbb{R}^2$ be a vector with positive entries. Let $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ be any representation of N , then let us define

$$\tilde{\Phi}_d(N, \underline{\mu}) = \max_{t \in [0, 2\pi]} \Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t).$$

The main theorem of this section is the following.

Theorem 3.35. *Let N be a 2×2 positive definite matrix with positive entries, and let $\underline{\mu} \in \mathbb{R}^2$ be a vector with positive entries. Then*

$$\Phi_d(N, \underline{\mu}) = \tilde{\Phi}_d(N, \underline{\mu}).$$

As a preparation for the proof we introduce some notations. We will also use the notions and tools from Sect. 3.4. The equation $\text{BP}(h) = h$ using the substitution $R = \frac{h_1}{h_2}$ becomes

$$R = \frac{\mu_1}{\mu_2} \left(\frac{N_{11}R + N_{12}}{N_{12}R + N_{22}} \right)^{d-1}.$$

Call $\mathcal{R}_{N, \underline{\mu}}$ the set of non-negative solutions of this equation. By a simple calculation we have

$$\begin{aligned} \overline{\mathbb{F}}(R, N, \underline{\mu}) &= \exp\left(\tilde{\Phi}_{N, \underline{\mu}, d}(h)\right) = \left(\sum_{\sigma} \mu_{\sigma} \left(\sum_{\sigma'} N_{\sigma, \sigma'} h_{\sigma'}\right)^d\right) \left(\sum_{\sigma, \sigma'} N_{\sigma, \sigma'} h_{\sigma} h_{\sigma'}\right)^{-\frac{d}{2}} \\ &= \mu_1 \left[\frac{N_{11}R + N_{12}}{\sqrt{N_{11}R^2 + 2N_{12}R + N_{22}}} \right]^d + \mu_2 \left[\frac{N_{12}R + N_{22}}{\sqrt{N_{11}R^2 + 2N_{12}R + N_{22}}} \right]^d. \end{aligned}$$

We know that

$$\Phi_d(N, \underline{\mu}) = \max_{R \in \mathcal{R}_{N, \underline{\mu}}} \overline{\mathbb{F}}(R, N, \underline{\mu})$$

Let us also choose a representation $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$, and let

$$R(t) = -\frac{b_2(t)}{b_1(t)} \quad \text{and} \quad S(t) = \frac{a_1(t)}{a_2(t)}.$$

Lemma 3.36. Let $a_1, a_2, b_1, b_2, \mu_1, \mu_2 \in \mathbb{R}$, then for every $t \in [0, 2\pi]$ such that $b_1(t), a_2(t) \neq 0$ we have

$$S(t) = \frac{N_{11}R(t) + N_{12}}{N_{12}R(t) + N_{22}}.$$

Furthermore,

$$|a_1(t)| = \frac{|N_{11}R(t) + N_{12}|}{\sqrt{N_{11}R(t)^2 + 2N_{12}R(t) + N_{22}}}$$

and

$$|a_2(t)| = \frac{|N_{12}R(t) + N_{22}|}{\sqrt{N_{11}R(t)^2 + 2N_{12}R(t) + N_{22}}},$$

and if t_0 maximizes $\Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t)$, then

$$R(t_0) = \frac{\mu_1}{\mu_2} \left(\frac{N_{11}R(t_0) + N_{12}}{N_{12}R(t_0) + N_{22}} \right)^{d-1}.$$

Proof. We have

$$\begin{aligned} \frac{N_{11}R(t) + N_{12}}{N_{12}R(t) + N_{22}} &= \frac{N_{11} \left(-\frac{b_2(t)}{b_1(t)} \right) + N_{12}}{N_{12} \left(-\frac{b_2(t)}{b_1(t)} \right) + N_{22}} = \frac{-N_{11}b_2(t) + N_{12}b_1(t)}{-N_{12}b_2(t) + N_{22}b_1(t)} \\ &= \frac{(a_1^2 + b_1^2)(a_2 \sin(t) - b_2 \cos(t)) + (a_1a_2 + b_1b_2)(-a_1 \cos(t) + b_1 \sin(t))}{(a_1a_2 + b_1b_2)(a_2 \sin(t) - b_2 \cos(t)) + (a_2^2 + b_2^2)(-a_1 \cos(t) + b_1 \sin(t))} \\ &= \frac{b_1(a_2b_1 - a_1b_2) \sin(t) + a_1(a_2b_1 - a_1b_2) \cos(t)}{b_2(a_2b_1 - a_1b_2) \sin(t) + a_2(a_2b_1 - a_1b_2) \cos(t)} \\ &= \frac{b_1 \sin(t) + a_1 \cos(t)}{b_2 \sin(t) + a_2 \cos(t)} \\ &= \frac{a_1(t)}{a_2(t)} \\ &= S(t) \end{aligned}$$

Note that we have $(a_2b_1 - a_1b_2)^2 = (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1a_2 + b_1b_2)^2 = N_{11}N_{22} - N_{12}^2 \neq 0$. Next let us prove that

$$|a_1(t)| = \frac{|N_{11}R(t) + N_{12}|}{\sqrt{N_{11}R(t)^2 + 2N_{12}R(t) + N_{22}}}.$$

Let us multiply both sides with the denominator of the right hand side, and take the square of both sides. Note that by Lemma 3.9 and the decomposition $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ we have $N_{11} = a_1(t)^2 + b_1(t)^2$, $N_{12} = a_1(t)a_2(t) + b_1(t)b_2(t)$ and $N_{22} = a_2(t)^2 + b_2(t)^2$ are true for every t . For ease of notation let

$$\text{LHS} = a_1(t)^2(N_{11}R(t)^2 + 2N_{12}R(t) + N_{22}).$$

Then

$$\begin{aligned}
 \text{LHS} &= a_1(t)^2(N_{11}R(t)^2 + 2N_{12}R(t) + N_{22}) \\
 &= a_1(t)^2 \left((a_1(t)^2 + b_1(t)^2) \frac{b_2(t)^2}{b_1(t)^2} - 2(a_1(t)a_2(t) \right. \\
 &\quad \left. + b_1(t)b_2(t)) \frac{b_2(t)}{b_1(t)} + (a_2(t)^2 + b_2(t)^2) \right) \\
 &= a_1(t)^4 \frac{b_2(t)^2}{b_1(t)^2} - 2a_1(t)^3 a_2(t) \frac{b_2(t)}{b_1(t)} + a_1(t)^2 a_2(t)^2 \\
 &= \left(-a_1(t)^2 \frac{b_2(t)}{b_1(t)} + a_1(t)a_2(t) \right)^2 \\
 &= \left(-(a_1(t)^2 + b_1(t)^2) \frac{b_2(t)}{b_1(t)} + (a_1(t)a_2(t) + b_1(t)b_2(t)) \right)^2 \\
 &= (N_{11}R(t) + N_{22})^2
 \end{aligned}$$

The proof of the third identity follows similarly and we omit it.

If t_0 maximizes $\Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t)$, in fact, we only need $\Phi'_{\underline{a}, \underline{b}, \underline{\mu}}(t_0) = 0$, then we have

$$R(t_0) = \frac{\mu_1}{\mu_2} S(t_0)^{d-1} = \frac{\mu_1}{\mu_2} \left(\frac{N_{11}R(t_0) + N_{12}}{N_{12}R(t_0) + N_{22}} \right)^{d-1}.$$

□

Now we are ready to prove Theorem 3.35.

Proof of Theorem 3.35. Note that $\tilde{\Phi}_d(N, \underline{\mu})$ does not depend on which representation $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ we choose so we can assume by part (ii) of Lemma 3.2 that $a_1, a_2, b_1, b_2 > 0$. Then

$$\Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t) = \mu_1(a_1 \cos(t) + b_1 \sin(t))^d + \mu_2(a_2 \cos(t) + b_2 \sin(t))^d$$

is maximized at some $t_0 \in [0, \frac{\pi}{2}]$. Then $S(t_0) > 0$, and by $R(t_0) = \frac{\mu_1}{\mu_2} S(t_0)^{d-1}$ we get that $R(t_0) > 0$. Thus $R(t_0) \in \mathcal{R}_{N, \underline{\mu}}$ and we have

$$\tilde{\Phi}_d(N, \underline{\mu}) = \Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t_0) = \overline{\mathbb{F}}(R(t_0), N, \underline{\mu}) \leq \max_{R \in \mathcal{R}_{N, \underline{\mu}}} \overline{\mathbb{F}}(R, N, \underline{\mu}) = \Phi_d(N, \underline{\mu}).$$

On the other hand, $\Phi_d(N, \underline{\mu}) = \max_{R \in \mathcal{R}_{N, \underline{\mu}}} \overline{\mathbb{F}}(R, N, \underline{\mu}) = \overline{\mathbb{F}}(R_0, N, \underline{\mu})$ for some R_0 . Then by Lemma 3.8 there exists a $t' \in [0, 2\pi]$ such that $R(t') = R_0$. Note that from

$$\frac{a_1(t')}{a_2(t')} = S(t') = \frac{N_{11}R_0 + N_{12}}{N_{12}R_0 + N_{22}} > 0,$$

we get that $a_1(t'), a_2(t')$ have the same sign. By changing t' to $t' + \pi$ if necessary we can ensure that they are both positive. Hence we have

$$a_1(t') = \frac{N_{11}R_0 + N_{12}}{\sqrt{N_{11}R_0^2 + 2N_{12}R_0 + N_{22}}} \text{ and } a_2(t') = \frac{N_{12}R_0 + N_{22}}{\sqrt{N_{11}R_0^2 + 2N_{12}R_0 + N_{22}}}.$$

Then

$$\Phi_d(N, \underline{\mu}) = \overline{\mathbb{F}}(R_0, N, \underline{\mu}) = \Phi_{a,b,\underline{\mu}}(t') \leq \max_{t \in [0, 2\pi]} \Phi_{a,b,\underline{\mu}}(t) = \tilde{\Phi}_d(N, \underline{\mu}).$$

□

We end this section with a lemma that we will use later.

Lemma 3.37. *Let Q be a 2×2 matrix with positive entries and positive determinant, and let k be an integer. Then the equation*

$$R = \left(\frac{Q_{11}R + Q_{12}}{Q_{21}R + Q_{22}} \right)^k$$

has at most 3 non-negative solutions.

Proof. Let

$$f(R) = \left(\frac{Q_{11}R + Q_{12}}{Q_{21}R + Q_{22}} \right)^k - R.$$

Then $\frac{\partial^2}{\partial R^2} f$ is given as

$$\left(\frac{Q_{11}R + Q_{12}}{Q_{21}R + Q_{22}} \right)^k \frac{k(Q_{11}Q_{22} - Q_{12}Q_{21})((k-1)Q_{11}Q_{22} - (k+1)Q_{12}Q_{21} - 2Q_{11}Q_{21}R)}{(Q_{11}R + Q_{12})^2(Q_{21}R + Q_{22})^2}.$$

Let

$$R^* = \frac{(k-1)Q_{11}Q_{22} - (k+1)Q_{12}Q_{21}}{2Q_{11}Q_{21}}.$$

If $R^* \leq 0$, then f is concave of $[0, \infty)$ and so it has at most two solutions.

If $R^* > 0$, then $\frac{\partial^2}{\partial R^2} f$ is positive if $0 \leq R < R^*$, and negative if $R > R^*$. If $f(R^*) > 0$, then f has at most 2 solutions on $(0, R^*)$ and has at most 1 solution on (R^*, ∞) . If $f(R^*) < 0$, then f has at most 1 solution on $(0, R^*)$ and has at most 2 solutions on (R^*, ∞) . Finally, if $f(R^*) = 0$, then f has at most 1 solution on $(0, R^*)$ and has at most 1 solutions on (R^*, ∞) . So in all cases it has at most 3 solutions. □

3.7. The vector $\underline{v}(t_0)$. Let t_0 be the maximizer of the function $\Phi_{\underline{a}, \underline{b}, \underline{\mu}}$. In this section we study the vector $\underline{v}(t_0) = (r_0(t_0), r_1(t_0), \dots, r_d(t_0))$. First we need a simple lemma.

Lemma 3.38. *Let $\underline{a}, \underline{b}, \underline{\mu} \in \mathbb{R}^r$. For $j = 1, \dots, r$ let*

$$a_j(t) = a_j \cos(t) + b_j \sin(t) \quad \text{and} \quad b_j(t) = -a_j \sin(t) + b_j \cos(t).$$

Let $r_j(t) = \sum_{k=1}^r \mu_k a_k(t)^{d-j} b_k(t)^j$, then

$$r_0(t) = \sum_{j=0}^d \binom{d}{j} r_j(0) \cos(t)^{d-j} \sin(t)^j.$$

Proof. We have

$$\begin{aligned}
 r_0(t) &= \sum_{k=1}^r \mu_k a_k(t)^d \\
 &= \sum_{k=1}^r \mu_k (a_k \cos(t) + b_k \sin(t))^d \\
 &= \sum_{k=1}^r \mu_k \sum_{j=0}^d \binom{d}{j} a_k^{d-j} b_k^j \cos(t)^{d-j} \sin(t)^j \\
 &= \sum_{j=0}^d \binom{d}{j} \cos(t)^{d-j} \sin(t)^j \left(\sum_{k=1}^r \mu_k a_k^{d-j} b_k^j \right) \\
 &= \sum_{j=0}^d \binom{d}{j} r_j(0) \cos(t)^{d-j} \sin(t)^j
 \end{aligned}$$

□

Consider the vector $\underline{v}(t_0) = (r_0(t_0), r_1(t_0), \dots, r_d(t_0))$. We show that $r_1(t_0) = 0$ and $r_j(t_0) \geq 0$ if j is even, and the numbers $r_j(t_0)$ have the same sign for odd $j \geq 3$. This follows from the following more general lemma.

Lemma 3.39. *Let $\mu_1, \mu_2 > 0$ and $a_1, a_2, b_1, b_2 \in \mathbb{R}$ such that $a_1, a_2, b_1, b_2 > 0$. Let $t_0 \in [0, \frac{\pi}{2}]$ be the maximizer of $\Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t) = \mu_1 a_1(t)^d + \mu_2 a_2(t)^d$. Let*

$$r_j(t) = \mu_1 a_1(t)^{d-j} b_1(t)^j + \mu_2 a_2(t)^{d-j} b_2(t)^j.$$

Then $r_1(t_0) = 0$ and either

- (i) $r_j(t_0) \geq 0$ for $j = 0, \dots, d$ or
- (ii) $r_j(t_0) \geq 0$ for even j , and $r_j(t_0) \leq 0$ for odd j .

Proof. Observe that $\frac{\partial}{\partial t} a_1(t) = b_1(t)$ and $\frac{\partial}{\partial t} a_2(t) = b_2(t)$ and so

$$\frac{\partial}{\partial t} r_0(t) = d\mu_1 a_1(t)^{d-1} b_1(t) + d\mu_2 a_2(t)^{d-1} b_2(t) = dr_1(t).$$

Hence if t_0 maximizes $r_0(t)$, then $r_1(t_0) = 0$.

To prove the inequalities we need to study several cases. First of all, $r_1(t_0) = 0$ implies that $a_1(t_0)b_1(t_0) = 0$ if and only if $a_2(t_0)b_2(t_0) = 0$. If $a_1(t_0)b_1(t_0) = a_2(t_0)b_2(t_0) = 0$, then $r_1(t_0) = r_2(t_0) = \dots = r_{d-1}(t_0) = 0$. We know that $r_0(t_0) \geq r_0(0) = \mu_1 a_1^d + \mu_2 a_2^d > 0$. Finally, $r_d(t_0) = \mu_1 b_1(t_0)^d + \mu_2 b_2(t_0)^d \geq 0$ if d is even. If d is odd, then no matter what the sign of $r_d(t_0)$ case (i) or (ii) is satisfied.

So we can assume that $a_1(t_0)b_1(t_0) \neq 0$ and $a_2(t_0)b_2(t_0) \neq 0$. By symmetry we can assume that $|\mu_2 a_2(t_0)^d| \geq |\mu_1 a_1(t_0)^d|$. Note that

$$r_j(t_0) = \mu_1 a_1(t_0)^d \left(\frac{b_1(t_0)}{a_1(t_0)} \right)^j \left(1 + \frac{\mu_2 a_2(t_0)^d}{\mu_1 a_1(t_0)^d} \left(\frac{a_1(t_0)b_2(t_0)}{a_2(t_0)b_1(t_0)} \right)^j \right).$$

From $r_1(t_0) = 0$ we get that

$$\left| \frac{a_1(t_0)b_2(t_0)}{a_2(t_0)b_1(t_0)} \right| = \left| \frac{\mu_1 a_1(t_0)^d}{\mu_2 a_2(t_0)^d} \right| \leq 1.$$

Note that if $\mu_2 a_2(t_0)^d = \mu_1 a_1(t_0)^d > 0$, then $\frac{a_1(t_0)b_2(t_0)}{a_2(t_0)b_1(t_0)} = -1$. Then for odd j we have $r_j(t_0) = 0$, and for even j all terms are positive in the above product, so $r_j(t_0) > 0$.

If $\mu_2 a_2^d > \mu_1 a_1^d$, then $\left| \frac{a_1(t_0)b_2(t_0)}{a_2(t_0)b_1(t_0)} \right| < 1$, and so the last term is positive for all $j \geq 2$. Then if $\frac{b_1(t_0)}{a_1(t_0)} > 0$, then $r_j(t_0) \geq 0$ for all j , and if $\frac{b_1(t_0)}{a_1(t_0)} < 0$, then $r_j(t)$ is positive for even j and negative for odd $j \geq 3$. We are done. \square

Remark 3.40. Note that besides $r_1(t_0) = 0$ we also have $\frac{r_2(t_0)}{r_0(t_0)} \leq \frac{1}{d-1}$ since $\frac{\partial^2}{\partial t^2} r_0(t) = d((d-1)r_2(t) - r_0(t))$ should be non-positive at $t = t_0$.

The following theorem is a strengthening of Theorem 3.25 with a new proof.

Theorem 3.41. *Let N be a 2×2 positive definite matrix with positive entries and let $\underline{\mu} \in \mathbb{R}_{>0}^2$. For any d -regular graph G we have $Z_G(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)}$. Furthermore, if G contains $\varepsilon v(G)$ cycles of length at most g , then there exists a $\delta = \delta(d, N, \underline{\mu}, \varepsilon, g) > 0$ such that $Z_G(N, \underline{\mu}) \geq ((1 + \delta)\Phi_d(N, \underline{\mu}))^{v(G)}$.*

The proof presented below is strongly inspired by the work of Chertkov and Chernyak [10] on loop series and gauge transformation. The paper of Borbényi and Csikvári [5] contains a similar proof about the number of Eulerian orientations in regular graphs.

Proof. By part (ii) of Lemma 3.2 we can choose $\underline{a}, \underline{b} \in \mathbb{R}^2$ such that $a_1, a_2, b_1, b_2 > 0$. Then $r_j(0) = \mu_1 a_1^{d-j} b_1^j + \mu_2 a_2^{d-j} b_2^j > 0$ for all $j \in \{0, 1, \dots, d\}$. This implies that

$$\Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t) = r_0(t) = \sum_{j=0}^d \binom{d}{j} r_j(0) \cos(t)^{d-j} \sin(t)^j$$

has a maximizer t_0 in the interval $[0, \frac{\pi}{2}]$. Indeed, for any $t \in [0, 2\pi]$ there is a $t' \in [0, \frac{\pi}{2}]$ such that $\cos(t') = |\cos(t)|$ and $\sin(t') = |\sin(t)|$, so $|\Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t)| \leq \Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t')$. Thus the conditions of Lemma 3.39 are satisfied. We have

$$Z_G(N, \underline{\mu}) = F_G(r_0(t_0), \dots, r_d(t_0)) = \sum_{A \subseteq E(G)} \prod_{v \in V} r_{d_A(v)}(t_0).$$

For each $A \subseteq E(G)$ the number of vertices with odd $d_A(v)$ is even, so by Lemma 3.39 each term in the sum is non-negative. Then taking $A = \emptyset$ we get that

$$Z_G(N, \underline{\mu}) \geq r_0(t_0)^{v(G)} = \Phi_d(N, \underline{\mu})^{v(G)}.$$

This completes the proof of the first part.

To prove the second part first observe that $r_2(t_0) > 0$. Indeed, since $a_1, a_2, b_1, b_2 > 0$ and $t_0 \in [0, \frac{\pi}{2}]$ we get that $a_1(t_0), a_2(t_0) > 0$ and thus

$$r_2(t_0) = \mu_1 a_1(t_0)^{d-2} b_1(t_0)^2 + \mu_2 a_2(t_0)^{d-2} b_2(t_0)^2 = 0$$

would imply that $b_1(t_0) = b_2(t_0) = 0$ which then implies that

$$N_{11}N_{22} - N_{12}^2 = (a_1(t_0)b_2(t_0) - a_2(t_0)b_1(t_0))^2 = 0$$

contradicting the positive definiteness of N . Also observe that if G contains $\varepsilon v(G)$ cycles of length at most g , then it also contains $\varepsilon' v(G)$ vertex-disjoint cycles of length at most g for some ε' depending on d, g and ε , but not depending on $v(G)$. Then

$$Z_G(N, \underline{\mu}) \geq \Phi_d(N, \underline{\mu})^{v(G)} \left(1 + \left(\frac{r_2(t_0)}{r_0(t_0)} \right)^g \right)^{\varepsilon' v(G)}.$$

Indeed, we can consider those sets $A \subseteq E(G)$ that consists of the union of some of the vertex-disjoint cycles of length at most g . Here we also use the fact that $0 \leq \frac{r_2(t_0)}{r_0(t_0)} \leq \frac{1}{d-1} \leq 1$ by Remark 3.40. Hence $1 + \delta = \left(1 + \left(\frac{r_2(t_0)}{r_0(t_0)} \right)^g \right)^{\varepsilon'}$ satisfies the claim of the theorem. \square

Remark 3.42. Theorem 3.41 implies that if $(G_n)_n$ is a sequence of d -regular graphs such that it is not essentially large girth, then

$$\limsup_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(N, \underline{\mu}) > \ln \Phi_d(N, \underline{\mu}).$$

Since $Z_G(q, w) \geq Z_G^{(2)}(q, w)$ for $q \geq 2$ and $w \geq 0$ this kind of stability statement is also true for $Z_G(q, w)$.

3.8. Mixed state. In this section we introduce a concept that is strongly related to the phase transition of the random cluster model. We will see that there exists a $w_c = w_c(q)$ such that if $0 \leq w \leq w_c$, then $\Phi_{d,q,w} = q \left(1 + \frac{w}{q} \right)^{d/2}$, and if $w > w_c$, then $\Phi_{d,q,w} > q \left(1 + \frac{w}{q} \right)^{d/2}$. The problem with such a statement is that it depends on the parametrization (q, w) , for a general model $(N, \underline{\mu})$ it does not make sense. On the other hand, there is a concept that makes sense even for general $(N, \underline{\mu})$, where N is a 2×2 positive definite matrix.

Definition 3.43. We say that $(N, \underline{\mu})$ exhibits a mixed state for a fixed positive integer d if $R_c(N, \underline{\mu}) = 1$.

Note that $R_c(N, \underline{\mu}) = 1$ does not depend on which representation $N = \underline{a}\underline{a}^T + \underline{b}\underline{b}^T$ we choose. We also know that $R = R_c(N, \underline{\mu})$ is a solution of

$$(N_{11}N_{22} - N_{12}^2)R^4 + (-N_{22}^2T + 2N_{12}^2 - N_{11}^2T^{-1})R^2 + (N_{11}N_{22} - N_{12}^2) = 0,$$

where $T = \left(\frac{\mu_2}{\mu_1} \right)^{2/d}$. This shows that $(N, \underline{\mu})$ exhibits a mixed state for d if

$$2(N_{11}N_{22} - N_{12}^2) - (N_{22}^2T - 2N_{12}^2 + N_{11}^2T^{-1}) = 0.$$

The main lemma of this section is the following.

Lemma 3.44. *Let $N = \underline{aa}^T + \underline{bb}^T$ for some $a_1, a_2, b_1, b_2 \in \mathbb{R}$ and let $\mu_1, \mu_2 > 0$. Suppose that $(N, \underline{\mu})$ exhibits a mixed state for d , that is, for some t_1 we have*

$$a_1(t_1) = \left(\frac{\mu_2}{\mu_1}\right)^{1/d} (-b_2(t_1)) \text{ and } -b_1(t_1) = \left(\frac{\mu_2}{\mu_1}\right)^{1/d} a_2(t_1).$$

Let $t_2 = 2t_1 - \frac{\pi}{2}$. Then for every $t \in \mathbb{R}$ we have

$$\mu_1 a_1(t)^d = \mu_2 a_2(t_2 - t)^d \text{ and } \mu_2 a_2(t)^d = \mu_1 a_1(t_2 - t)^d.$$

In particular,

$$\Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t) = \Phi_{\underline{a}, \underline{b}, \underline{\mu}}(t_2 - t).$$

Proof. Note that for any $u_1, u_2 \in \mathbb{R}$ we have

$$a_1(u_1 + u_2) = a_1(u_1) \cos(u_2) + b_1(u_1) \sin(u_2).$$

To prove $\mu_1 a_1(t)^d = \mu_2 a_2(t_2 - t)^d$ it will be more convenient to prove the statement in the form

$$a_2(t_1 - t) = \left(\frac{\mu_1}{\mu_2}\right)^{1/d} a_1\left(t + t_1 - \frac{\pi}{2}\right).$$

This is indeed true,

$$\begin{aligned} a_2(t_1 - t) &= a_2(t_1) \cos(-t) + b_2(t_1) \sin(-t) \\ &= a_2(t_1) \cos(t) - b_2(t_1) \sin(t) \\ &= \left(\frac{\mu_1}{\mu_2}\right)^{1/d} (-b_1(t_1) \cos(t) + a_1(t_1) \sin(t)) \\ &= \left(\frac{\mu_1}{\mu_2}\right)^{1/d} \left(a_1(t_1) \cos\left(t - \frac{\pi}{2}\right) + b_1(t_1) \sin\left(t - \frac{\pi}{2}\right)\right) \\ &= \left(\frac{\mu_1}{\mu_2}\right)^{1/d} a_1\left(t + t_1 - \frac{\pi}{2}\right) \end{aligned}$$

By symmetry the other claim is also true. □

3.9. Specialization to $N = M'_2$ and $\underline{\mu} = \underline{v}_2$. In this section we collected some results that are specialized to $N = M'_2$ and $\underline{\mu} = \underline{v}_2$. In particular, we will choose $a_1 = a_2 = \sqrt{1 + \frac{w}{q}}$, $b_1 = \sqrt{\frac{(q-1)w}{q}}$ and $b_2 = -\sqrt{\frac{w}{q(q-1)}}$.

Lemma 3.45. *Let $q \geq 2, w \geq 0$. Let $a_1 = a_2 = \sqrt{1 + \frac{w}{q}}$, $b_1 = \sqrt{\frac{(q-1)w}{q}}$, $b_2 = -\sqrt{\frac{w}{q(q-1)}}$, $v_1 = 1$ and $v_2 = q - 1$. Let $r_j(0) = v_1 a_1^{d-j} b_1^j + v_2 a_2^{d-j} b_2^j$. Then we have $r_j(0) \geq 0$ for $j = 0, 1, \dots, d$ and $r_1(0) = 0$.*

Proof. We have

$$\begin{aligned} r_j(0) &= \left(1 + \frac{w}{q}\right)^{(d-j)/2} \left(\frac{(q-1)w}{q}\right)^{j/2} + (q-1)(-1)^j \left(1 + \frac{w}{q}\right)^{(d-j)/2} \left(\frac{w}{q(q-1)}\right)^{j/2} \\ &= \left(1 + \frac{w}{q}\right)^{(d-j)/2} \left(\frac{w}{q(q-1)}\right)^{j/2} ((q-1)^j + (-1)^j(q-1)) \end{aligned}$$

This is 0 if $j = 1$, and positive if $j \neq 1$. □

Recall that

$$\Phi_{d,q,w} = \max_{t \in [0, 2\pi]} \Phi_{d,q,w}(t).$$

The next lemma shows that it is enough to consider the interval $[0, \frac{\pi}{2}]$ to find the maximum when $q \geq 2$ and $w \geq 0$.

Lemma 3.46. *If $q \geq 2$ and $w \geq 0$, then there is a $t_0 \in [0, \frac{\pi}{2}]$ for which $\Phi_{d,q,w} = \Phi_{d,q,w}(t_0)$.*

Proof. By Lemma 3.38 we have

$$\Phi_{d,q,w}(t) = r_0(t) = \sum_{j=0}^d \binom{d}{j} r_j(0) \cos(t)^{d-j} \sin(t)^j.$$

By Lemma 3.45 we have $r_j(0) \geq 0$ for all $j \in \{0, 1, \dots, d\}$ if $q \geq 2$ and $w \geq 0$. For any $t \in [0, 2\pi]$ there is a $t' \in [0, \frac{\pi}{2}]$ such that $|\cos(t)| = \cos(t')$ and $|\sin(t)| = \sin(t')$, thus

$$|\Phi_{d,q,w}(t)| \leq \sum_{j=0}^d \binom{d}{j} r_j(0) |\cos(t)|^{d-j} |\sin(t)|^j = \Phi_{d,q,w}(t').$$

Hence $\max_{t \in [0, 2\pi]} \Phi_{d,q,w}(t) = \max_{t \in [0, \frac{\pi}{2}]} \Phi_{d,q,w}(t)$. □

The next lemma will be useful to get even more precise bounds on $\tan(t_0)$.

Lemma 3.47. *Let $q \geq 2$ and $w \geq 0$. Let $\bar{t} \in [0, \frac{\pi}{2}]$ such that*

$$\left. \frac{\partial}{\partial t} \Phi_{d,q,w}(t) \right|_{t=\bar{t}} = 0.$$

Then we have $a_{q,w,1}(\bar{t}), a_{q,w,2}(\bar{t}), b_{q,w,1}(\bar{t}) > 0$ and $b_{q,w,2}(\bar{t}) < 0$. In particular, this is true if $\bar{t} = t_0$ maximizing the function $\Phi_{d,q,w}(t)$ in the interval $[0, \frac{\pi}{2}]$.

Proof. Note that $\Phi_{d,q,w}(t) = a_{q,w,1}(t)^d + (q-1)a_{q,w,2}(t)^d$, and its derivative is

$$\frac{\partial}{\partial t} \Phi_{d,q,w}(t) = d(b_{q,w,1}(t)a_{q,w,1}(t)^{d-1} + b_{q,w,2}(t)a_{q,w,2}(t)^{d-1}).$$

Note that $a_{q,w,1}(t) > 0$ and $b_{q,w,2}(t) < 0$ for all $t \in [0, \frac{\pi}{2}]$. Suppose for contradiction that $b_{q,w,1}(\bar{t}) < 0$. Then from $b_{q,w,1}(\bar{t})a_{q,w,1}(\bar{t})^{d-1} + b_{q,w,2}(\bar{t})a_{q,w,2}(\bar{t})^{d-1} = 0$ we also get that $a_{q,w,2}(\bar{t})^{d-1} < 0$, that is, $a_{q,w,2}(\bar{t}) < 0$ and d is even. Then

$$\left(\frac{a_{q,w,1}(\bar{t})}{-a_{q,w,2}(\bar{t})} \right)^{d-1} = \frac{(q-1)(-b_{q,w,2}(\bar{t}))}{-b_{q,w,1}(\bar{t})}.$$

Note that $a_{q,w,1}(t) > -a_{q,w,2}(t)$ for all $t \in [0, \frac{\pi}{2}]$, and so

$$\left(\frac{a_{q,w,1}(\bar{t})}{-a_{q,w,2}(\bar{t})}\right)^{d-1} > \frac{a_{q,w,1}(\bar{t})}{-a_{q,w,2}(\bar{t})}.$$

By $a_{q,w,1}(t)b_{q,w,1}(t) + (q - 1)a_{q,w,2}(t)b_{q,w,2}(t) = -q \cos(t) \sin(t)$ we have

$$\frac{a_{q,w,1}(\bar{t})}{-a_{q,w,2}(\bar{t})} > \frac{(q - 1)(-b_{q,w,2}(\bar{t}))}{-b_{q,w,1}(\bar{t})},$$

but then

$$\left(\frac{a_{q,w,1}(\bar{t})}{-a_{q,w,2}(\bar{t})}\right)^{d-1} > \frac{a_{q,w,1}(\bar{t})}{-a_{q,w,2}(\bar{t})} > \frac{(q - 1)(-b_{q,w,2}(\bar{t}))}{-b_{q,w,1}(\bar{t})}$$

leads to a contradiction. Hence $b_{q,w,1}(\bar{t}) > 0$. But then $a_{q,w,1}(\bar{t})a_{q,w,2}(\bar{t}) + b_{q,w,1}(\bar{t})b_{q,w,2}(\bar{t}) = 1$ implies that $a_{q,w,2}(\bar{t}) > 0$. \square

Finally, we collected some claims about the derivatives of $\Phi_{d,q,w}(t)$ at $t = 0$.

Lemma 3.48. *We have*

$$\frac{\partial}{\partial t} \Phi_{d,q,w}(t) \Big|_{t=0} = 0 \quad \text{and} \quad \frac{\partial^2}{\partial t^2} \Phi_{d,q,w}(t) \Big|_{t=0} = \left(1 + \frac{w}{q}\right)^{d/2-1} ((d - 2)w - q).$$

In particular, if $w < \frac{q}{d-2}$, then the function $\Phi_{d,q,w}(t)$ has a local maximum at $t = 0$, and if $w > \frac{q}{d-2}$ the function $\Phi_{d,q,w}(t)$ has a local minimum at $t = 0$.

3.9.1. Mixed state and phase transition. In this section we discuss the mixed state and phase transition of $Z_G^{(2)}(q, w)$.

We know that $(N, \underline{\mu})$ exhibits mixed state for some d if

$$2(N_{11}N_{22} - N_{12}^2) - (N_{22}^2T - 2N_{12}^2 + N_{11}^2T^{-1}) = 0,$$

where $T = \left(\frac{\mu_2}{\mu_1}\right)^{1/d}$. Applying this equation to $(M'_2, \underline{\nu}_2)$ we get that

$$\begin{aligned} & 2\left((1+w)\left(1 + \frac{w}{q-1}\right) - 1\right) - \left(\left(1 + \frac{w}{q-1}\right)^2 (q-1)^{2/d} - 2 + (1+w)^2 (q-1)^{-2/d}\right) \\ & = -(((q-1)^{-1/d} - (q-1)^{1/d-1})w - ((q-1)^{1/d} - (q-1)^{-1/d}))^2. \end{aligned}$$

Note that if $q = 2$, then this is constant 0, so in the case of $q = 2$, the spin system $(M'_2, \underline{\nu}_2)$ always exhibits mixed state. If $q \neq 2$, then the solution $w_c = w_c(q)$ of this equation is

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

Note that by L'Hôpital's rule we have

$$\lim_{q \rightarrow 2^+} w_c(q) = \frac{2}{d-2},$$

so we will define $w_c(2) = \frac{2}{d-2}$ even though the spin system $(M'_2, \underline{\nu}_2)$ itself always exhibits mixed state for every w if $q = 2$.

The main theorem of this section is the following. It asserts that the mixed state also describes a phase transition in the value of $\Phi_{d,q,w}$.

Theorem 3.49. *Let $q \geq 2$. If $0 \leq w \leq w_c$, then $\Phi_{d,q,w} = q \left(1 + \frac{w}{q}\right)^{d/2}$. If $w > w_c$, then $\Phi_{d,q,w} > q \left(1 + \frac{w}{q}\right)^{d/2}$.*

Before we start to prove Theorem 3.49 we need a lemma about the curve $(q, w_c(q))$. For a visualization of this lemma see the dashed curve on Fig. 4.

Lemma 3.50. *For every $q \geq 2$ let $x(q) = 1 + \frac{q}{w_c(q)}$ and $y(q) = 1 + w_c(q)$. Then the curve $(x(q), y(q))$ on the (x, y) plane is the graph of a monotone increasing function. Furthermore, $w_c(q) \leq \frac{q}{d-2}$.*

Proof. We have

$$y = w_c(q) + 1 = \frac{q - 2}{(q - 1)^{1-2/d} - 1},$$

and $q = (x(q) - 1)(y(q) - 1)$. Thus

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial q} \cdot \frac{\partial q}{\partial x} = \frac{(q - 1)^{1-2/d} - 1 - (1 - 2/d)(q - 1)^{-2/d}(q - 2)}{((q - 1)^{1-2/d} - 1)^2} \cdot (y - 1).$$

Clearly, $y - 1 = w > 0$ so we only need to show that the first term is also positive. We can rewrite the numerator as

$$\frac{2}{d}(q - 1)^{-2/d}(q - 2) + (q - 1)^{-2/d} - 1 = (q - 1)^{-2/d} \left(\frac{2}{d}(q - 2) + 1 - (q - 1)^{2/d} \right).$$

This is 0 if $q = 2$ and its derivative is $\frac{2}{d}(1 - (q - 1)^{-2/d}) \geq 0$ for $q \geq 2$. Hence $\frac{dy}{dx} > 0$.

The second part follows from the first part since if we follow any hyperbola $(x - 1)(y - 1) = q$ while decreasing x , and hence increasing y , it intersects the curve $(x(q), y(q))$ before hitting the line $x = d - 1$. The intersection of the hyperbola $(x - 1)(y - 1) = q$ and the line $x = d - 1$ is at $y = 1 + \frac{q}{d-2}$, thus $w_c(q) \leq \frac{q}{d-2}$.

We decompose the proof of Theorem 3.49 into three propositions dealing with $w = w_c$, $0 \leq w < w_c$ and $w > w_c$.

Proposition 3.51. *For $q \geq 2$ and $w = w_c$ we have $\Phi_{d,q,w} = q \left(1 + \frac{w}{q}\right)^{d/2}$.*

Proof. Let us assume that $q > 2$, the statement for $q = 2$ follows by continuity. We show that for $w = w_c$ the function $\Phi_{d,q,w}(t)$ has a global maximizer in $(0, \pi/2)$ with value $\Phi_{d,q,w}(0) = q \left(1 + \frac{w}{q}\right)^{d/2}$. We know that for $w = w_c$ there is a $t_1 \in [0, \frac{\pi}{2}]$ such that

$$\frac{a_{q,w,2}(t_1)}{-b_{q,w,1}(t_1)} = \left(\frac{v_1}{v_2}\right)^{1/d}.$$

This means that

$$\frac{\sqrt{1 + \frac{w}{q}} \cos(t_1) - \sqrt{\frac{w}{q(q-1)}} \sin(t_1)}{\sqrt{1 + \frac{w}{q}} \sin(t_1) - \sqrt{\frac{(q-1)w}{q}} \cos(t_1)} = (q - 1)^{-1/d} \leq 1.$$

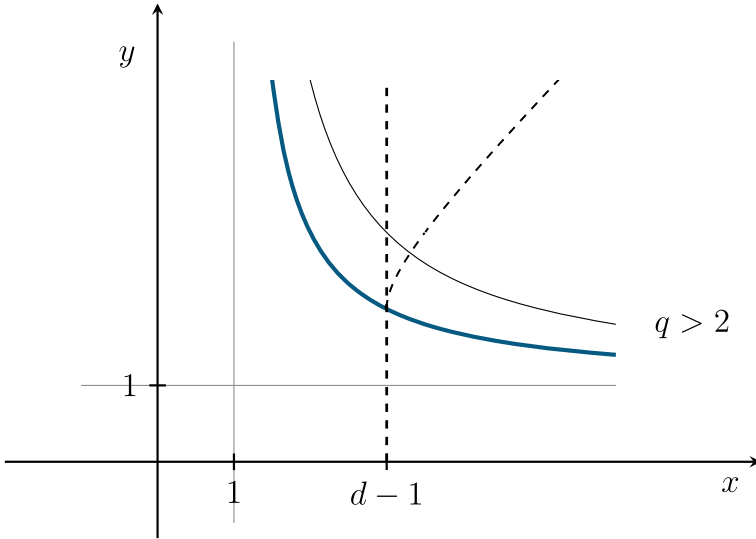


Fig. 4. A generic hyperbola intersecting the line $x = d - 1$ and curve of the phase transition

Note that if $q = 2$, then $t_1 = \frac{\pi}{4}$ is a solution. If $q > 2$, then

$$\frac{\sin(t_1)}{\cos(t_1)} \geq \frac{\sqrt{1 + \frac{w}{q}} + \sqrt{\frac{(q-1)w}{q}}}{\sqrt{1 + \frac{w}{q}} + \sqrt{\frac{w}{q(q-1)}}} > 1$$

showing that $t_1 \in (\frac{\pi}{4}, \frac{\pi}{2})$. Let $t_2 = 2t_1 - \frac{\pi}{2} \in (0, \frac{\pi}{2})$. By Lemma 3.44 we know that $\Phi_{d,q,w_c}(t) = \Phi_{d,q,w_c}(t_2 - t)$. (For an example of the graph of a function $\Phi_{d,q,w_c}(t)$ see Fig. 5)

We know that

$$\left. \frac{\partial}{\partial t} \Phi_{d,q,w_c}(t) \right|_{t=0} = 0.$$

This immediately implies that

$$\left. \frac{\partial}{\partial t} \Phi_{d,q,w_c}(t) \right|_{t=t_2} = 0.$$

The equation $\Phi_{d,q,w_c}(t) = \Phi_{d,q,w_c}(t_2 - t)$ also implies that

$$\left. \frac{\partial}{\partial t} \Phi_{d,q,w_c}(t) \right|_{t=t_2/2} = 0.$$

This means that a computation similar to the one in Sect. 3.6 gives that if

$$R(t) = -(q - 1) \frac{b_{q,w,2}(t)}{b_{q,w,1}(t)} \quad \text{and} \quad S(t) = \frac{a_{q,w,1}(t)}{a_{q,w,2}(t)} = \frac{(1 + w)R(t) + q - 1}{R(t) + q + w - 1},$$

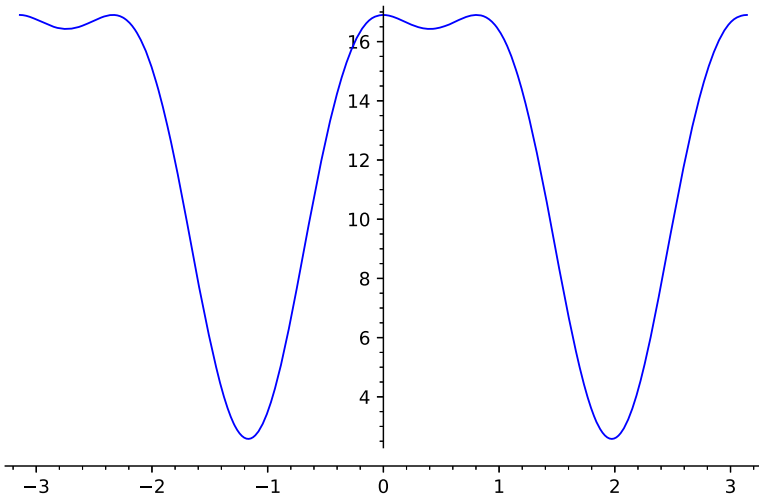


Fig. 5. For $d = 4$ and $q = 10$ we have $w_c = 3$. The graph of the trigonometric polynomial $\Phi_{4,10,3}(t)$ is depicted in the figure

then the values $R = R(0), R(\frac{t_2}{2}), R(t_2)$ are all solutions of the equation

$$R = \left(\frac{(1+w)R + q - 1}{R + w + q - 1} \right)^{d-1}.$$

The values $R(0), R(\frac{t_2}{2}), R(t_2)$ are at least 1, because by Lemma 3.47 they are non-negative, and $S(t) > 1$ whenever $t \in [0, \frac{\pi}{2}]$ and both $a_{q,w,1}(t), a_{q,w,2}(t) > 0$. The equation

$$R = \left(\frac{(1+w)R + q - 1}{R + w + q - 1} \right)^{d-1}$$

has at most 3 solutions satisfying $R \geq 1$ by Lemma 3.37 which means that there is no other $t' \in (0, \frac{\pi}{2})$ that is a local maximizer or minimizer of $\Phi_{d,q,w}(t)$. Note that $\frac{d^2}{dt^2} \Phi_{d,q,w_c} \Big|_{t=0} < 0$ since $w_c(q) < \frac{q}{d-2}$ by Lemma 3.50. So at $\frac{t_2}{2}$ we have a local minimum, and at t_2 we have a local maximum. Hence

$$\Phi_{d,q,w_c} = \Phi_{d,q,w_c}(0) = \Phi_{d,q,w_c}(t_2) = q \left(1 + \frac{w}{q} \right)^{d/2}.$$

□

Proposition 3.52. For $q \geq 2$ and $0 \leq w \leq w_c$ we have $\Phi_{d,q,w} = q \left(1 + \frac{w}{q} \right)^{d/2}$.

Proof. We will describe the pairs (q, w) for which $\Phi_{d,q,w} = q \left(1 + \frac{w}{q} \right)^{d/2}$. To do this it is better to use the Tutte polynomial $T_G(x, y)$ instead of $Z_G(q, w)$ with $q = (x-1)(y-1)$

and $w = y - 1$. Recall that the connection between the Tutte polynomial and the partition function of the random cluster model is the following:

$$T_G(x, y) = (x - 1)^{-k(E)}(y - 1)^{-v(G)}Z_G((x - 1)(y - 1), y - 1).$$

Then for $q \geq 2$ and an essentially large girth sequence of d -regular graphs $(G_n)_n$ the statement

$$\lim_{n \rightarrow \infty} Z_{G_n}^{(2)}(q, w)^{1/v(G_n)} = \lim_{n \rightarrow \infty} Z_{G_n}(q, w)^{1/v(G_n)} = q \left(1 + \frac{w}{q}\right)^{d/2}$$

is equivalent with

$$\lim_{n \rightarrow \infty} T_{G_n}(x, y)^{1/v(G_n)} = x \left(1 + \frac{1}{x - 1}\right)^{d/2-1}.$$

This is independent of y . The Tutte polynomial has only non-negative coefficients [25], so if this limit value holds true for (x, y_1) and (x, y_2) , then so for every $y \in [y_1, y_2]$. Note that for $x \geq d - 1$ and $y = 1$ this was indeed proved by Bencs and Csikvári [4]. In fact, we do not even need to use this result since for $q = 1$ this statement is trivial. By Lemma 3.50 the curve $(q, w_c(q))$ for $q \geq 2$ reparametrized with x and y is the graph of a monotone increasing function on the interval $[d - 1, \infty)$, see the dashed line on Fig. 1. In particular, for $q \geq 2$ the part of the hyperbola $(x - 1)(y - 1) = q$ with $0 \leq w = y - 1 \leq w_c$ goes under this curve implying $\Phi_{d,q,w} = q \left(1 + \frac{w}{q}\right)^{d/2}$. \square

Remark 3.53. We remark that the same argument also gives that if $1 < q < 2$ and $0 \leq w \leq \frac{q}{d-2}$, then for an essentially large girth sequence of d -regular graphs $(G_n)_n$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}^{(2)}(q, w) = \lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w) = q \left(1 + \frac{w}{q}\right)^{d/2}.$$

Proposition 3.54. *For $q \geq 2$ and $w > w_c$ we have $\Phi_{d,q,w} > q \left(1 + \frac{w}{q}\right)^{d/2}$. Furthermore, the function $\frac{\partial}{\partial w} \Phi_{d,q,w}$ has a discontinuity at $w = w_c$ if $q > 2$.*

Proof. Consider the function

$$h(w, t) = \left(1 + \frac{w}{q}\right)^{-d/2} \Phi_{d,q,w}(t).$$

We show that it is a strictly monotone increasing function in w for every $t \in (0, \frac{\pi}{2})$. By definition

$$h(w, t) = \left(\cos(t) + \sqrt{\frac{(q - 1)w}{q + w}} \sin(t)\right)^d + (q - 1) \left(\cos(t) - \sqrt{\frac{w}{(q - 1)(q + w)}} \sin(t)\right)^d.$$

Then $\frac{\partial h}{\partial w}$ is given as

$$\frac{dq\sqrt{q-1}\sin(t)}{\sqrt{w(q+w)^3}} \left(\left(\cos(t) + \sqrt{\frac{(q-1)w}{q+w}} \sin(t) \right)^{d-1} - \left(\cos(t) - \sqrt{\frac{w}{(q-1)(q+w)}} \sin(t) \right)^{d-1} \right).$$

This is positive if $t \in (0, \frac{\pi}{2})$ since

$$\cos(t) + \sqrt{\frac{(q-1)w}{q+w}} \sin(t) > \left| \cos(t) - \sqrt{\frac{w}{(q-1)(q+w)}} \sin(t) \right|.$$

Note that for $q > 2$ there is a $t_0(w_c) \in (0, \frac{\pi}{2})$ such that $\Phi_{d,q,w_c}(t_0(w_c)) = \Phi_{d,q,w_c}(0)$, that is, $h(w_c, t_0(w_c)) = q$. Then for $w > w_c$ we have $h(w, t_0(w_c)) > q$ which gives that

$$\Phi_{d,q,w} \geq \Phi_{d,q,w}(t_0(w_c)) > q \left(1 + \frac{w}{q}\right)^{d/2}.$$

For $q = 2$ we know that $w_c = \frac{2}{d-2}$ and for $w > \frac{q}{d-2} = \frac{2}{d-2}$ we have $\frac{\partial}{\partial t} \Phi_{d,q,w}(t) \Big|_{t=0} = 0$ and $\frac{\partial}{\partial t^2} \Phi_{d,q,w}(t) \Big|_{t=0} < 0$, so at $t = 0$ we have a local minimum, thus $\Phi_{d,q,w} > \Phi_{d,q,w}(0)$ for $w > w_c$.

Next we prove the claim about $\frac{\partial}{\partial w} \Phi_{d,q,w}$. Let $w > w_c$ such that $w - w_c$ is small enough, namely it satisfies

$$\begin{aligned} h(w, t_0(w_c)) &\geq h(w_c, t_0(w_c)) + \frac{1}{2}(w - w_c) \frac{\partial}{\partial w} h(w, t_0(w_c)) \Big|_{w=w_c} \\ &= q + \frac{1}{2}(w - w_c) \frac{\partial}{\partial w} h(w, t_0(w_c)) \Big|_{w=w_c}. \end{aligned}$$

Then

$$\begin{aligned} \frac{\Phi_{d,q,w} - \Phi_{d,q,w_c}}{w - w_c} &\geq \frac{1}{w - w_c} \left(\left(1 + \frac{w}{q}\right)^{d/2} h(w, t_0(w_c)) - q \left(1 + \frac{w_c}{q}\right)^{d/2} \right) \\ &\geq \frac{1}{w - w_c} \left(\left(1 + \frac{w}{q}\right)^{d/2} \left(q + \frac{1}{2}(w - w_c) \frac{\partial}{\partial w} h(w, t_0(w_c)) \Big|_{w=w_c} \right) - q \left(1 + \frac{w_c}{q}\right)^{d/2} \right) \end{aligned}$$

From this it follows that

$$\frac{\partial}{\partial w^+} \Phi_{d,q,w} \Big|_{w=w_c} \geq \frac{\partial}{\partial w^-} \Phi_{d,q,w} \Big|_{w=w_c} + \frac{1}{2} \left(1 + \frac{w_c}{q}\right)^{d/2} \frac{\partial}{\partial w} h(w, t_0(w_c)) \Big|_{w=w_c}.$$

□

Remark 3.55. It is well-known that if $q = 2$, then there is a second order phase transition, that is, $\frac{\partial}{\partial w} \Phi_{d,2,w}$ is continuous, but $\frac{\partial^2}{\partial w^2} \Phi_{d,2,w}$ is discontinuous at $w = w_c = \frac{2}{d-2}$. For details see Chapter 4.8 of [3].

3.10. Examples. In this section we give some examples for the theorems we proved.

Example 3.56. Let $d = 8, q = 5$ and $w = 1$. Then the vector

$$\underline{v}(0) = (10.368, 0, 1.728, 1.058, 0.936, 0.749, 0.615, 0.501, 0.409),$$

where we kept only the first three digits everywhere. Note that $10.368 = 5 \cdot (1 + \frac{1}{5})^{8/2}$. So for every 8-regular graph G we have

$$Z_G^{(2)}(5, 1) = F_G(10.368, 0, 1.728, 1.058, 0.936, 0.749, 0.615, 0.501, 0.409).$$

Using $t_0 = 0.6619549492373429$ we get the vector

$$\underline{v}(t_0) = (16.277, 0, 0.433, -0.496, 0.581, -0.679, 0.794, -0.929, 1.086)$$

again only keeping the first 3 digits everywhere. A more precise value of the first coordinate is 16.277748757985485, and so this is $\Phi_{8,5,1}$. Note that the sign structure of $\underline{v}(t_0)$ shows that

$$\begin{aligned} Z_G^{(2)}(5, 1) &= F_G(16.277, 0, 0.433, -0.496, 0.581, -0.679, 0.794, -0.929, 1.086) \\ &\geq 16.277^{v(G)} \end{aligned}$$

for every 8-regular graph G .

Example 3.57. Let $d = 4, q = 5$ and $w = 3$. Then

$$\underline{v}(0) = (12.8, 0, 4.8, 4.409, 5.85)$$

where we again kept only the first three digits everywhere. This time $t_0 = 0.8316331320342567$ and $\Phi_{4,5,3} = 16.315621073058985$ while

$$\underline{v}(t_0) = (16.315, 0, 1.878, -3.867, 8.176).$$

In this case $t_1 = 1.06627054934707$ and the corresponding vector

$$\underline{v}(t_1) = (15.010, -2.835, 0.994, -2.454, 11.249).$$

For the complete graph K_5 on 5 vertices the subgraph counting polynomial looks like as follows:

$$\begin{aligned} F_{K_5}(15.010, -2.835, 0.994, -2.454, 11.249 \mid z) &= 180176.234z^{20} + 85764.618z^{18} \\ &+ 28876.392z^{16} + 15784.587z^{14} + 10454.536z^{12} + 9093.510z^{10} + 13949.743z^8 \\ &+ 28103.222z^6 + 68600.498z^4 + 271865.398z^2 + 762087.303 \end{aligned}$$

All zeros of this polynomial have absolute value approximately 1.0747696. Of course, we could have used any 4-regular graph instead of K_5 (see Fig. 6).

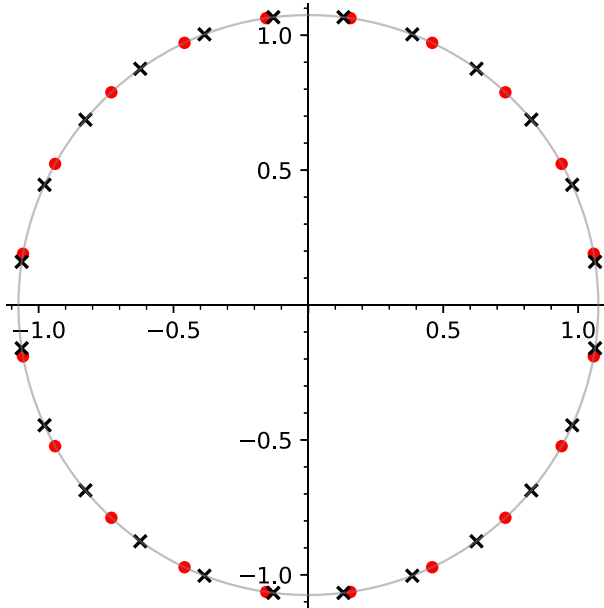


Fig. 6. The zeros of $F_G(15.010, -2.835, 0.994, -2.454, 11.249 | z)$, where G is K_5 (red) and G is the octahedron (black x)

Example 3.58. Let $d = 4$ and $q = 5$ again, but let $w = w_c = 2$. Then

$$\underline{v}(0) = (9.8, 0, 2.8, \sqrt{5.04}, 2.6),$$

$t_0 = 0.5575988373258864$ and

$$\underline{v}(t_0) = (9.8, 0, 2.8, -\sqrt{5.04}, 2.6).$$

We have $t_1 = 1.06419757674722$ and

$$\underline{v}(t_1) = (8, -\sqrt{4.5}, 1, -\sqrt{4.5}, 8).$$

One can check that

$$F_{K_5}(8, -\sqrt{4.5}, 1, -\sqrt{4.5}, 8 | z) = 32768z^{20} + 23040z^{18} + 11070z^{16} + 6647.5z^{14} + 4620z^{12} + 3927z^{10} + 4620z^8 + 6647.5z^6 + 11070z^4 + 23040z^2 + 32768$$

and all of its zeros have absolute value 1.

4. Selected Remarks About the Interval $1 < q < 2$

In this section we collected several remarks about the interval $1 < q < 2$.

4.1. *Two different quantities.* In this section we aim to explain a seemingly negligible thing that makes the interval $q \geq 2$ and $1 < q < 2$ really different.

Once again let $N = M'_2$ and $\underline{\mu} = \underline{\nu}_2$, and parametrize the distribution h in the Bethe recursion as follows:

$$h = \left(\frac{R}{R + q - 1}, \frac{q - 1}{R + q - 1} \right).$$

Then

$$\text{BP}(h)_1 = \frac{1}{z_h} \left(\frac{(1 + w)R + q - 1}{R + q - 1} \right)^{d-1}$$

and

$$\text{BP}(h)_2 := \frac{1}{z_h} (q - 1) \left(\frac{R + w + q - 1}{R + q - 1} \right)^{d-1}.$$

If $\text{BP}(h) = h$, then by dividing the Bethe recursions for h_1 and h_2 we get that

$$R = \left(\frac{(1 + w)R + q - 1}{R + w + q - 1} \right)^{d-1}.$$

We remark that if we study the Potts model $N = M = wI_q + J_q$ and $\mu \equiv 1$ with

$$h = \left(\frac{R}{R + q - 1}, \frac{1}{R + q - 1}, \dots, \frac{1}{R + q - 1} \right),$$

then we would have arrived to the same equation. Let $\mathcal{R}_{d,q,w}$ be the set of non-negative solutions of this equation. Let $\mathcal{R}_{d,q,w}^*$ be the solutions satisfying also that $R \geq 1$. Let us introduce the notation

$$\begin{aligned} \overline{\mathbb{F}}(R, d, q, w) &= \left(\frac{(1 + w)R + q - 1}{\sqrt{(1 + w)(R^2 + q - 1) + 2R(q - 1) + (q - 1)(q - 2)}} \right)^d + \\ &+ (q - 1) \left(\frac{R + q + w - 1}{\sqrt{(1 + w)(R^2 + q - 1) + 2R(q - 1) + (q - 1)(q - 2)}} \right)^d \end{aligned}$$

Then we know that

$$\Phi_d(M'_2, \nu_2) = \max_{R \in \mathcal{R}_{d,q,w}} \overline{\mathbb{F}}(R, d, q, w).$$

For later use let us also introduce

$$\Phi_d^*(M'_2, \nu_2) = \max_{R \in \mathcal{R}_{d,q,w}^*} \overline{\mathbb{F}}(R, d, q, w).$$

Similarly, we can consider the pair

$$\Phi_{d,q,w} = \max_{t \in [0, 2\pi]} \Phi_{d,q,w}(t) \text{ and } \Phi_{d,q,w}^* = \max_{t \in [0, \frac{\pi}{2}]} \Phi_{d,q,w}(t).$$

In case of $q \geq 2$ we have

$$\Phi_d(M'_2, \nu_2) = \Phi_{d,q,w} = \Phi_d^*(M'_2, \nu_2) = \Phi_{d,q,w}^*.$$

But when $1 < q < 2$ we have

$$\Phi_d(M'_2, \nu_2) = \Phi_{d,q,w} > \Phi_d^*(M'_2, \nu_2) = \Phi_{d,q,w}^*.$$

While it is still true that for an essentially large girth sequence of d -regular graphs we have

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}^{(2)}(q, w) = \Phi_{d,q,w},$$

we actually believe that

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w) = \Phi_{d,q,w}^*.$$

This means that the rank 2 approximation is not good enough in the interval $1 < q < 2$.

Nevertheless, by Remark 3.53 we know that for $1 < q < 2$ and $0 \leq w \leq \frac{q}{d-2}$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_{G_n}(q, w) = q \left(1 + \frac{w}{q} \right)^{d/2}.$$

We remark that this result is compatible with the conjecture

$$\lim_{n \rightarrow \infty} \frac{1}{v(G_n)} \ln Z_G(q, w) = \Phi_{d,q,w}^*$$

since for the function $\Phi_{d,q,w}(t)$ we have

$$\frac{\partial}{\partial t} \Phi_{d,q,w}(t) \Big|_{t=0} = 0 \quad \text{and} \quad \frac{\partial^2}{\partial t^2} \Phi_{d,q,w}(t) \Big|_{t=0} = d \left(1 + \frac{w}{q} \right)^{d/2-1} ((d-2)w - q)$$

which is negative if $w < \frac{q}{d-2}$ and positive $w > \frac{q}{d-2}$. So in the first case we get that $t = 0$ is a local maximum, in the second case it is a local minimum.

Acknowledgements. We are very grateful to the anonymous reviewers for the careful reading and the suggestions leading to a significant improvement in the presentation of this paper.

Funding Open access funding provided by ELKH Alfréd Rényi Institute of Mathematics.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Abért, M., Hubai, T.: Benjamini–Schramm convergence and the distribution of chromatic roots for sparse graphs. *Combinatorica* **35**(2), 127–151 (2015)
2. Bandyopadhyay, A., Gamarnik, D.: Counting without sampling: asymptotics of the log-partition function for certain statistical physics models. *Random Struct. Algorithms* **33**(4), 452–479 (2008)
3. Baxter, R.J.: *Exactly Solved Models in Statistical Mechanics*. Elsevier, Amsterdam (2016)
4. Bencs, F., Csikvári, P.: Evaluations of Tutte polynomials of regular graphs. *J. Comb. Theory Ser. B* **157**, 500–523 (2022)
5. Borbényi, M., Csikvári, P.: Counting degree-constrained subgraphs and orientations. *Discrete Math.* **343**(6), 111842 (2020)
6. Borgs, C., Chayes, J., Kahn, J., Lovász, L.: Left and right convergence of graphs with bounded degree. *Random Struct. Algorithms* **42**(1), 1–28 (2013)
7. Carlson, C., Davies, E., Kolla, A.: Efficient algorithms for the Potts model on small-set expanders. arXiv preprint [arXiv:2003.01154](https://arxiv.org/abs/2003.01154) (2020)
8. Carlson, C., Davies, E., Fraiman, N., Kolla, A., Potukuchi, A., Yap, C.: Algorithms for the ferromagnetic Potts model on expanders. arXiv preprint [arXiv:2204.01923](https://arxiv.org/abs/2204.01923) (2022)
9. Chertkov, M., Chernyak, V.Y.: Loop calculus in statistical physics and information science. *Phys. Rev. E* **73**(6), 065102 (2006)
10. Chertkov, M., Chernyak, V.Y.: Loop series for discrete statistical models on graphs. *J. Stat. Mech. Theory Exp.* **2006**(06), P06009 (2006)
11. Csikvári, P., Frenkel, P.E.: Benjamini–Schramm continuity of root moments of graph polynomials. *Eur. J. Comb.* **52**, 302–320 (2016)
12. Dembo, A., Montanari, A.: Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20**(2), 565–592 (2010)
13. Dembo, A., Montanari, A., Sly, A., Sun, N.: The replica symmetric solution for Potts models on d-regular graphs. *Commun. Math. Phys.* **327**(2), 551–575 (2014)
14. Dembo, A., Montanari, A., Sun, N.: Factor models on locally tree-like graphs. *Ann. Probab.* **41**(6), 4162–4213 (2013)
15. Galanis, A., Štefankovič, D., Vigoda, E., Yang, L.: Ferromagnetic Potts model: refined #BIS-hardness and related results. *SIAM J. Comput.* **45**(6), 2004–2065 (2016)
16. Grimmett, G.: The random-cluster model. In: *Probability on Discrete Structures*, pp. 73–123. Springer, Berlin (2004)
17. Helmuth, T., Jenssen, M., Perkins, W.: Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs. arXiv preprint [arXiv:2006.11580](https://arxiv.org/abs/2006.11580) (2020)
18. Lee, T.-D., Yang, C.-N.: Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev.* **87**(3), 410 (1952)
19. McKay, B.D., Wormald, N.C., Wysocka, B.: Short cycles in random regular graphs. *Electron. J. Comb.* **11**, R66–R66 (2004)
20. Mezard, M., Montanari, A.: *Information, Physics, and Computation*. Oxford University Press, Oxford (2009)
21. Ruožzi, N.: The Bethe partition function of log-supermodular graphical models. In: *Advances in Neural Information Processing Systems*, 25 (2012)
22. Ruožzi, N.: Beyond log-supermodularity: lower bounds and the Bethe partition function. In: *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence* (2013)
23. Sly, A., Sun, N.: The computational hardness of counting in two-spin models on d-regular graphs. In: *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pp. 361–369. IEEE (2012)
24. Sly, A., Sun, N.: Counting in two-spin models on d-regular graphs. *Ann. Probab.* **42**(6), 2383–2416 (2014)
25. Tutte, W.T.: A contribution to the theory of chromatic polynomials. *Can. J. Math.* **6**, 80–91 (1954)
26. Wagner, D.G.: Weighted enumeration of spanning subgraphs with degree constraints. *J. Comb. Theory Ser. B* **99**(2), 347–357 (2009)
27. Yang, C.-N., Lee, T.-D.: Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Phys. Rev.* **87**(3), 404 (1952)