Journal of Combinatorics
Volume 3, Number 1, 49–62, 2012

Tree-minimal graphs are almost regular

D. Dellamonica Jr.*, P. Haxell†, T. Łuczak‡, D. Mubayi§, B. Nagle¶, Y. Person‖, V. Rödl** and M. Schacht††

For all fixed trees T and any graph G we derive a counting formula for the number $\hat{N}_T(G)$ of homomorphisms from T to G in terms of the degree sequence of G.

As a consequence we obtain that any n-vertex graph G with edge density $p = p(n) \gg n^{-1/(t-2)}$, which contains at most $(1 + o(1))p^{t-1}n^t$ copies of some fixed tree T with $t \geq 3$ vertices must be almost regular, i.e., $\sum_{v \in V(G)} |\deg(v) - pn| = o(pn^2)$.

1. Introduction

For graphs F and G, let $N_F(G)$ denote the number of labeled copies of F in G. A well-known conjecture, due to Erdős and Simonovits (see [8]) and Sidorenko [6, 7], asserts that for every bipartite graph F and every $p > 0$, if an n-vertex graph G contains at least $p\binom{n}{2}$ edges, then

$$N_F(G) \geq (1 - o(1))p^{e_F} n^{v_F}. \tag{1}$$

This conjecture is known to hold for several classes of graphs F, including forests, even cycles, and complete bipartite graphs [7], Boolean cubes [5] and bipartite graphs F which contain a vertex that is connected to every vertex in the other vertex class [3].

The bound in (1) is asymptotically best possible, as for example the random graph $G(n, p)$ and, more generally, *quasi-random graphs* of density p

*Supported by a CAPES/Fulbright scholarship.
†Partially supported by the NSERC.
‡Partially supported by the Foundation for Polish Science.
§Supported in part by NSF grants DMS 0653946 and DMS 0969092.
¶Supported by NSF grant DMS 1001781.
‖Supported by GIF grant no. I-889-182.6/2005.
**Supported by NSF grant DMS 0800070.
††Supported through the Heisenberg-Programme of the Deutsche Forschungsgemeinschaft (DFG Grant SCHA 1263/4-1).

show (see, e.g. [2, 10]). Skokan and Thoma [9] asked to what extent the converse holds, and posed the following problem: *if F is a bipartite graph, but not a forest, and if $N_F(G) \leq (1+o(1))p^{e_F}n^{v_F}$, does G have to be quasi-random, in the sense of* [2]*?* It is believed that the answer to this question is positive, and this is known as the *forcing conjecture for quasi-random graphs.*

The exclusion of forests F in the forcing conjecture is clearly necessary, as there are examples of regular graphs G which (therefore) minimize $N_F(G)$ but which are not quasi-random. We study this case and address the following problem:

> *What structure on G is enforced when $N_T(G) \leq (1+o(1))p^{t-k}n^t$ for a forest T with t vertices and k components?*

Since $N_T(G) = (1+o(1))N_{T'}(G) \cdot N_{T''}(G)$, when T is the (vertex) disjoint union of T' and T'', it suffices to consider connected graphs T only. Moreover, the case when T consists of only one edge is rather uninteresting, as in this case $N_T(G) = 2|E(G)|$. Consequently, from now on we restrict ourselves to trees T with at least three vertices.

For an n-vertex graph $G = (V, E)$ and $v \in V$, let d_v denote the degree of v in G and let $d = d(G) = \frac{1}{n}\sum_{v \in V} d_v$ denote the average degree of G. Clearly, if $|E| \geq p\binom{n}{2}$, then $d(G) \geq p(n-1)$. It is easy to see that for every tree T, and every $p = p(n) \gg 1/n$, the following holds: if $G = (V, E)$ is an *almost regular* n-vertex graph, that is, $\sum_{v \in V} |d_v - d| \leq o(pn^2)$ and $|E| = p\binom{n}{2}$, then $N_T(G) \leq (1+o(1))p^{t-1}n^t$. Our first result provides a converse.

Theorem 1. *For all $\varepsilon \in (0,1]$ and integers $t \geq 3$, there exists $\delta > 0$ so that whenever $p \gg n^{-1/(t-2)}$, the following statement holds. Let T be a tree with t vertices, and let $G = (V, E)$ be a graph with n vertices and $|E| = p\binom{n}{2}$ edges. If $N_T(G) \leq (1+\delta)p^{t-1}n^t$, then $\sum_{v \in V} |d_v - d| \leq \varepsilon pn^2$.*

For the proof of Theorem 1 we consider not only copies of T, but more generally, homomorphisms (edge-preserving maps) $h\colon V(T) \to V(G)$. Let $\hat{N}_T(G)$ denote the number of such homomorphisms. Note that

$$N_T(G) \leq \hat{N}_T(G) \leq N_T(G) + O(pn^{t-1}). \tag{2}$$

To see the error term $O(pn^{t-1})$ above, consider a non-injective homomorphic image S of T in G. The image S contains at least one edge (of which there are $p\binom{n}{2}$) and at most $t-3$ other vertices (disjoint from the chosen edge).

The bound in (2) restricts the proof of Theorem 1 to $p \gg n^{-1/(t-2)}$. We believe Theorem 1 is valid for much smaller values of p.

Problem 2. *Does Theorem 1 remain true as long as $p \gg 1/n$?*

Theorem 1 is a consequence of the main result of the paper, which extends a result of Alon, Hoory and Linial [1] from paths to arbitrary trees.

Theorem 3. *Let T be a tree with $t \geq 3$ vertices, and let $G = (V, E)$ be a graph with n vertices and average degree d. Then,*

$$\hat{N}_T(G) \geq nd \prod_{v \in V} d_v^{\frac{(t-2)d_v}{nd}}.$$

We prove Theorem 3 in Section 3 using ideas from [1].

We believe that the counting formula for homomorphisms of T in G given in Theorem 3 remains more or less valid for injective homomorphisms of T in G (i.e. labeled copies of T in G), provided the minimum degree $\delta(G)$ of G is sufficiently large.

Problem 4. *Prove a lower bound for $N_T(G)$, the number of labeled copies of a tree T in a graph G with sufficiently large $\delta(G)$, which would be analogous to that proven for $\hat{N}_T(G)$ in Theorem 3.*

In fact, the proof of Theorem 3 can be altered to give such a result when T is the path P_3 with three edges. We will sketch the proof of Theorem 5 below in Section 5.

Theorem 5. *Let P_3 denote the path with 3 edges and let G be a graph with n vertices, average degree d and minimum degree at least 3. Then*

$$N_{P_3}(G) \geq nd \prod_{v \in V} (d_v - 2)^{\frac{2d_v}{nd}}.$$

Problem 4 and Theorem 5 are related to a result of Erdős and Simonovits [4]. They proved that the number of walks of length ℓ that are not paths in G is a negligible proportion of the total number of walks of length ℓ as the average degree d of G goes to infinity. In our notation,

$$\hat{N}_{P_\ell}(G) - N_{P_\ell}(G) = o(\hat{N}_{P_\ell}(G)) \quad \text{as } d \to \infty.$$

Together with the special case of Theorem 3 when $T = P_\ell$, which was earlier proved in [1], the Erdős-Simonovits result implies that

$$N_{P_\ell}(G) \geq (1 - o(1))\, nd \prod_{v \in V(G)} d_v^{\frac{(\ell-1)d_v}{nd}} \quad \text{as } d \to \infty. \tag{3}$$

The dependence of the $o(1)$ term above on d was not very good (and not even really made explicit in [4]). Theorem 5 can thus be viewed as an improvement of (3) in the case $\ell = 3$.

2. Proof of Theorem 1

In this section, we deduce Theorem 1 from Theorem 3. For that, we will also use the following consequence of Jensen's inequality.

Lemma 6. *Let $G = (V, E)$ be a graph with $|V| = n$ vertices and average degree $d = d(G)$. Suppose*

$$\frac{1}{n}\sum_{v\in V} d_v \log d_v < \frac{\gamma^2}{\gamma+2} d + d\log d \tag{4}$$

for some $\gamma > 0$. Then

$$\frac{1}{n}\sum_{v\in V}\left|\frac{d_v}{d} - 1\right| < \gamma.$$

For completeness, we include a proof of Lemma 6 at the end of the section and first we deduce Theorem 1 from Theorem 3 and Lemma 6.

Proof of Theorem 1. Let $\varepsilon \in (0, 1]$ and $t \geq 3$ be given. With $\gamma = \varepsilon$, let $\delta = \gamma^2/(3(\gamma+2))$, and let $p \gg n^{-1/(t-2)}$. Let T be a tree with t vertices and let $G = (V, E)$ be a graph with $|V| = n$ vertices, with $|E| = p\binom{n}{2}$ edges, and average degree $d = p(n-1)$. Suppose

$$N_T(G) \leq (1+\delta)p^{t-1}n^t.$$

From Theorem 3 and (2), we infer

$$\prod_{v\in V} d_v^{\frac{(t-2)d_v}{nd}} \leq \frac{\hat{N}_T(G)}{nd} \leq \frac{1}{nd}\left((1+\delta)p^{t-1}n^t + O(pn^{t-1})\right).$$

Since $p \gg n^{-1/(t-2)}$, we have $p^{t-1}n^t \gg pn^{t-1}$, and therefore,

$$\prod_{v\in V} d_v^{\frac{(t-2)d_v}{nd}} \leq \frac{(1+2\delta)}{nd}p^{t-1}n^t \leq (1+3\delta)d^{t-2}.$$

Taking natural logarithms then yields

$$\frac{t-2}{nd}\sum_{v\in V} d_v \log d_v \leq 3\delta + (t-2)\log d,$$

which implies,

$$\frac{1}{n}\sum_{v\in V} d_v \log d_v \le \frac{3\delta}{t-2}d + d\log d \le \frac{\gamma^2}{\gamma+2}d + d\log d. \tag{5}$$

Therefore, Lemma 6 and our choice of constants yield

$$\sum_{v\in V} |d_v - d| \le \gamma dn \le \varepsilon p n^2\,,$$

which proves Theorem 1. □

Proof of Lemma 6. We proceed by contradiction. In fact, since the RHS of (4) is increasing in γ, we may assume

$$\frac{1}{n}\sum_{v\in V}\left|\frac{d_v}{d} - 1\right| = \gamma. \tag{6}$$

Observe that (4) may be rewritten as

$$\frac{1}{n}\sum_{v\in V}\frac{d_v}{d}\log\frac{d_v}{d} < \frac{\gamma^2}{\gamma+2}.$$

We show, however, that the assumption in (6) implies

$$\frac{1}{n}\sum_{v\in V}\frac{d_v}{d}\log\frac{d_v}{d} \ge \frac{\gamma^2}{\gamma+2}, \tag{7}$$

a contradiction. It remains to prove (7).

Set

$$W = \{v\in V\colon\ d_v > d\},$$

and write, as usual, $W^C = V\setminus W$ for the complement. Since $\sum_{v\in V}(d_v - d) = 0$, it follows that

$$\sum_{v\in W}\left(\frac{d_v}{d} - 1\right) = \sum_{v\in W^C}\left(1 - \frac{d_v}{d}\right).$$

Consequently, (6) implies that

$$\sum_{v\in W}\left(\frac{d_v}{d} - 1\right) = \frac{\gamma n}{2} = \sum_{v\in W^C}\left(1 - \frac{d_v}{d}\right). \tag{8}$$

Write $\alpha_1 = |W|/n$ and $\alpha_2 = 1 - \alpha_1 = |W^C|/n$.

Setting $\phi(x) = x\log x$ (a convex function), Jensen's inequality implies that

$$\frac{1}{|W|}\sum_{v\in W}\phi\Big(\frac{d_v}{d}\Big) \ge \phi\Big(\frac{1}{|W|}\sum_{v\in W}\frac{d_v}{d}\Big) \overset{(8)}{=} \phi\Big(1+\frac{\gamma}{2\alpha_1}\Big).$$

Similarly,

$$\frac{1}{|W^C|}\sum_{v\in W^C}\phi\Big(\frac{d_v}{d}\Big) \ge \phi\Big(\frac{1}{|W^C|}\sum_{v\in W^C}\frac{d_v}{d}\Big) \overset{(8)}{=} \phi\Big(1-\frac{\gamma}{2\alpha_2}\Big).$$

From the two inequalities above, we obtain

$$\frac{1}{n}\sum_{v\in V}\frac{d_v}{d}\log\frac{d_v}{d} \ge \alpha_1\,\phi\Big(1+\frac{\gamma}{2\alpha_1}\Big) + \alpha_2\,\phi\Big(1-\frac{\gamma}{2\alpha_2}\Big). \tag{9}$$

Now, set

$$y_1 = 1+\frac{\gamma}{2\alpha_1} \quad\text{and}\quad y_2 = 1-\frac{\gamma}{2\alpha_2}, \tag{10}$$

and observe that $\alpha_1 y_1 + \alpha_2 y_2 = 1$. By Taylor's theorem, there exist reals $\xi_1 \in (1, y_1)$ and $\xi_2 \in (y_2, 1)$ such that for $i = 1, 2$, we have

$$\phi(y_i) = \phi(1) + (y_i - 1)\phi'(1) + \frac{(y_i-1)^2}{2}\phi''(\xi_i).$$

Since $\phi(1) = 0$, $\phi'(1) = 1$, and $\phi''(\xi_i) = 1/\xi_i$, we have

$$\begin{aligned}
\alpha_1\,\phi(y_1) + \alpha_2\,\phi(y_2) &= \frac{\alpha_1}{2\xi_1}(y_1-1)^2 + \frac{\alpha_2}{2\xi_2}(y_2-1)^2 \\
&> \frac{\alpha_1}{2y_1}(y_1-1)^2 + \frac{\alpha_2}{2}(y_2-1)^2 \\
&\overset{(10)}{=} \frac{\gamma^2}{4(\gamma+2\alpha_1)} + \frac{\gamma^2}{8\alpha_2} \\
&= \frac{\gamma^2(\gamma+2)}{8(\gamma+2\alpha_1)(1-\alpha_1)}. \qquad (11)
\end{aligned}$$

The denominator is maximized with $\alpha_1 = 1/2-\gamma/4$, there equaling $(\gamma+2)^2$. Hence

$$\frac{1}{n}\sum_{v\in V}\frac{d_v}{d}\log\frac{d_v}{d} \overset{(9)}{\ge} \alpha_1\,\phi(y_1) + \alpha_2\,\phi(y_2) \overset{(11)}{\ge} \frac{\gamma^2}{\gamma+2}\,,$$

proving (7). □

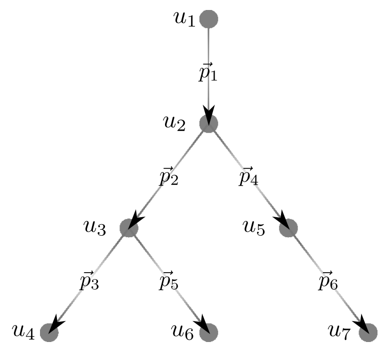

Figure 1: Tree labeling and orientation.

3. Proof of Theorem 3

Let $T = (U, P)$ be a tree with $t \geq 3$ vertices, and let $G = (V, E)$ be a graph with $|V| = n$ vertices and average degree $d = d(G)$. We begin the proof by describing some notation.

We will consider the digraph $\vec{G} = (V, \vec{E})$ obtained from G by including, for each edge $\{v, v'\} \in E$, both arcs (v, v') and (v', v) into $\vec{E}$.

Moreover, we will label the vertices and orient the edges of T. For that we fix a leaf u_1 of T and let u_2 be the unique neighbor of u_1 in T. Furthermore, fix a labeling $u_3, \ldots, u_t$ of $U \setminus \{u_1, u_2\}$ such that, for every $i = 1, \ldots, t$, the induced subgraph $T_i = T[\{u_1, \ldots, u_i\}]$ is connected. Let $P = \{p_1, \ldots, p_{t-1}\}$ be the corresponding labeling of the edges of T defined by the property that $E(T_i) = \{p_1, \ldots, p_{i-1}\}$ for every $i = 2, \ldots, t$. For every edge $p = \{u_i, u_j\}$ with $i < j$, we denote by $\vec{p}$ the oriented pair (u_i, u_j). Note that, from these definitions, it follows that, for every $j = 1, \ldots, t-1$, we have $\vec{p}_j = (u_i, u_{j+1})$ for some $i < j$. Finally, we denote by $\vec{T} = (U, \vec{P})$ the oriented tree with $\vec{P} = \{\vec{p}\colon\ p \in P\}$ (see Figure 1).

Now, let $\Omega = \Omega_T(G)$ denote the family of all homomorphisms $\omega\colon\ U \to V$ (so that $|\Omega| = \hat{N}_T(G)$). We develop some notation for Ω. To begin, if $\vec{p} = (u, u') \in \vec{P}$, $\vec{e} = (v, v') \in \vec{E}$, and $\omega \in \Omega$, we write $\omega(\vec{p}) = \vec{e}$ if $\omega(u) = v$ and $\omega(u') = v'$. In this way, we may view the image of an element $\omega \in \Omega$ as a $(t-1)$-tuple $\omega = (\omega(\vec{p}_1), \ldots, \omega(\vec{p}_{t-1}))$. For $\vec{e}_1 \in \vec{E}$, define $\Omega_{\vec{e}_1} \subseteq \Omega$ to be the set of those homomorphisms $\omega \in \Omega$ for which $\omega(\vec{p}_1) = \vec{e}_1$. For $\omega \in \Omega_{\vec{e}_1}$, we write $\omega_- = (\omega(\vec{p}_2), \ldots, \omega(\vec{p}_{t-1}))$. For an arc $(v, v') \in \vec{E}$, we shall sometimes abuse notation and write $(v, v') \in \omega_-$ to mean $(v, v') \in \{\omega(\vec{p}_2), \ldots, \omega(\vec{p}_{t-1})\}$, and we denote by $m_{\omega_-}(v, v')$ the number of appearances of $(v, v') \in \{\omega(\vec{p}_2), \ldots, \omega(\vec{p}_{t-1})\}$, i.e.,

$$m_{\omega_-}(v, v') = |\{j = 2, \ldots, t-1\colon\ \omega(\vec{p}_j) = (v, v')\}|.$$

We consider the following procedure for generating a random element $\boldsymbol{\omega} = \{\boldsymbol{\omega}(\vec{p}_1), \ldots, \boldsymbol{\omega}(\vec{p}_{t-1})\} \in \Omega$. Select $\vec{e}_1 = \boldsymbol{\omega}(\vec{p}_1) \in \vec{E}$ uniformly at random. For $2 \le j < t$, suppose $\vec{e}_1 = \boldsymbol{\omega}(\vec{p}_1), \ldots, \vec{e}_{j-1} = \boldsymbol{\omega}(\vec{p}_{j-1}) \in \vec{E}$ have been selected. Then, choose $\vec{e}_j = \boldsymbol{\omega}(\vec{p}_j) = (v, v') \in \vec{E}$ by selecting v' uniformly at random from the set of neighbors of v in G. Moreover, a random element $\boldsymbol{\omega} \in \Omega_{\vec{e}_1}$ is generated similarly, only that we condition on $\boldsymbol{\omega}(\vec{p}_1) = \vec{e}_1$.

For fixed $\vec{e}_1 = (v_1, v_2) \in \vec{E}$ and fixed $\omega \in \Omega_{\vec{e}_1}$, we denote by $\mathbb{P}_{\vec{e}_1}(\omega)$ the probability of $\mathbb{P}(\boldsymbol{\omega} = \omega)$, for a random homomorphism $\boldsymbol{\omega}$ from the probability space $\Omega_{\vec{e}_1}$. It follows from the definition of $\Omega_{\vec{e}_1}$ that for every $\vec{e}_1 \in \vec{E}$ and every $\omega = (\vec{e}_1 = (v_1, v_1'), \ldots, \vec{e}_{t-1} = (v_{t-1}, v_{t-1}')) \in \Omega$, we have

$$\mathbb{P}_{\vec{e}_1}(\omega) = \prod_{j=2}^{t-1} \frac{1}{d_{v_j}} = \prod_{(v,v') \in \omega_-} \left(\frac{1}{d_v}\right)^{m_{\omega_-}(v,v')}. \tag{12}$$

We now estimate the quantity $\hat{N}_T(G) = |\Omega| = \sum_{\vec{e}_1 \in \vec{E}} |\Omega_{\vec{e}_1}|$. To that end, observe that

$$\frac{|\Omega|}{nd} = \sum_{\vec{e}_1 \in \vec{E}} \frac{|\Omega_{\vec{e}_1}|}{nd} = \sum_{\vec{e}_1 \in \vec{E}} \sum_{\omega \in \Omega_{\vec{e}_1}} \frac{\mathbb{P}_{\vec{e}_1}(\omega)}{nd} \left(\frac{1}{\mathbb{P}_{\vec{e}_1}(\omega)}\right),$$

where $\sum_{\vec{e}_1 \in \vec{E}} \sum_{\omega \in \Omega_{\vec{e}_1}} (\mathbb{P}_{\vec{e}_1}(\omega)/(nd)) = 1$. Applying the Arithmetic-Geometric Mean Inequality[1] and using (12), we obtain

$$\begin{aligned} \frac{|\Omega|}{nd} &= \sum_{\vec{e}_1 \in \vec{E}} \sum_{\omega \in \Omega_{\vec{e}_1}} \frac{\mathbb{P}_{\vec{e}_1}(\omega)}{nd} \left(\frac{1}{\mathbb{P}_{\vec{e}_1}(\omega)}\right) \\ &\ge \prod_{\vec{e}_1 \in \vec{E}} \prod_{\omega \in \Omega_{\vec{e}_1}} \left(\frac{1}{\mathbb{P}_{\vec{e}_1}(\omega)}\right)^{\mathbb{P}_{\vec{e}_1}(\omega)/(nd)} \\ &\overset{(12)}{=} \prod_{\vec{e}_1 \in \vec{E}} \prod_{\omega \in \Omega_{\vec{e}_1}} \prod_{(v,v') \in \omega_-} d_v^{m_{\omega_-}(v,v')\mathbb{P}_{\vec{e}_1}(\omega)/(nd)}. \end{aligned} \tag{13}$$

For a fixed $\vec{e} = (v, v') \in \vec{E}$, we now collect the terms $d_v^{\mathbb{P}_{\vec{e}_1}(\omega)/(nd)}$ in the triple product above. To that end, observe that a factor $d_v^{\mathbb{P}_{\vec{e}_1}(\omega)/(nd)}$ appears for

[1] The generalized AM-GM inequality states that for positive $x_1, \ldots, x_m$ and non-negative $c_1, \ldots, c_m$, where $c_1 + \cdots + c_m = 1$,

$$\sum_{i=1}^m c_i x_i \ge \prod_{i=1}^m x_i^{c_i}.$$

every $\vec{e_1} \in \vec{E}$, $\omega \in \Omega_{\vec{e}_1}$ and $2 \le s \le t-1$ for which $\omega(\vec{p}_s) = \vec{e}$. Set, therefore, for $s = 2, \dots, t-1$ and $\vec{e} = (v, v') \in \vec{E}$

$$g_s(\vec{e}) = \frac{1}{nd} \sum_{\vec{e}_1 \in \vec{E}} \sum \left\{ \mathbb{P}_{\vec{e}_1}(\omega)\colon\ \omega \in \Omega_{\vec{e}_1} \text{ and } \omega(\vec{p}_s) = \vec{e} \right\} \tag{14}$$

so that (13) is equivalently

$$|\Omega| \ge nd \prod_{\vec{e}=(v,v')\in\vec{E}} d_v^{\sum_{s=2}^{t-1} g_s(\vec{e})}. \tag{15}$$

In a moment, we prove that for a fixed $\vec{e} \in \vec{E}$ and $2 \le s \le t-1$,

$$g_s(\vec{e}) = \frac{1}{nd}. \tag{16}$$

Applying (16) to (15) yields

$$|\Omega| \ge nd \prod_{(v,v')\in\vec{E}} d_v^{\frac{t-2}{nd}} = nd \prod_{v\in V} d_v^{\frac{(t-2)d_v}{nd}},$$

as promised by Theorem 3. It remains to prove (16).

Proof of (16). Fix $\vec{e} \in \vec{E}$ and an integer s between 2 and $t-1$. We claim that $g_s(\vec{e})$ is the probability that $\vec{e}$ will appear at the s-th step of a randomly generated $\boldsymbol{\omega} \in \Omega$, i.e., $g_s(\vec{e}) = \mathbb{P}(\boldsymbol{\omega}(\vec{p}_s) = \vec{e})$ (where this probability occurs in the space Ω). Indeed,

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\omega}(\vec{p}_s) = \vec{e}) &= \sum_{\vec{e}_1\in\vec{E}} \mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_s) = \vec{e} \,\big|\, \boldsymbol{\omega}(\vec{p}_1) = \vec{e}_1\big) \mathbb{P}(\boldsymbol{\omega}(\vec{p}_1) = \vec{e}_1) \\
&= \frac{1}{nd} \sum_{\vec{e}_1\in\vec{E}} \mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_s) = \vec{e} \,\big|\, \boldsymbol{\omega}(\vec{p}_1) = \vec{e}_1\big) \\
&= \frac{1}{nd} \sum_{\vec{e}_1\in\vec{E}} \mathbb{P}_{\vec{e}_1}\big(\boldsymbol{\omega}(\vec{p}_s) = \vec{e}\big) \\
&= \frac{1}{nd} \sum_{\vec{e}_1\in\vec{E}} \sum \left\{ \mathbb{P}_{\vec{e}_1}(\omega)\colon\ \omega \in \Omega_{\vec{e}_1} \text{ and } \omega(\vec{p}_s) = \vec{e} \right\} \\
&= g_s(\vec{e}). \qquad (17)
\end{aligned}$$

We prove (16) by induction on s. We can easily extend the definition of $g_s(\vec{e})$ for the case $s = 1$ by setting

$$g_1(\vec{e}) = \frac{1}{nd} \sum_{\vec{e}_1 \in \vec{E}} \sum \left\{ \mathbb{P}_{\vec{e}_1}(\omega) \colon \omega \in \Omega_{\vec{e}_1} \text{ and } \omega(\vec{p}_1) = \vec{e} \right\} \tag{18}$$

and, clearly, $g_1(\vec{e}) = 1/(nd)$, which establishes the base case.

Assume, for each $1 \le r < s$ and each $\vec{e}\,' \in \vec{E}$, that $g_r(\vec{e}\,') = 1/(nd)$. Now, let $\vec{p}_s = (u_r, u_{s+1}) \in \vec{P}$, for some $1 \le r \le s$. Consider the arc $\vec{p}_{r-1} = (u_q, u_r) \in \vec{P}$, where $1 \le q \le r-1$. Observe that $\boldsymbol{\omega}(\vec{p}_s) = \vec{e} = (v, v')$ only if $\boldsymbol{\omega}(\vec{p}_{r-1}) = \vec{e}\,' = (v'', v)$ for some vertex v'' in the neighborhood $\Gamma_G(v)$ of v in the undirected graph G. Thus,

$$\begin{aligned} g_s(\vec{e}) &= \mathbb{P}(\boldsymbol{\omega}(\vec{p}_s) = \vec{e}) \\ &= \sum_{v'' \in \Gamma_G(v)} \mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_s) = \vec{e} \,\big|\, \boldsymbol{\omega}(\vec{p}_{r-1}) = (v'', v)\big) \mathbb{P}(\boldsymbol{\omega}(\vec{p}_{r-1}) = (v'', v)). \end{aligned}$$

For every $v'' \in \Gamma_G(v)$, the induction hypothesis gives

$$\mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_{r-1}) = (v'', v)\big) = \frac{1}{nd},$$

and by the definition of the probability space Ω, we have

$$\mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_s) = \vec{e} \,|\, \boldsymbol{\omega}(\vec{p}_{r-1}) = (v'', v)\big) = \frac{1}{d_v}.$$

Thus,

$$g_s(\vec{e}) = |\Gamma_G(v)| \cdot \frac{1}{d_v} \cdot \frac{1}{nd} = \frac{1}{nd},$$

as desired. □

4. Proof of Theorem 5

The proof is similar to the proof of Theorem 3 and we only outline the differences here. We modify the probability space Ω to include only labeled copies of P_3:

(i) the first arc $\vec{e}_1 = (v_1, v_2)$ is chosen uniformly at random in $\vec{E}$;
(ii) the second arc $\vec{e}_2 = (v_2, v_3)$ is selected uniformly among all arcs $(v_2, v) \in \vec{E}$ with $v \neq v_1$;

(iii) the third arc $\vec{e}_3 = (v_3, v_4)$ is selected uniformly among all arcs $(v_3, v) \in \vec{E}$ with $v \notin \{v_1, v_2\}$.

Note that the minimum degree condition of G is necessary to guarantee that the process above is feasible. Moreover, it is clear that $|\Omega| = N_{P_3}(G)$, since $N_{P_3}(G)$ denotes the number of labeled copies of P_3 in G.

In order to bound $|\Omega|$, we shall obtain an inequality along the lines of (13). Let $\omega = (\vec{e}_1 = (v_1, v_2), \vec{e}_2 = (v_2, v_3), \vec{e}_3 = (v_3, v_4)) \in \Omega$. Replacing (12) by the obvious bound

$$\mathbb{P}_{\vec{e}_1}(\omega) \geq \frac{1}{d_{v_2} - 2} \cdot \frac{1}{d_{v_3} - 2},$$

we obtain (using the arithmetic-geometric mean as before) the following analog of (13)

$$\frac{|\Omega|}{nd} \geq \prod_{\vec{e}_1 \in \vec{E}} \prod_{\omega \in \Omega_{\vec{e}_1}} \prod_{(v', v) \in \omega_-} (d_v - 2)^{\frac{\mathbb{P}_{\vec{e}_1}(\omega)}{nd}}, \tag{19}$$

where this time ω_- is the subpath $(\vec{e}_1, \vec{e}_2)$ of the path $\omega = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$. We remark that, in (19), there is a subtle difference in how we organized the products. In (19), the elements being multiplied are the degrees (minus two) of the *tail* vertex of each arc besides $\vec{e}_3$, while in (13), they are the degrees of the *head* vertex of each arc besides $\vec{e}_1$. The reason for this subtle change, will become clear in a moment (see (20) below). Moreover, note that the $m_{\omega_-(v',v)}$-term in the exponent of (13) does not appear in (19), since by definition of Ω no arc $\vec{e} \in \vec{E}$ can appear more than once in ω.

Let $g_s(\vec{e})$ for $s = 1, \ldots, 3$, be defined as in (14) and (18). Note that the right-hand side of (19) reduces to

$$\prod_{\vec{e}=(v',v) \in \vec{E}} (d_v - 2)^{g_1(\vec{e}) + g_2(\vec{e})}.$$

The reason for changing the definition of ω_- and multiplying degrees of tails instead of heads above is that, as we will show below,

$$g_1(\vec{e}) = g_2(\vec{e}) = \frac{1}{nd} \tag{20}$$

while $g_3(\vec{e})$ may be different than $1/nd$ depending on the structure of G (triangles in G may force $g_3(\vec{e}) \neq 1/nd$). In fact, this is the reason the current method does not seem to apply to paths of length four or larger.

In order to verify (20), we first note that by definition of $g_1(\vec{e})$ in (18), we have $g_1(\vec{e}) = \frac{1}{nd}$. Moreover, (17) extends to this case here so that $g_2(\vec{e}) = \mathbb{P}(\boldsymbol{\omega}(\vec{p}_2) = \vec{e})$ for every $\vec{e} \in \vec{E}$. The definition of Ω guarantees

$$\mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_2) = \vec{e}\,\big|\,\boldsymbol{\omega}(\vec{p}_1) = (v'', v')\big) = \frac{1}{d_{v'} - 1}$$

for every $\vec{e} = (v', v) \in \vec{E}$ with $v \neq v''$. We therefore have

$$\begin{aligned} g_2(\vec{e}) &= \mathbb{P}(\boldsymbol{\omega}(\vec{p}_2) = \vec{e}) \\ &= \sum_{v'' \in \Gamma_G(v') \setminus \{v\}} \mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_2) = \vec{e}\,\big|\,\boldsymbol{\omega}(\vec{p}_1) = (v'', v')\big)\mathbb{P}\big(\boldsymbol{\omega}(\vec{p}_1) = (v'', v')\big) \\ &= (d_{v'} - 1) \cdot \frac{1}{d_{v'} - 1} \cdot \frac{1}{nd} = \frac{1}{nd}, \end{aligned}$$

which proves (20).

Combining (19) and (20) we obtain

$$|\Omega| \geq nd \prod_{\vec{e} = (v', v) \in \vec{E}} (d_v - 2)^{\frac{2}{nd}} = nd \prod_{v \in V} (d_v - 2)^{\frac{2d_v}{nd}},$$

and Theorem 5 is proved. □

Acknowledgement

This work was done as a group project of all authors at a Focused Research Group meeting at the Banff International Research Station in May 2010. The authors thank BIRS for their support. We also thank Jacques Verstraëte for pointing out the relation of our work to [4].

References

[1] Alon, N., Hoory, S. and Linial, N. (2002). The Moore bound for irregular graphs. *Graphs Combin.* **18**(1) 53–57. MR1892433 1, 1, 1

[2] Chung, F. R. K., Graham, R. L. and Wilson, R. M. (1989). Quasi-random graphs. *Combinatorica* **9**(4) 345–362. MR1054011 1

[3] Conlon, D., Fox, J. and Sudakov, B. (2010). An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.* **20**(6) 1354–1366. MR2738996 1

[4] Erdős, P. and Simonovits, M. (1982). Compactness results in extremal graph theory. *Combinatorica* **2**(3) 275–288. MR0698653 1, 1, 4

[5] Hatami, H. (2010). Graph norms and Sidorenko's conjecture. *Israel J. Math.* **175** 125–150. MR2607540 1

[6] Sidorenko, A. F. (1991). Inequalities for functionals generated by bipartite graphs. *Diskret. Mat.* **3**(3) 50–65. MR1138091 1

[7] Sidorenko, A. F. (1993). A correlation inequality for bipartite graphs. *Graphs Combin.* **9**(2) 201–204. MR1225933 1, 1

[8] Simonovits, M. (1984). Extremal graph problems, degenerate extremal problems and super-saturated graphs. In: *Progress in Graph Theory* (Waterloo, Ont., 1982), Academic Press, Toronto, ON, pp. 419–437. MR0776819 1

[9] Skokan, J. and Thoma, L. (2004). Bipartite subgraphs and quasi-randomness. *Graphs Combin.* **20**(2) 255–262. MR2080111 1

[10] Thomason, T. (1989). Dense expanders and pseudo-random bipartite graphs. In: *Graph Theory and Combinatorics* (Cambridge, 1988), *Discrete Math.* **75**(1–3) 381–386. MR1001409 1

D. Dellamonica Jr.
Department of Mathematics and Computer Science
Emory University
Atlanta, GA 30322
USA
E-mail address: ddellam@mathcs.emory.edu

P. Haxell
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, ON N2L 3G1
Canada
E-mail address: pehaxell@math.uwaterloo.ca

T. Łuczak
Department of Discrete Mathematics
Faculty of Mathematics and Computer Science
Adam Mickiewicz University
Umultowska 87, 61-614 Poznań
Poland
E-mail address: tomasz@amu.edu.pl

Department of Mathematics and Computer Science
Emory University
Atlanta, GA 30322
USA

D. Mubayi
Department of Mathematics, Statistics, and Computer Science
University of Illinois
Chicago, IL 60607
USA
E-mail address: mubayi@math.uic.edu

B. Nagle
Department of Mathematics and Statistics
University of South Florida
Tampa, FL 33620
USA
E-mail address: bnagle@math.usf.edu

Y. Person
Institut für Mathematik
Freie Universität Berlin
Arnimallee 3, D-14195 Berlin
Germany
E-mail address: person@zedat.fu-berlin.de

V. Rödl
Department of Mathematics and Computer Science
Emory University
Atlanta, GA 30322
USA
E-mail address: rodl@mathcs.emory.edu

M. Schacht
Fachbereich Mathematik
Universität Hamburg
Bundesstrasse 55, D-20146 Hamburg
Germany
E-mail address: schacht@math.uni-hamburg.de

Received December 8, 2010