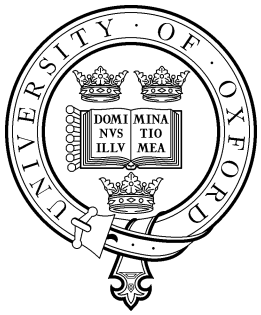
Trees and Graphs: Congestion, Polynomials and Reconstruction



Hiu-Fai Law

St. Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2011

Abstract

Spanning tree congestion was defined by Ostrovskii (2004) as a measure of how well a network can perform if only minimal connection can be maintained. We compute the parameter for several families of graphs. In particular, by partitioning a hypercube into pieces with almost optimal edge-boundaries, we give tight estimates of the parameter thereby disproving a conjecture of Hruska (2008). For a typical random graph, the parameter exhibits a zigzag behaviour reflecting the feature that it is not monotone in the number of edges. This motivates the study of the most congested graphs where we show that any graph is close to a graph with small congestion.

Next, we enumerate independent sets. Using the independent set polynomial, we compute the extrema of averages in trees and graphs. Furthermore, we consider inverse problems among trees and resolve a conjecture of Wagner (2009). A result in a more general setting is also proved which answers a question of Alon, Haber and Krivelevich (2011).

After briefly considering polynomial invariants of general graphs, we specialize into trees. Three levels of tree distinguishing power are exhibited. We show that polynomials which do not distinguish rooted trees define typically exponentially large equivalence classes. On the other hand, we prove that the rooted Ising polynomial distinguishes rooted trees and that the Negami polynomial determines the subtree polynomial, strengthening results of Bollobás and Riordan (2000) and Martin, Morin and Wagner (2008). The top level consists of the chromatic symmetric function and it is proved to be a complete invariant for caterpillars.

Acknowledgements

I am very grateful to my supervisor Alex Scott for his encouragement and guidance. I would like to thank Colin McDiarmid and John Talbot for examining the thesis and their valuable comments. I would also like to thank Mikhail Ostrovskii for a series of fruitful joint work and Bhalchandra Thatte for helpful discussions and aid in programming. Support from the other members of the group and friends beyond, Matt, Kitty, Tom, Atsushi, Puck, Oliver, Lutz, Philip, Kerstin and Paul has managed to keep me afloat, for that I thank them.

Thanks are due to the generous support of the Croucher Foundation and the administrator Ms. Elaine Sit who has been most supportive throughout the years.

Finally, I thank my parents and brothers for their love and care.

Authorship

Chapter 1 is joint work with Mikhail Ostrovskii, [59, 60]. Chapters 3, 4 are joint work with Alex Scott.

Unless duly acknowledged, the rest represents my own work from Michaelmas 2007 to Trinity 2011.

To my parents

Contents

1 Spanning tree congestion **1**

1.1 Tools and Examples 4

1.1.1 Duality for planar graphs 4

1.1.2 Edge-isoperimetry 7

1.1.3 Platonic graphs and cross-polytopes 8

1.2 Hypercube 10

1.2.1 Upper bound for $s(Q_d)$ 10

1.2.2 Construction of a spanning tree 11

1.2.3 Estimates of the edge congestion 13

1.2.4 Lower bound for $s(Q_d)$ 14

1.3 Cartesian products of complete graphs 18

1.3.1 Hamming graphs 18

1.3.2 Two-dimensional 19

1.4 Grid and torus 20

1.4.1 Lower bound 21

1.4.2 Upper bound 22

1.5 Complete multipartite graphs 25

1.6 Random graphs 28

1.6.1 Spanning tree congestion 28

1.6.2 Tree congestion 35

1.7 Maximum spanning tree congestion 36

1.8 Maximum spanning tree congestion of planar graphs 41

1.8.1 Estimates from above for planar graphs 42
1.8.2 Planar graphs with large spanning tree congestion 43

2 Enumeration of independent sets **45**
2.1 Preliminaries 46
2.2 Average order of an independent set 48
2.2.1 Trees 49
2.2.2 Graphs 51
2.3 Inverse problems 53
2.3.1 The number of independent sets 57
2.3.2 General F-matchings 58
2.3.3 Asymptotic behaviour of divisibility properties 63
2.3.4 Trees in other residue classes 65
2.3.5 Maximum number of (F, r)-matchings 68

3 Graph reconstruction with polynomials **70**
3.1 Definitions 71
3.2 Decorations 72
3.3 Independent set polynomial 73
3.4 Ising polynomial 74
3.4.1 Magnetically unique graphs 76
3.4.2 Isomagnetic graphs 77

4 Tree reconstruction with polynomials **81**
4.1 Level one polynomials 82
4.1.1 Independent set polynomial 82
4.1.2 Matching polynomial 83
4.1.3 (F, r)-matching polynomials 83
4.1.4 Reliability polynomial 83
4.1.5 Path sequences 84
4.1.6 Dominating set polynomial 85

4.2 Switching . . . 86
4.2.1 Unlabelled trees . . . 88
4.2.2 Labelled trees . . . 90
4.3 Level two polynomials . . . 91
4.3.1 Ising polynomial and trees . . . 93
4.3.2 Negami and subtree polynomials . . . 95
4.4 Level three polynomials . . . 97

5 Independent set polynomial of trees **101**
5.1 Spiders . . . 102
5.2 Well-covered trees . . . 103
5.3 Independent set polynomial deck . . . 106
5.3.1 Information encoded . . . 106
5.3.2 Graphs with the same I-deck . . . 107
5.3.3 Large family of trees sharing the same I-deck . . . 109

Summary and future work **111**

Chapter 1

Spanning tree congestion

Let G be a connected graph and T be a spanning tree. For $e = uv \in E(G)$, let the *detour* of e be the unique path joining u and v in T. The *congestion* of $e \in E(T)$, denoted by $\mathrm{ec}_G(e, T)$ (or $\mathrm{ec}(e, T)$ if G is clear), is the number of detours containing it. The *edge congestion* $\mathrm{ec}(G : T)$ of T is the maximum congestion over all edges of T. The *spanning tree congestion* $s(G)$ is the minimum of edge congestion of T over all spanning trees. A spanning tree T in G satisfying $\mathrm{ec}(G : T) = s(G)$ is called a *minimum congestion spanning tree*.

We rephrase the definition of congestion in terms of edge-boundary. For $S \subset V(G)$, the *edge-boundary* ∂S is defined as $E(S, S^c) = \{uv \in E(G) \mid u \in S, v \notin S\}$. Denote the connected components of $T - e$ by A_e and B_e, then $\mathrm{ec}(e, T) = |\partial A_e| = |\partial B_e|$. Writing $\deg_S(v)$ for the degree of a vertex v in the subgraph $G[S]$ induced by S, we have

$$\mathrm{ec}(e, T) = \sum_{v \in A_e} \deg_G(v) - \deg_{A_e}(v) = \sum_{v \in A_e} \deg_G(v) - 2e(A_e). \tag{1.1}$$

Spanning tree congestion was introduced by Ostrovskii [78] in 2004, though some versions had been studied before, see [54, 90]. See also [8, 54] and references therein for other related parameters. Since then there are a number of papers studying the parameters in different classes of graphs, [18, 19, 20, 25, 47, 48, 55, 57, 70, 79, 93], where many interesting results were obtained.

The etymology of the definitions was explained in [20] as follows. Imagine that

edges of G are roads, and edges of T are those roads which are cleaned from snow after snowstorms. If we assume that each edge in G bears the same amount of traffic between its vertices, and that after a snowstorm each driver takes the corresponding (unique) detour in T, then $\mathrm{ec}(G : T)$ describes the traffic congestion at the most congested road of T.

The definition of $s(G)$ can also be phrased in terms of layout. A *layout* of G into a *host graph* H consists of an injection $\rho : V(G) \to V(H)$ and an injection of $E(G)$ into paths of H such that $\rho(uv)$ ends with $\rho(u)$ and $\rho(v)$. The number of $\rho(uv)$ containing an edge $e \in E(H)$ is the congestion of e, maximum of which is the edge congestion of H. Then we consider the minimum of edge congestion over all host graphs within some family. For spanning tree congestion, H is a spanning tree of G and ρ restricts to the identity on $V(G)$. Well studied cases include *linear cutwidth* when H is a path; *cyclic cutwidth* when H is a cycle. Layout problems are studied in connection with circuit designs; see [27] for a survey.

The related parameter of tree congestion was also considered in [78]. In this case, H is a tree of order $|G|$ and the congestion of e in H is equal to $|\partial A_e|$ where A_e is a component of $H - e$. Since H is not necessarily a subgraph, A_e may not be connected in G. The *tree congestion* $t(G)$ is defined as the minimum edge congestion among all trees of order $|G|$.

Our focus is on spanning tree congestion but it is useful to contrast with tree congestion. Ostrovskii [78] gave a characterization of $t(G)$ as follows. Let $m(u, v)$ be the maximum number of edge-disjoint paths joining two vertices u, v of G. Then $t(G) = \max_{u,v} m(u, v)$. The lower bound can be obtained by the max-flow-min-cut theorem since any edge of a tree T induces a cut in G of size at least $m(G)$; whereas the upper bound can be obtained by a Gomory-Hu tree which is a tree such that each edge $e = uv$ induces a minimum cut separating u, v in G. Thus, it is easily seen that $t(P_n^d) = 2d$ for $n > 3$ and $t(P_3^d) = 2d - 1$ where P_n^d is the d-dimensional grid on a path of n vertices. Moreover, simple counting in [78] shows that the tree congestion of a k-regular graph is k. Thus $t(Q_d) = d, t(\mathbb{Z}_n^d) = 2d$ where $Q_d, \mathbb{Z}_n^d$ are the d-dimensional hypercube and torus respectively. However, computation of $s(Q_d)$

appears more difficult. In fact, only a few exact values of $s(Q_d), s(\mathbb{Z}_n^d), s(P_n^d)$ are known. For such a graph, we partition it into close-to-optimal blocks, construct a spanning tree for each block and then join them to a centroid, see Sections 1.2 to 1.4. This way we obtain estimates tight up to a constant factor. In Section 1.5 we apply edge-isoperimetry of a complete multipartite graph G, which was studied by Piliposyan and Muradyan [82], to compute $s(G)$.

Since $t(G)$ can be found by solving $|G| - 1$ flow problems for which there are efficient algorithms (see [33]), the problem of deciding if $t(G) \leq k$ given integer k, is solvable in polynomial time. On the other hand, the complexity of computing spanning tree congestion is quite different. Otachi, Bodlaender and van Leeuwen [80] showed that for fixed d, k, the problem of $s(G) \leq k$ can be solved in linear time for graphs with $\Delta(G) \leq d$; but it is NP-hard in general even for planar graphs.

Spanning tree congestion differs from tree congestion or linear cutwidth in another way, namely, it is not monotone in the number of edges. Deleting edges does not increase tree congestion or linear cutwidth but it may increase spanning tree congestion, because a 'good' spanning tree may be deleted. For example, $s(K_{2n} - nK_2) = 4n - 6$ whereas $s(K_{2n}) = 2n - 1$, see [78]; another example is a typical dense random graph where we a.a.s. have $s(\mathcal{G}(n,p)) \sim \lceil 1/p \rceil pn$, see Section 1.6. This makes it interesting to study

$$\mu(n) = \max\{s(G) : \ G \text{ is a graph of order } n\}.$$

Examples in [78] show that $\mu(n) = \Omega(n^{3/2})$; and it was conjectured that $\mu(n) = o(n^2)$. This is in contrast with tree congestion where we always have $t(G) \leq \Delta(G) \leq n - 1$.

While a typical graph has linear order spanning tree congestion, there are much more congested graphs. How far is a congested graph from a less congested one? We answer this in Section 1.7 with Szemerédi's Regularity Lemma, thereby proving the conjecture in [78] that $\mu(n) = o(n^2)$. Subsequently, the correct order was determined by Löwenstein, Rautenbach and Regen [70] that $\mu(n) = \Theta(n^{3/2})$.

Within the family of planar graphs, we give tight bounds on the maximum spanning tree congestion in Section 1.8.

Sections 1.3, 1.4, 1.5 and 1.8 are joint work with M.I. Ostrovskii.

1.1 Tools and Examples

Upper bounds on spanning tree congestion can be derived from explicit constructions of which we will see many examples later. For lower bounds, we have two tools, namely, duality and isoperimetry. Duality is further developed by Ostrovskii [79] to compute the spanning tree congestion of triangular grids (see also [59]). A first estimate is given in [78]

$$\frac{e(G)}{|G|-1} \leq s(G) \leq e(G) - |G| + 2.$$

1.1.1 Duality for planar graphs

In this section, we exploit duality of planar graphs to give another formulation of congestion.

Recall that a graph is *planar* if it can be drawn on the plane such that no two edges cross. A particular drawing is called a *plane* graph. If we delete the edges and vertices in a plane graph, then the plane becomes a union of connected components which are called the *faces*. The *boundary* of a face is the set of edges in its closure. The Jordan Curve Theorem states that a simple closed curve separates the plane into two components. Thus, each edge of a cycle is always in the boundary of two faces. Any planar graph has a unique unbounded face.

In general, a planar graph may be realized by different drawings. To circumvent this complication and to ensure that the dual of a graph is a graph, we restrict ourselves to planar graphs which are 3-connected. For a proof the following theorem, see for example [28].

Theorem 1.1. *A 3-connected planar graph has a unique collection of the sizes of the face boundaries over all drawings on the plane.*

By Theorem 1.1 and the definition of 3-connectivity, it follows that the boundaries of two faces share at most one edge. Furthermore, since there are no bridges, any edge is on the boundary of two faces.

Next, we recall the notion of a dual. Let G be a 3-connected planar graph. Let the sets of vertices, edges, faces of G be denoted by $V(G), E(G), F(G)$ respectively. The *dual graph* G^* of G is defined by $V(G^*) = F(G)$ and F_1, F_2 are adjacent in G^* iff they share an edge in G. For a spanning tree T of G, the dual tree T^* is defined as: $e = F_1 \cap F_2 \in T$ iff $e^* = F_1^* F_2^* \notin T^*$.

The following results are well known, see for example [28, chapter 4].

Lemma 1.2. *The dual graph of a 3-connected planar graph is a 3-connected planar graph.*

Lemma 1.3. *A subgraph T of G is a spanning tree iff T^* is a spanning tree of G^*.*

By Lemma 1.2, G^* is also a 3-connected planar graph, so that we can consider $(G^*)^*$. For a spanning subgraph T of G, let e be an edge in G but not in T. Write $T+e$ for the subgraph of G with edge set $E(T) \cup e$ and $T-e$ for that with $E(T-e) = E(T) \setminus \{e\}$.

Lemma 1.4. *For an edge e of a spanning tree T of a 3-connected planar graph G, $\mathrm{ec}(e, T)$ is equal to the length of the cycle in $T^* + e^*$.*

Proof. Note that $\mathrm{ec}(e, T) = e(A_e, B_e)$ is equal to the number of edges each of which when added to $T - e$ gives a connected subgraph of G.

Since T is a spanning tree, by Lemma 1.3, T^* is a spanning tree. Consider the unique cycle in $T^* + e^*$. Removal of any edge f^* of this cycle breaks the cycle but does not disconnect $T^* + e^*$. Hence $T^* + e^* - f^*$ is a spanning tree. Again by Lemma 1.3, $T - e + f$ is connected which implies the result. □

In view of Lemma 1.4, we define the dual notion of congestion in planar graphs. For a spanning tree T^* of G^*, define the *cycle-stretch* of $e^* \in E(T^*)$ to be the length of the cycle in $T^* + e^*$. Let the *cycle-stretch* $\mathrm{cs}(G^* : T^*)$ of T^* be the maximum of the cycle-stretch over all edges. Let the *cycle-stretch* $\mathrm{cs}(G^*)$ be the minimum of $\mathrm{cs}(G^* : T^*)$ over all spanning trees of G^*. By Lemma 1.4,

$$\mathrm{cs}(G^*) = s(G). \tag{1.2}$$

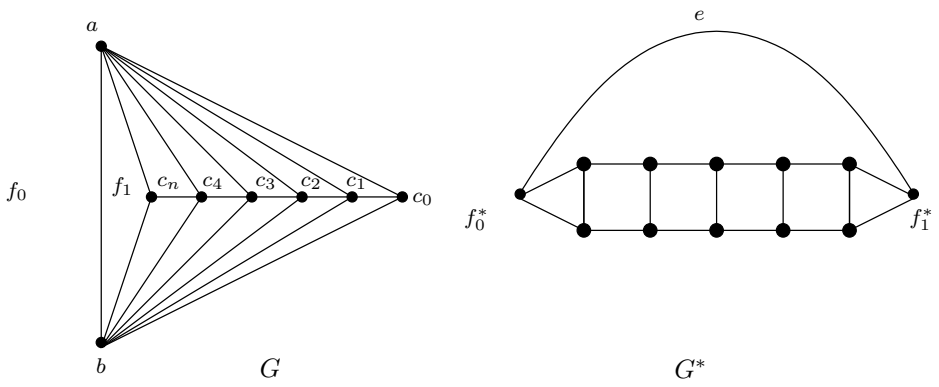


Figure 1.1: A triangulation and its dual

This dual version has also been observed in [79]. We remark that the formulation of congestion in terms of cycle-stretch applies to matroids which may be of interest for further studies. Another remark is that for general graphs there is the notion of *stretch*, see [83], which measures the maximum distortion of distance when a graph is embedded into another. This might open up possibilities of other estimates of $s(G)$.

We illustrate (1.2) by an example. Recall that a *triangulation* is a planar graph such that each face is a triangle.

Lemma 1.5. *Let G be a triangulation defined by $V(G) = \{a, b, c_0, c_1, \cdots, c_n\}$ and $E(G) = \{ab, bc_0, ac_0, ac_i, bc_i, c_{i-1}c_i : i = 1, \cdots, n\}$ (see Figure 1.1). Then $s(G) = n + 2$.*

Proof. Note that G is a 3-connected graph with $n+3$ vertices, $2n+2$ faces and $3n+3$ edges. Let the bounded face whose boundary contains ab be f_1 and the unbounded face be f_0.

As the second shortest path between f_0^* and f_1^* in G^* has length $n+1$, $\mathrm{cs}(G^* : T^*) \geq n+2$ for any spanning tree T^*. The lower bound is attained by the spanning tree which contains a path from f_0^* to f_1^* of length $n+1$ and edges so that all other vertices are leaves. By (1.2), $s(G) = n + 2$. □

We remark that $s(G^*) = 3$ so that the difference between $s(G)$ and $s(G^*)$ can be large.

1.1.2 Edge-isoperimetry

The second tool is edge-isoperimetry which applies to general graphs. We need some definitions. Recall that for $S \subset V(G)$, the edge-boundary $\partial S = E(S, S^c)$ is the set of edges with one endvertex in S and the other in S^c. To solve an *edge-isoperimetric problem (EIP)* is to compute for a given graph G and positive integer k,

$$\theta_G(k) = \min\{|\partial S| : |S| \subset V(G), |S| = k\}. \tag{1.3}$$

See Harper's monograph [44] or Leader's survey [61] for a reference to EIP. Since we are interested in connected subgraphs, the following function is more relevant. Let

$$f_G(k) = \min\{|\partial S| : \ S \subset V(G), \ |S| = k, \ G[S] \text{ is connected}\}. \tag{1.4}$$

In many cases the two functions are identical, but not for some multipartite graphs, see Section 1.5.

Sets minimizing the edge-boundary can sometimes be characterized. For example, in the hypercube, there exists a linear ordering of the vertices such that an initial segment I of length k characterizes sets of vertices of size k with minimum edge-boundaries. Such an ordering is often found with techniques of compression or stability, see [61, 7, 44].

We proceed to the notion of centroid which originates from Jordan [52]. The *weight at* u is the largest order of a component of $T - u$. A vertex u is called a *centroid vertex* if its weight is minimal. The *weight* $w(T)$ *of* T is defined as the weight at a centroid vertex. Note that either there is a centroid vertex or two adjacent centroid vertices in which case $2w(T) = |T|$.

For a spanning tree T of G, a characterization of $w(T)$ is given in [78]; we include a proof for completeness. Let A_e and B_e denote the components of $T - e$ for $e \in E(T)$.

Lemma 1.6. $w(T) = \max\{\min\{|A_e|, |B_e|\} : e \in E(T)\}$.

Proof. Suppose $e = uv$ is an edge incident with a centroid u. Assume $u \in A_e$. Then $|A_e| \geq |B_e|$; otherwise, the weight of T at u is $|B_e|$ while that at v is at most $|B_e| - 1$,

which contradicts that u is a centroid. Hence,

$$\begin{aligned} w(T) &= \max\{\min\{|A_e|, |B_e|\} : e \in E(T) \text{ is incident with } u\} \\ &\leq \max\{\min\{|A_e|, |B_e|\} : e \in E(T)\}. \end{aligned}$$

For an edge e in a component C of $T - u$, $\min\{|A_e|, |B_e|\} \leq |C| \leq w(T)$. Hence, $\max\{\min\{|A_e|, |B_e|\} : e \in E(T)\} \leq w(T)$. □

Edge-isoperimetry is related to spanning tree congestion in the following way. As there are at most $\Delta(T)$ components when a centroid is deleted,

$$\left\lceil \frac{|V(G)| - 1}{\Delta(G)} \right\rceil \leq w(T) \leq \frac{|G|}{2}. \tag{1.5}$$

Let $c_G = \left\lceil \frac{|V(G)|-1}{\Delta(G)} \right\rceil$. By Lemma 1.6, there is an edge in any spanning tree T disconnecting a component of order at least c_G. Thus

$$s(G) \geq \min_{c_G \leq k \leq |G|/2} f_G(k). \tag{1.6}$$

This inequality was observed in [78], where it was used to estimate the spanning tree congestion from below using the Cheeger constants and isoperimetric dimension constants. In [55] and [57] inequality (1.6) was combined with the edge-isoperimetric results for hypercubes from [23]. In [20] and [55] inequality (1.6) was combined with the edge-isoperimetric results for cubic grids from [11].

1.1.3 Platonic graphs and cross-polytopes

We illustrate the use of duality and edge-isoperimetry by computing the spanning tree congestion of the Platonic graphs and the cross-polytopes. Denote the tetrahedron by K_4, the octahedron by $\maltese_3$, the cube by Q_3, the dodecahedron by $\mathscr{D}$ and the icosahedron by $\mathscr{I}$. The solution of the EIP for $\mathscr{D}$ and $\mathscr{I}$ can be found in [44].

Proposition 1.7. *$s(K_4) = 3, s(\maltese_3) = 6, s(Q_3) = 4, s(\mathscr{D}) = 6$ and $s(\mathscr{I}) = 10$.*

Proof. We prove here the lower bounds; matching upper bounds are obtained from the spanning trees in Figure 1.2.

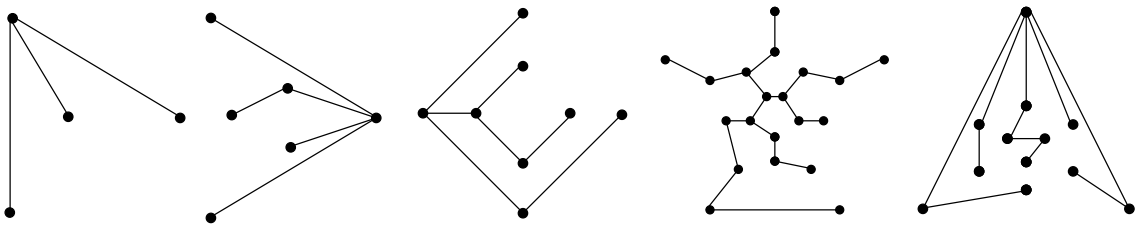


Figure 1.2: Minimum congestion spanning trees of the Platonic graphs

1. The tetrahedron is self-dual with girth 3; so that (1.2) implies the result.

2. Let T be a spanning tree of $\boxplus_3$ such that $\mathrm{cs}(\boxplus_3^* : T^*) < 6$. Since $Q_3 = \boxplus_3^*$ has no C_3 or C_5, $\mathrm{cs}(Q_3 : T^*) = 4$. Since each edge of T needs 3 edges of T^* to complete to a copy of C_4, a total of $5 \times 3 = 15$ edges in T^* is needed (counting multiplicity). However, each edge of Q_3 is on two squares so that only $7 \times 2 = 14$ edges are available in T^*. So $s(\boxplus_3) \geq 6$.

3. Let T be a spanning tree of Q_3 such that $\mathrm{cs}(\boxplus_3 : T^*) = 3$. Then each edge of T needs 2 edges of T^* to complete to a triangle, so that at least 14 edges are needed. However, since each edge of the octahedron is on two triangles, only 10 edges available. Hence, $s(Q_3) \geq 4$.

4. By (1.5), $w(T) \geq 7$ for any spanning tree T of $\mathscr{D}$. Using the table below and (1.6), we have $s(\mathscr{D}) \geq 6$.

5. By (1.5), $w(T) \geq 3$ for any spanning tree T of $\mathscr{I}$. Using the table below, we conclude by (1.6) that $s(\mathscr{I}) \geq 9$. However, if $\mathrm{ec}(\mathscr{I} : T) \leq 9$, then for each $e \in E(T)$, $T - e$ has a component inducing a K_n for $n \leq 3$. In particular, the depth of T rooted at a centroid is 2, contradicting the fact that the radius of $\mathscr{I}$ is 3. So $s(\mathscr{I}) \geq 10$.

$\lvert S\rvert$	5	6	7	8	9	10
$\min \lvert\partial_{\mathscr{D}} S\rvert$	5	6	7	6	7	6

$\lvert S\rvert$	2	3	4	5	6
$\min \lvert\partial_{\mathscr{I}} S\rvert$	8	9	10	11	10

□

The *cross-polytope* $\maltese_d$ has $V(\maltese_d) = \{\pm e_i, i = 1, 2, \cdots, d\}$ where $\{e_1, \cdots, e_d\}$ forms the standard basis of $\mathbb{R}^d$ and $E(\maltese_d)$ consists of all pairs except $\{e_i, -e_i\}$. Note that $\maltese_d \cong K_{2d} - dK_2$. In [78], $s(\maltese_d)$ was computed directly.

Proposition 1.8. *For every $d > 1$, $s(\maltese_d) = 4d - 6$.*

Proof. The EIP of $\maltese_d$ has a nested solution $(e_1, e_2, \cdots, e_d, -e_d, -e_{d-1}, \cdots, -e_1)$, see [44]. This means that for any set A of $a \leq d$ vertices, $|\partial A| \geq |\partial\{e_1, e_2, \cdots, e_a\}| = (2d-2)a - a(a-1)$. As $a(2d-1-a)$ is nondecreasing in $a \leq d$ and $c_{\maltese_d} = 2$, by (1.6), $s(\maltese_d) \geq 2(2d-3)$. The lower bound is attained by any spanning tree of depth 2 rooted at a centroid whose degree is $2d-2$. □

1.2 Hypercube

The d-dimensional hypercube Q_d is the graph whose vertices are the 0-1 vectors of length d; and two vertices are adjacent if they differ at exactly one coordinate. Hruska [47] conjectured that $s(Q_d) = 2^{d-1}$ for any integer $d \geq 1$. In this section we disprove the conjecture by showing that

$$s(Q_d) = \Theta\left(\frac{\log_2 d}{d} 2^d\right).$$

More precisely, we show that $s(Q_d) = 2^{d-1}$ for $1 \leq d \leq 6$, and for $d > 7$,

$$\frac{\log_2 d}{d} 2^d \leq s(Q_d) \leq \frac{\lfloor \log_2 d \rfloor}{2^{\lfloor \log_2 d \rfloor}} 2^d + (d-2)\lfloor \log_2 d \rfloor + (d - \lfloor \log_2 d \rfloor)(d - 2\lfloor \log_2 d \rfloor).$$

For $d = 7$, our general construction yields a tree with edge congestion larger than 2^6; but with ad-hoc construction, we have $59 \leq s(Q_7) \leq 60$. Note that our upper bound is smaller than 2^{d-1} when $d > 7$. This section is based on [57].

1.2.1 Upper bound for $s(Q_d)$

Our aim is to construct spanning trees of Q_d for $d \geq 8$ with edge congestion smaller than what Hruska [47] conjectured $s(Q_d) = 2^{d-1}$. The procedure is as follows. Let $k = \lfloor \log_2 d \rfloor$. By fixing the first k entries, we decompose Q_d into 2^k subcubes of

dimension $(d-k)$. We use a spanning tree of Q_k to build a tree that connects the 2^k subcubes of Q_d. The latter tree is then extended to a spanning tree of Q_d by constructing spanning trees in each of the $(d-k)$-dimensional subcubes.

We shall adopt the following notation. For any subgraph G of Q_m and $v \in Q_n$, we write (G, v) for the subgraph of Q_{m+n} obtained by adding v after the last entry of each vertex of G. For any integer $n > 0$, let 0_n denote the vector of n zeros and $e(i, n)$ be the vector of length n whose only nonzero entry is 1 at the ith entry. Let $e_i = e(i, d-k)$.

The following simple fact is useful when we estimate congestion.

Lemma 1.9. *Let $f(x) = (d-x)2^x$. Then f is increasing for $x \leq d-2$ and $f(d-1) = f(d-2)$.*

Proof. Since $f'(x) = (d-x)2^x \ln 2 - 2^x = 2^x(d \ln 2 - x \ln 2 - 1) \geq 2^x(2 \ln 2 - 1) > 0$, f is increasing for $x \leq d-2$. □

We build a family of spanning trees which is a major constituent of our spanning trees of Q_d. Let $T_1 = Q_1 = \{0, 1\}$. For $d > 1$, let $T_d = (T_{d-1}, 0) \cup (T_{d-1}, 1) \cup \{0_d, e(d, d)\}$. Then T_d is a spanning tree of Q_d for all $d > 1$. We call T_d the *standard spanning tree* of Q_d *centered* at 0_d. Observe that when an edge of T_d is deleted, there is always a subcube component (note that $Q_0 \cong K_1$ and $Q_1 \cong K_2$). Hence, by (1.1), the congestion of an edge has the form $(d-i)2^i$ which is maximized when $i = d-1$ by Lemma 1.9. Thus $\mathrm{ec}(Q_d, T_d) = 2^{d-1}$ and so $s(Q_d) \leq 2^{d-1}$.

1.2.2 Construction of a spanning tree

For $d \geq 8$, let $k = \lfloor \log_2 d \rfloor$. Consider $Q_d = Q_k \times Q_{d-k}$. More precisely, let $Q_d = \bigcup_{x \in Q_k} \mathcal{Q}^x$ where $\mathcal{Q}^x = \{(x, y) : y \in Q_{d-k}\}$ for each $x \in Q_k$. For each $x \in Q_k$, let the *corner* be $(x, 0_{d-k})$. We shall build a subdivided star whose leaves lie in distinct subcubes of Q_d. Then we span the remaining vertices of each subcube by a standard spanning tree centered at such a leaf.

Let T be a spanning tree of Q_k rooted at $u = 0_k$ such that $d_T(u) = k$ and the depth of T is k. Note that the depth is the smallest possible as Q_k has radius k. For

definiteness, we take T to be a standard spanning tree of Q_k centered at u. Clearly, T has at least k leaves and any vertex other than u has fewer than $2^k - k$ descendants. This will imply that we do not use all the neighbours of a corner when we subsequently construct paths to join the subcubes of Q_d.

Let $v_1, \cdots, v_k$ be k leaves of T such that all the u–v_i paths in T are edge-disjoint. For each $1 \leq i \leq k$, we construct a path from $r = 0_d$ to $(v_i, 0_{d-k})$ in Q_d. Indeed, let $P(v_i, 0_{d-k})$ be the path $(u, 0_{d-k})$ $(x_1, 0_{d-k})$ $(x_2, 0_{d-k}) \cdots (v_i, 0_{d-k})$ where $ux_1x_2 \cdots v_i$ is the path from u to v_i in T. Next, consider the remaining $2^k - k$ vertices of T, namely, $q_1 = u, \cdots, q_{2^k-k}$. For $2 \leq j \leq 2^k - k$, let $P(q_j, e_j)$ be the path $(u, 0_{d-k})$ (u, e_j) (x_1, e_j) $(x_2, e_j) \cdots (q_j, e_j)$ where $ux_1x_2 \cdots q_j$ is the path from u to q_j in T. For $q_1 = u$, let $P(q_1, e_1)$ be the path $(u, 0_{d-k})$ (u, e_1).

Let *the frame* $\mathcal{F}$ be $\bigcup_{i=1}^{k} P(v_i, 0_{d-k}) \cup \bigcup_{j=1}^{2^k-k} P(q_j, e_j)$. As the paths P are edge-disjoint and share only u, every leaf of $\mathcal{F}$ lies in a unique subcube $\mathcal{Q}^x$. We call these leaves *connecting vertices* because they are used to connect the subcubes to r. Note that every vertex in $\mathcal{F}$ is either a corner or a neighbour of one. Moreover, since we reserve the corners for constructing paths corresponding to our initial set of k leaves of T, when a corner is a connecting vertex, none of the other vertices of the corresponding subcube is in $\mathcal{F}$.

The next step of the construction is to attach a spanning tree at the connecting vertex for each $\mathcal{Q}^x$. When the connecting vertex is a corner, a standard spanning tree centered at the corner suffices. Otherwise, the connecting vertex is adjacent to a corner, in which case we slightly modify the standard spanning tree. Note that the corner and some (but not all) of its neighbours may lie on paths of $\mathcal{F}$.

Lemma 1.10. *For any edge $\{u, v\}$ of Q_n, there is a standard spanning tree centered at u such that v and its neighbours other than u are leaves.*

Proof. Since Q_n is edge-transitive, we may assume $u = 0_n$ and $v = e(1, n)$. We claim that T_n is a required standard spanning tree. It is clearly true for $n \leq 2$. Suppose it is true for n. By construction, $T_{n+1} = (T_n, 0) \cup (T_n, 1) \cup \{0_{n+1}, e(n+1, n+1)\}$. It follows by induction that $e(1, n+1)$ and n of its neighbours which are in $(T_n, 0)$ are

leaves of T_{n+1}. For the remaining neighbour $e(1, n+1) + e(n+1, n+1)$, being a leaf of $(T_n, 1)$ makes it a leaf of T_{n+1}. □

Suppose the connecting vertex of $\mathcal{Q}^x$ is (x, e_1) which is adjacent to the corner $(x, 0_{d-k})$. By Lemma 1.10, there is a standard spanning tree centered at (x, e_1) such that $(x, 0_{d-k})$ and all (x, e_i) where $2 \leq i \leq d-k$ are leaves. We delete any of these leaves if it lies in $\mathcal{F}$. Thus we obtain a tree $\mathcal{S}^x$ covering all vertices of $\mathcal{Q}^x$ except those vertices of degree 2 on $\mathcal{F}$.

Applying the same procedure to each subcube of Q_d, we obtain a spanning tree $\mathcal{T} = \bigcup_{x \in Q_k} \mathcal{S}^x \cup \mathcal{F}$ of Q_d.

1.2.3 Estimates of the edge congestion

Let m be the number of vertices of $\mathcal{Q}^x$ which are in $\mathcal{F}$, where $m \leq 2^k - k \leq d - k$. For the edge e of $\mathcal{F}$ incident with the connecting vertex of $\mathcal{Q}^x$, using (1.1),

$$\mathrm{ec}(e, \mathcal{T}) \leq d(2^{d-k} - m) - [(d-k)2^{d-k} - 2(d-k)m] = k2^{d-k} + dm - 2km.$$

Suppose $f = wu, e = uv$ are edges on $\mathcal{F}$ such that w is closer to $r = 0_d$ on the frame. Then it is easy to check $|\partial A_f| - |\partial A_e| \leq d - 2$, where A_e and A_f are the components which do not contain r. In other words, when we retrace an edge $e \in T$ on the frame back to the root, we increase the congestion by $d-2$ for each level. As $\mathcal{F}$ originates from T whose depth is k, the total increase is at most $k(d-2)$. Hence, the congestion of an edge on the frame is at most $k2^{d-k} + k(d-2) + (d-k)(d-2k)$.

As $\mathcal{S}^x$ is obtained from a standard spanning tree by pruning a number of leaves all of which are neighbours of the connecting vertex of $\mathcal{Q}^x$, for any $e \in E(\mathcal{S}^x)$, $\mathcal{S}^x - e$ contains a component which is either a subcube or a subcube less a vertex. Moreover, the dimension of such a cube is at most $d-k-1$. Thus $\mathrm{ec}(e, \mathcal{T}) = (d-a)2^a$ or $d(2^b - 1) - 2(b2^{b-1} - b) = (d-b)2^b + 2b - d$, where $a, b \leq d-k-1$. By Lemma 1.9, $\mathrm{ec}(e, \mathcal{T}) \leq (k+1)2^{d-k-1} + (d-2k-2)$.

Observe that the maximum congestion occurs at an edge incident with the root. Thus $\mathrm{ec}(Q_d : \mathcal{T}) \leq k2^{d-k} + k(d-2) + (d-k)(d-2k)$. Hence, we have proved the following.

Theorem 1.11. *For any integer $d \geq 8$,*

$$s(Q_d) \leq \frac{\lfloor \log_2 d \rfloor}{2^{\lfloor \log_2 d \rfloor}} 2^d + (d-2)\lfloor \log_2 d \rfloor + (d - \lfloor \log_2 d \rfloor)(d - 2\lfloor \log_2 d \rfloor).$$

We believe that the upper bound is tighter in the sense that $s(Q_d)$ has jumps. Compare $s(G_{n,p})$ in Section 1.6.

1.2.4 Lower bound for $s(Q_d)$

We apply (1.6) to derive a lower bound of $s(Q_d)$. The EIP of Q_d was first solved by Harper [43]. For $d \geq 7$, we use the following approximate version of due to Chung, Füredi, Graham and Seymour [23].

Theorem 1.12. *For any nonempty $X \subset Q_d$, $|\partial X| \geq |X|(d - \log_2 |X|)$.*

Note that although the theorem is true for all nonempty X, the estimate is much tighter when $|X| \leq 2^{d-1}$.

Theorem 1.13. *For any integer $d > 1$, $s(Q_d) \geq \dfrac{\log_2 d}{d} 2^d$.*

Proof. Let T be a spanning tree of Q_d and u a centroid vertex. Since Q_d is d-regular, by (1.5), $w(T) \geq (2^d - 1)/d$. By Lemma 1.6, there exists an edge $e \in E(T)$ such that both components of $T - e$ have orders at least $(2^d - 1)/d$.

We improve the lower bound of $w(T)$ to $2^d/d$ by showing that $d \nmid (2^d - 1)$. Indeed, we may assume that $d > 1$ is odd. Let p be the smallest prime factor of d. Then the order $k > 1$ of 2 in the unit group $U(\mathbb{Z}_p)$ satisfies $k|(p-1)$ and so $k \nmid d$. Hence, $p \nmid (2^d - 1)$ and so $d \nmid (2^d - 1)$. As the order of a component is an integer, we may assume there exists $e \in E(T)$ such that both $|A_e|, |B_e|$ are at least $2^d/d$. By Theorem 1.12,

$$\mathrm{ec}(e, T) = |\partial A_e| \geq \frac{2^d}{d}\left(d - \log_2 \frac{2^d}{d}\right) = \frac{\log_2 d}{d} 2^d.$$

Indeed, let $g(x) = x(d - \log_2 x)$. Since edge-boundary is symmetric under taking complement, we may assume $x \leq 2^{d-1}$. Since $g'(x) = d - 1/\ln 2 - \log_2 x$, g is increasing for $2^d/d \leq x \leq 2^{d-1/\ln 2}$ and decreasing for $2^{d-1/\ln 2} \leq x \leq 2^{d-1}$. Therefore, the minimum of g over $[2^{d-1/\ln 2}, 2^{d-1}]$ is $g(2^{d-1}) = g(2^{d-2}) = 2^{d-1} \geq g(2^d/d)$, as $2^{d-2} \leq$

$2^{d-1/\ln 2}$. Thus, $g(x) \geq g(2^d/d)$ for $x \in [2^d/d, 2^{d-1}]$. Hence, $\mathrm{ec}(e,T) \geq (\log_2 d/d)2^d$. Since T is arbitrary, we have $s(Q_d) \geq (\log_2 d/d)2^d$. □

In the following, we determine the exact values of $s(Q_d)$ for $1 \leq d \leq 6$ using a sharp bound of the edge-boundary of a subset of Q_d, due to Harper [43], Lindsey [68], Bernstein [6] and Hart [45]. The binary order of Q_d is the natural order when each vertex is viewed as a binary number. The results on $s(Q_d)$ for $d \leq 6$ have been obtained independently by Hunter [48].

Theorem 1.14. *Let $A \subset Q_d$, and let I be the set of the first $|A|$ vertices of Q_d in the binary order. Then $|\partial A| \geq |\partial I|$.*

Let $b(m)$ be the maximum number of edges induced by a set of m vertices of Q_d. Suppose e is an edge of a spanning tree T of Q_d. By (1.1), $\mathrm{ec}(e,T) \geq d|A_e| - 2b(|A_e|) = d|B_e| - 2b(|B_e|)$.

Lemma 1.15. *For any $0 \leq m \leq 2^k, k > 0$, $b(2^k + m) = b(2^k) + b(m) + m$; $b(1) = 0$ and $b(2) = 1$.*

Proof. Let $d > k$ and $X \cup Y$ be the first $2^k + m$ vertices of Q_d, where $X = \{(x, 0_{d-k}) : x \in Q_k\}$ and vertices of Y take the form of $(y, 1, 0_{d-k-1})$. By Theorem 1.14, $e(X \cup Y) = b(2^k + m)$. Ignoring the $(k+1)$st entry onward, we view Y as the first m vertices of Q_k, and again by Theorem 1.14, $e(Y) = b(m)$; moreover, each vertex in Y is adjacent to exactly one vertex in X, which itself spans $b(2^k)$ edges. Thus, $b(2^k + m) = b(m) + m + b(2^k)$. □

Theorem 1.16. *For $1 \leq d \leq 6$, $s(Q_d) = 2^{d-1}$; $59 \leq s(Q_7) \leq 60$.*

Proof. We compute the following table using the recursion in Lemma 1.15.

m	1	2	3	4	5	6	7	8	9	10	11	12
$b(m)$	0	1	2	4	5	7	9	12	13	15	17	20
$3m-2b(m)$	**3**	4	5	4								
$4m-2b(m)$	4	**6**	8	8	10	10	10	8				
$5m-2b(m)$	5	8	11	12	**15**	16	17	16	19	20	21	20
$6m-2b(m)$	6	10	14	16	20	22	24	24	28	**30**	32	32
$7m-2b(m)$	7	12	17	20	25	28	31	32	37	40	43	44
m	13	14	15	16	17	18	19	20	21	22	23	24
$b(m)$	22	25	28	32	33	35	37	40	42	45	48	52
$5m-2b(m)$	21	20	19	16								
$6m-2b(m)$	34	34	34	32	36	38	40	40	42	42	42	40
$7m-2b(m)$	47	48	49	48	53	56	**59**	60	63	64	65	64
m	25	26	27	28	29	30	31	32	33	34	35	36
$b(m)$	54	57	60	64	67	71	75	80	81	83	85	88
$6m-2b(m)$	42	42	42	40	40	38	36	32				
$7m-2b(m)$	67	68	69	68	69	68	67	64	69	72	75	76
m	37	38	39	40	41	42	43	44	45	46	47	48
$b(m)$	90	93	96	100	102	105	108	112	115	119	123	128
$7m-2b(m)$	79	80	81	80	83	84	85	84	85	84	83	80
m	49	50	51	52	53	54	55	56	57	58	59	60
$b(m)$	130	133	136	140	143	147	151	156	159	163	167	172
$7m-2b(m)$	83	84	85	84	85	84	83	80	81	80	79	76
m	61	62	63	64								
$b(m)$	176	181	186	192								
$7m-2b(m)$	75	72	79	64								

For any spanning tree T of Q_6, $w(T) \geq 2^6/6 > 10$. By (1.6),

$$s(Q_6) \geq \min_{11 \leq m \leq 32} 6m - 2b(m) = 32.$$

A standard spanning tree of Q_d shows that $s(Q_d) \leq 2^{d-1}$; so $s(Q_6) = 32$.

For $d = 3, 4, 5$, a similar analysis with critical orders of the components being $1, 2, 5$ respectively shows that $s(Q_d) = 2^{d-1}$. Clearly, $s(Q_1)$ is 1 and $s(Q_2)$ is 2.

Since the weight of any spanning tree of Q_7 is at least $128/7 > 18$, by the table and (1.6), $s(Q_7) \geq 59$. We produce a spanning tree of Q_7 with edge congestion 60 in Figure 1.3. Indeed, note that any edge disconnecting a component of order at most 11 has congestion at most 57. Moreover, the most congested edges are incident with the centroid.

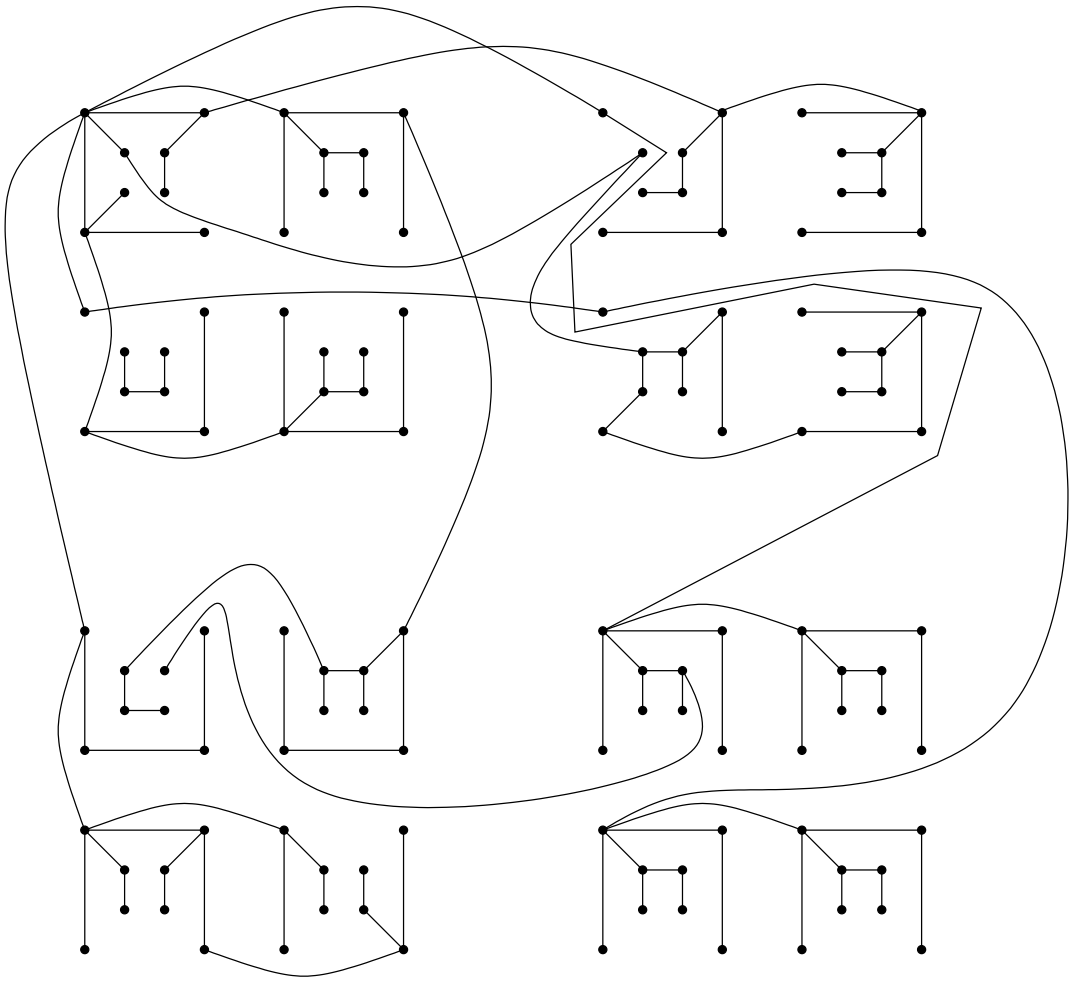


Figure 1.3: A spanning tree of Q_7 with edge congestion 60

□

1.3 Cartesian products of complete graphs

In this section we estimate spanning tree congestion for products of complete graphs. This problem was studied by Kozawa and Otachi [56] who, in particular, computed the exact values of $s(K_m \times K_n)$ independently, with an approach similar to ours in Section 1.3.2. In the next section, we give a tight estimate of $s(K_n^d)$ for n fixed.

This section is joint work with M.I. Ostrovskii [60].

1.3.1 Hamming graphs

For $d, n \in \mathbb{N}$, a *Hamming graph* $H = H(d,n)$ is the product of d copies of K_n. That is, $V(H) = \{(v_i)_1^d, 0 \leq v_i \leq n-1, 1 \leq i \leq d\}$ and two vertices $(v_1, \cdots, v_d)$ and $(u_1, \cdots, u_d)$ are adjacent iff $u_i = v_i$ for all but one $i \in \{1, \cdots, d\}$. Note that Hamming graphs generalize hypercubes as $H(d,2) = Q_d$.

The lexicographic ordering of $V(H)$ orders $(v_i) < (u_i)$ if there exists $1 \leq m \leq d$ such that $v_1 = u_1, \cdots, v_m = u_m$, $v_{m+1} < u_{m+1}$. Then the initial segment of the ordering minimizes the edge-boundary, see for example, Harper [44]. Furthermore, an approximation is given by Squier, Torrence and Vogt [91] which extends Therorem 1.12.

Lemma 1.17. *For $A \subset V(H(d,n))$, $2e(A) \leq (n-1)|A| \log_n |A|$.*

Using this lemma, it was shown in [56] that $s(H(d,n)) \geq (n^d - 1) \log_n d/d$ for $n, d \geq 3$. For an upper bound, the same approach in Section 1.2 (see also [57]) can be used. Details are in [60].

Theorem 1.18. *If $d, n \geq 3$, then*

$$s(H(d,n)) \leq (n-1) \min \left\{ \left(\frac{\lfloor \log_n d(n-1) \rfloor}{n^{\lfloor \log_n d(n-1) \rfloor}} n^d + d^2(n-1) \right), n^{d-1} \right\}.$$

Thus for fixed n and large d

$$s(H(d,n)) = O\left(\frac{\log_n d}{d} \cdot n^d \right),$$

and so $s(H(d,n)) = \Theta\left(\frac{\log_n d}{d} \cdot n^d \right)$.

1.3.2 Two-dimensional

In this section, we compute $s(K_{n_1} \times K_{n_2})$ exactly by isoperimetry. It has also been obtained independently by Kozawa and Otachi [56].

Let $G = K_{n_1} \times K_{n_2} = \{(x, y) : 0 \le x \le n_1 - 1, 0 \le y \le n_2 - 1\}$. We construct a tree such that $(0, 0)$ is adjacent to all $(x, 0)$ and $(0, y)$ for $x, y \neq 0$; and (x, y) is adjacent to $(x, 0)$ if $M(x, y) = C$; to $(0, y)$ if $M(x, y) = R$ according to an n_1 by n_2 array M whose entry is either R or C.

Theorem 1.19. *For $2 \le n_1 \le n_2$, let $G = K_{n_1} \times K_{n_2}$ and $c_G = \lceil \frac{n_1 n_2 - 1}{n_1 + n_2 - 2} \rceil$; then*

$$s(G) = \begin{cases} c_G(n_1 + n_2 - 1 - c_G) & \text{if } n_1 - 1 > c_G, \\ n_1 n_2 - n_2 & \text{otherwise.} \end{cases}$$

Proof. We first apply edge-isoperimetry (see Harper [44]) to derive a lower bound. For $1 \le a \le n_1 n_2/2$, let $a = qn_2 + r$ where $0 \le r < n_2$ and $f(a) = \min_{A \subset V(G), |A| = a} |\partial A|$. Then

$$\begin{aligned} f(a) &= (n_1 + n_2 - 2)(qn_2 + r) - qn_2(n_2 - 1) - q(q - 1)n_2 - r(r - 1) - 2qr \\ &= qn_2(n_1 - q) + r(n_1 + n_2 - 1 - 2q - r). \end{aligned}$$

As $n_2 - 1 \ge r$ and $q \le n_1/2$, $n_1 - 1 - 2q \ge -1$, it follows that for $q \ge 1$, $f(qn_2 + r) \ge f(qn_2) \ge f(n_2)$. Hence, by (1.6), $s(G) \ge \min\{f(n_2), f(c_G)\}$.

We distinguish two cases according to whether the minimum is $f(n_2)$ or $f(c_G)$.

Suppose $n_1 - 1 > c_G$. Then we have $n_1 \ge 4$ so that $(n_1 + n_2 - 1)/2 \ge c_G$. Furthermore, $(n_1 + n_2 - 1)/2 - c_G \ge n_2 - (n_1 + n_2 - 1)/2$. It follows that $f(c_G) = c_G(n_1 + n_2 - 1 - c_G) \le n_1 n_2 - n_2 = f(n_2)$.

For a matching upper bound, since vertices in the same row/column form a clique, it suffices to construct a spanning tree from an array with at most $c_G - 1$ entries being R in each row and at most $c_G - 1$ entries being C in each column.

In other words, we want a bipartite graph with classes $V_1 \cup V_2$ such that $|V_1| = n_1 - 1$ and $|V_2| = n_2 - 1$, $d(v_2) \le c_G - 1$ for all $v_2 \in V_2$ and $d(v_1) \ge n_2 - c_G$ for all $v_1 \in V_1$. This can be constructed by assigning $c_G - 1$ neighbours to each vertex in V_2 one at a

$(0,0)$	$(1,0)$	$(2,0)$	$(3,0)$	$(4,0)$	$(5,0)$	$(6,0)$	$(7,0)$
$(0,1)$	R	R	R	R	R	C	C
$(0,2)$	R	R	R	C	C	R	R
$(0,3)$	R	C	C	R	R	R	R
$(0,4)$	C	R	R	R	R	R	C
$(0,5)$	R	R	R	R	C	C	R
$(0,6)$	R	R	C	C	R	R	R
$(0,7)$	C	C	R	R	R	R	R
$(0,8)$	R	R	R	R	R	C	C
$(0,9)$	R	R	R	C	C	R	R
$(0,10)$	R	C	C	R	R	R	R
$(0,11)$	C	R	R	R	R	R	C

Figure 1.4: A minimum congestion spanning tree of $G = K_8 \times K_{12}$ where $c_G = 6$

time, so that at any time the neighbourhoods of vertices in V_1 are as even as possible, e.g. cyclically.

On the other hand, if $n_1 - 1 \leq c_G$, then $f(n_2) \leq f(c_G)$. In this case, we construct a spanning tree whose array is all C's. It is easily checked that the congestion of this tree is $n_1 n_2 - n_2 \geq n_1 + n_2 - 2$ as $n_1 \geq 2$. □

1.4 Grid and torus

A *grid* P_n^d is a product of paths and a *torus* $\mathbb{Z}_n^d$ is a product of cycles. Hruska [47] computed $s(P_n^2)$ using duality of planar graphs. Castejón and Ostrovskii [20] estimated the growth of $s(P_n^3)$ and $s(\mathbb{Z}_n^3)$. Here we estimate $s(P_n^d), s(\mathbb{Z}_n^d)$ for fixed d. Indeed we approximate a large grid by the solid cube. A lower bound is derived in Section 1.4.1 using the edge-isoperimetric inequalities for the grid, proved by Bollobás and Leader [11].

An upper bound of $s(P_n^d)$ can be derived from partitions of $[0,1]^d$. A *brick* is a product of intervals each containing more than one points and a *rectilinear body* is

a union of bricks. In the following, we consider partition of $[0,1]^d$ into $2d$ rectilinear bodies each of which has a measure of at most $1/2$. For a rectilinear body B, let $|\partial_g B|$ denote the $(d-1)$-dimensional measure of the topological boundary of B in $[0,1]^d$; and $|\partial_t B|$ for that of the boundary of B in $(\mathbb{R}/\mathbb{Z})^d$. Let $m(B)$ be the d-dimensional measure of B. Let

$$\begin{aligned} g(d) &= \min\{\max\{|\partial_g B|, B \in \mathcal{B}\}, \mathcal{B} \text{ partitions } [0,1]^d\}, \\ t(d) &= \min\{\max\{|\partial_t B|, B \in \mathcal{B}\}, \mathcal{B} \text{ partitions } (\mathbb{R}/\mathbb{Z})^d\}. \end{aligned}$$

We shall concentrate on $g(d)$, any bound on $g(d)$ can be transferred to $t(d)$ upon multiplying by 2. Moreover, estimate on $g(d)$ can be transferred easily to information about $s(P_n^d)$. Indeed, given any partition attaining $g(d)$, for each corresponding block of P_n^d, we construct a spanning tree. These trees are then joined to a centroid (deleting edges if necessary to avoid cycles) to form a spanning tree. Moreover, this can be arranged such that congestion within each block is small compared to that between different blocks. Hence, the main term in the congestion corresponds to the maximum surface area of a piece. Using the isoperimetric inequalities from [11] and applying (1.6), we have the following.

Theorem 1.20. *For $d \geq 1$, $F_d(\frac{1}{2d})n^{d-1} \leq s(P_n^d) \leq g(d)n^{d-1} + Cn^{d-2}$, for some constant $C > 0$ and any large integer n, where $F_d(v) = \min\{rv^{1-1/r}, r(1-v)^{1-1/r}, r = 1, 2, \cdots, d\}$ for $0 \leq v \leq 1$.*

See also [20] for $d = 3$. We give partitions of the cube in Section 1.4.2 for an upper bound.

This section is joint work with M.I. Ostrovskii.

1.4.1 Lower bound

To derive a lower bound, we apply edge-isoperimetric inequalities of the grid due to Bollobás and Leader [11, Theorems 2, 7].

Theorem 1.21. *Let $F_d(v) = \min\{rv^{1-1/r}, r(1-v)^{1-1/r}, r = 1, 2, \cdots, d\}$. Then for a rectilinear body $A \subset [0,1]^d$, $|\partial_g A| \geq F_d(m(A))$ and $|\partial_t A| \geq 2F_d(m(A))$.*

In any partition of $[0,1]^d$ into $2d$ blocks, there is always a block with volume at least $\frac{1}{2d}$. Furthermore, as $F_d(v)$ is increasing from $[0,1/2]$, we have

$$\begin{aligned} g(d) &\geq F_d\left(\frac{1}{2d}\right), \\ t(d) &\geq 2F_d\left(\frac{1}{2d}\right). \end{aligned}$$

Since $r(2d)^{1/r-1} \leq r(1-1/(2d))^{1-1/r}$ for all $1 \leq r \leq d$, we need only find the minimum of $h(r) = r(2d)^{1/r-1} = \frac{r}{2d}(2d)^{1/r}$ over $[0,d]$. Differentiating,

$$2dh'(r) = (2d)^{1/r}\left(1 - \frac{\ln 2d}{r}\right).$$

Hence, h attains minimum $\frac{e\ln 2d}{2d}$ at $r = \ln 2d$. Let $r_0 = \lfloor \ln 2d \rfloor$, then

$$F_d\left(\frac{1}{2d}\right) \geq \min\left\{\left(\frac{r_0}{2d}\right)(2d)^{1/r_0}, \left(\frac{r_0+1}{2d}\right)(2d)^{1/(r_0+1)}\right\}.$$

A cleaner but weaker bound is given as follows.

Lemma 1.22. *For $d > 2$, $g(d) \geq \dfrac{e\ln 2d}{2d}$ and $t(d) \geq \dfrac{e\ln 2d}{d}$.*

1.4.2 Upper bound

Next, we construct partitions of $[0,1]^d$. Let $\mathcal{F}$ be a partition of Q_d into subcubes. For ease of presentation, we assign a vector to each subcube $F \in \mathcal{F}$ by

$$F(i) = \begin{cases} \text{the } i\text{th entry of } F \text{ if it is constant over } F; \\ *, \text{ otherwise.} \end{cases}$$

We suppress brackets and commas. For example, when $d = 4$, the vector $000*$ denotes the edge $\{(0,0,0,0),(0,0,0,1)\}$.

Given any real number $0 \leq a_i \leq 1$ for $i = 1, 2, \cdots, d$, we construct a partition of $[0,1]^d$ derived from $\mathcal{F}$ weighted by the *partition vector*, $(a_1, a_2, \cdots, a_d)$ as follows. For $F \in \mathcal{F}$, let

$$J_i(F) = \begin{cases} [0,1] & \text{if } F(i) = *; \\ [0,a_i] & \text{if } F(i) = 0; \\ [a_i,1] & \text{if } F(i) = 1. \end{cases}$$

Define the corresponding brick as $\prod_{i=1}^{d} J_i(F)$.

For any integer $d > 2$, let $k = \lfloor \log_2 2d \rfloor$ and $m = 2^{k+1} - 2d$. Let $v_1 < v_2 < \cdots < v_{2^k}$ be the binary order of the vertices of Q_k. Let $\mathcal{F}(m, d, k)$ be the partition $\{(v_i, *_{d-k}) : 1 \le i \le m\} \cup \{(v_i, 0, *_{d-k-1}), (v_i, 1, *_{d-k-1}) : i > m\}$. Note that there are exactly $2d$ elements in $\mathcal{F}(m, d, k)$.

Lemma 1.23. *For $d > 2$, let $k = \lfloor \log_2 2d \rfloor$ and $m = 2^{k+1} - 2d$. Using $\mathcal{F}(m, d, k)$, we derive the following upper bounds of $g(d)$:*

1. $\dfrac{4k^2 - 2}{(3k-2)2^k}$, *if $m \le 2^{k-1}$; attained by the partition vector $\left(\frac{2k-1}{6k-1}, \left(\frac{1}{2}\right)_k, 0_{d-k-1}\right)$;*
2. $\dfrac{k}{2^{k-1}}$, *if $m > 2^k - 2^{k-2}$; attained by the partition vector $\left(\left(\frac{1}{2}\right)_{k+1}, 0_{d-k-1}\right)$;*
3. $\dfrac{4k^3 - 12k^2 + 4k + 10 + (2k^2 - 4k)\sqrt{4k^2 - 8k + 5}}{2^{k-1}(3k-5)^2}$, *if $2^{k-1} < m \le 2^k - 2^{k-2}$; attained by the partition vector $\left(\left(\frac{4k-5-\sqrt{4k^2-8k+5}}{6k-10}\right)_2, \left(\frac{1}{2}\right)_{k-1}, 0_{d-k-1}\right)$.*

Proof. Consider $\mathcal{F}(m, d, k)$.

1. If $m \le 2^{k-1}$, the partition consists of (where * indicates directions that are not cut)

$$\begin{array}{cccccc} 0 & x & \cdots & x & * & *_{d-k-1} \\ 0 & x & \cdots & x & x & *_{d-k-1} \\ 1 & x & \cdots & x & x & *_{d-k-1}. \end{array}$$

   We optimize over p with the partition vector $(p, (1/2)_k, 0_{d-k-1})$:

$$\begin{aligned} & \min_p \max\{2^{1-k} + p(k-1)2^{2-k}, 2^{-k} + k(1-p)2^{1-k}, 2^{-k} + kp2^{1-k}\} \\ = \; & \frac{1}{2^k} \min_p \max\{2 + 4(k-1)p, 2k - 2kp + 1, 1 + 2kp\} \\ = \; & \frac{8k^2 - 4}{(6k-4)2^k}, \end{aligned}$$

   where the minimum is attained at $p = (2k-1)/(6k-4)$.

2. If $2^k > m > 2^k - 2^{k-2}$, the partition consists of

$$\begin{array}{cccccc} x & x & x & \cdots & x & * & *_{d-k-1} \\ 1 & 1 & x & \cdots & x & x & *_{d-k-1}; \end{array}$$

if $m = 2^k$ we have the first type but not the second. By symmetry, we may assume that the first two coordinates of the optimal partition vector are equal. Thus, we optimize over p with the partition vector $(p, p, (1/2)_{k-1}, 0_{d-k-1})$:

$$\begin{aligned} & \frac{1}{2^k} \min_p \max\{2kp - 2p, (2k-4)p(1-p) + 1, \\ & 2(k-2)(1-p)^2 + 2(1-p), (k-1)(1-p)^2 + (1-p)\} \\ = \quad & \frac{k}{2^{k-1}}, \end{aligned}$$

where the minimum is attained at $p = 1/2$.

3. If $2^{k-1} < m \leq 2^k - 2^{k-2}$, the partition consists of

$$\begin{array}{cccccc} 0 & 0 & x & \cdots & x & * & *_{d-k-1} \\ 0 & 1 & x & \cdots & x & * & *_{d-k-1} \\ 1 & 0 & x & \cdots & x & * & *_{d-k-1} \\ 1 & 0 & x & \cdots & x & x & *_{d-k-1} \\ 1 & 1 & x & \cdots & x & x & *_{d-k-1}. \end{array}$$

We optimize over the partition vector $(p, p, (1/2)_{k-1}, 0_{d-k-1})$, that is to solve

$$\begin{aligned} & \frac{1}{2^{k-1}} \min_p \max\{4p + 4(k-2)p, 2 + 4(k-2)p(1-p), \\ & 2(k-1)(1-p)^2 + 2(1-p)\}. \end{aligned}$$

The minimum is attained when the last two terms are equal, i.e. $(3k-5)p^2 + (5-4k)p + k - 1 = 0$, whose solution p is as stated.

□

By putting $p = 1/2$, we obtain a general but weaker bound.

Lemma 1.24. *For $d > 2$, $g(d) \leq \dfrac{2\lfloor \log_2 2d \rfloor}{2^{\lfloor \log_2 2d \rfloor}}$ and $t(d) \leq \dfrac{4\lfloor \log_2 2d \rfloor}{2^{\lfloor \log_2 2d \rfloor}}$.*

So far our partitions use bricks only. In particular, for $d = 3$, Lemma 1.23 gives the partition $\{000, 001, 010, 011, 10*, 11*\}$ with partition vector $(3/8, 1/2, 1/2)$, such that $g(3) \leq 7/8$ (which is the same as in [20]). In fact, it is possible to slightly

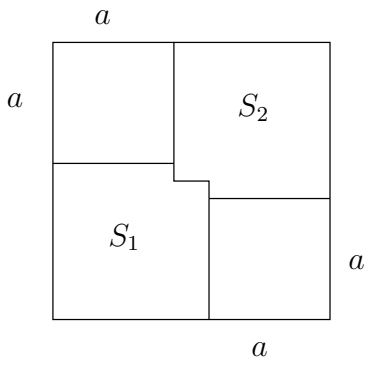


Figure 1.5: Cuts of $[0,1] \times [0,1]$ optimized at $a = \frac{-3+\sqrt{15}}{2}$

improve to $\sqrt{15} - 3 \approx 0.873$ by using rectilinear bodies, see [60]. For $a = \frac{-3+\sqrt{15}}{2}$ and S_1, S_2 as indicated in Figure 1.5, partition $[0,1]^3$ into

$$A_1 = [0,a] \times [1-a,1] \times [0,1], \quad A_2 = [1-a,1] \times [0,a] \times [0,1],$$

$$A_3 = S_1 \times \left[0, \frac{1}{2}\right], A_4 = S_1 \times \left[\frac{1}{2}, 1\right], A_5 = S_2 \times \left[0, \frac{1}{2}\right], A_6 = S_2 \times \left[\frac{1}{2}, 1\right].$$

Then the maximum surface area is $\sqrt{15} - 3$.

1.5 Complete multipartite graphs

Spanning tree congestion of complete multipartite graphs has been computed in [25] and [55]. We assume that the sizes $n_1, \cdots, n_k$ $(k \geq 2)$ of the classes satisfy $n_1 \geq n_2 \geq \cdots \geq n_k$ and let $N = \sum_{i=1}^{k} n_i$. Throughout this section, let $K = K(n_1, n_2, \cdots, n_k)$ denote the complete k-partite graph with n_i vertices in the ith class.

Theorem 1.25 ([25, 55])**.**

$$s(K(n_1, \cdots, n_k)) = \begin{cases} 2N - n_1 - n_2 - 2 & \text{if } n_k > 1 \\ N - n_{k-1} & \text{if } n_k = 1. \end{cases}$$

We are going to show that Theorem 1.25 can be easily derived from the isoperimetric result of Piliposyan and Muradyan [82]. For a slightly simplified version of the proof in English, see [59].

This section is joint work with M.I. Ostrovskii.

We first introduce some terminology.

Definition 1.26. An ordering of the vertex set of a complete multipartite graph K is called *PM-admissible* if it can be obtained according to the following inductive description (suggested in [82]).

1. If each class of the partition contains exactly one element, then each ordering is PM-admissible.

2. If there are classes containing two or more elements and $V_1, \cdots, V_q$ are the largest classes (all other classes are strictly smaller) then an ordering of $V(K)$ is PM-admissible if it can be described in the following way: There are elements $s_i, t_i \in V_i$, $s_i \neq t_i$, $(i = 1, \cdots, q)$ such that the ordering starts with $s_1, \cdots, s_q$ (ordered arbitrarily), ends with $t_1, \cdots, t_q$ (ordered arbitrarily), and between these two sets contains a PM-admissible ordering of the multipartite graph obtained from K if we delete the vertices $s_1, \cdots, s_q, t_1, \cdots, t_q$ from it.

Definition 1.27. We say that a subset $S \subset V(K)$ is a *PM-set* if it is an initial set of some PM-admissible ordering.

We denote the cardinality of the edge-boundary of a PM-set of cardinality α in K by $\mathrm{PM}_K(\alpha)$.

Theorem 1.28 ([82])**.** *For any complete multipartite graph K, let $S \subset V(K)$. Then $|\partial S| \geq PM_K(|S|)$.*

Corollary 1.29. *The function $\theta_K(a)$ defined by (1.3) is increasing for $1 \leq a \leq \lfloor N/2 \rfloor$.*

Proof. By Theorem 1.28, it suffices to show that $\mathrm{PM}(a) \leq \mathrm{PM}(a+1)$ when $a < N/2$. Suppose the $(a+1)$st vertex in a PM-admissible ordering is in V_j. Denote the initial set of cardinality k of this ordering by I_k. It is easy to check that $|\partial I_{a+1}| = |\partial I_a| + \sum_{i=1, i\neq j}^{k} (|V_i| - 2|V_i \cap I_a|)$. Observe that $|V_i| - 2|V_i \cap I_a|$ is non-negative unless I_a contains halves of all even classes, and the $(a+1)$st vertex of the ordering is the

$(n_j+1)/2$-th (we call it *middle*) vertex from an odd class V_j. But in such a case, I_a contains middle vertices of less than half of odd classes, and it is easy to check that $|\partial I_a| \leq |\partial I_{a+1}|$ is still satisfied. □

Corollary 1.30. *If $N \geq n_1+2$, then the function $f_K(a)$ defined by (1.4) is increasing for $1 \leq a \leq \lfloor N/2 \rfloor$.*

Proof. Note that if $n_1 \leq n_2+2$, then θ_K coincides with f_K (the second vertex of a PM-admissible ordering can be picked from V_2) so that we are done by Corollary 1.29. Otherwise, the two functions differ on the range $a \leq \lceil \frac{n_1-n_2}{2} \rceil$. Indeed, Theorem 1.28 implies that for $a > \lceil \frac{n_1-n_2}{2} \rceil$, the minimum edge-boundary is attained by subsets intersecting at least two classes; and in complete multipartite graphs such subsets induce connected subgraphs. For $a \leq \lceil \frac{n_1-n_2}{2} \rceil$, we claim that $f_K(a) = (a-1)(N-n_1)+N-n_2-2(a-1)$. For any $A \subset V(K)$ such that $|A \cap V_i| = a_i$ and $|A| = a$, we have

$$\begin{aligned} |\partial A| &= \sum_{i=1}^{k} (a_i(N-n_i) - a_i(a-a_i)) \\ &= aN - a^2 - \sum a_i(n_i - a_i). \end{aligned}$$

If we transfer one vertex of A from V_i to V_1 to get A', then

$$|\partial A'| = |\partial A| - (n_1 - n_i - 2(a_1 - a_i + 1)).$$

As $\frac{n_1-n_i}{2} \geq a \geq a_1 - a_i + 1$, we have $|\partial A'| \leq |\partial A|$. Hence, putting more vertices in V_1, we get a smaller edge-boundary. To get a connected subgraph, we can put only $a-1$ vertices in V_1 and, as is easy to see, the edge-boundary is minimized if the remaining vertex is in V_2. Hence, $f_K(a) = a(N-n_1-2)+n_1-n_2+2$ for $a \leq \lceil \frac{n_1-n_2}{2} \rceil$ which is increasing in a because $N-n_1-2 \geq 0$. □

Remark 1.31. *The case $N = n_1+1$ is exceptional. In this case the function f_K is decreasing. But then $K = K_{1,n_1}$ so that $s(K) = 1$.*

Proof of Theorem 1.25. First we consider the case $n_k > 1$. Pick a vertex v in V_2 and consider the following spanning tree T. The edge set $E(T)$ contains all edges joining v

with vertices in $V(K)\backslash V_2$. To get a spanning tree we add to $E(T)$ a matching between $V_2 - v$ and some subset of V_1. A straightforward inspection of all edges shows that $\mathrm{ec}(K:T) = N - n_1 + N - n_2 - 2$.

On the other hand, since $n_k > 1$, $\Delta_K < N - 1$ and $N \geq n_1 + 2$. Therefore, $c_K = \lceil (N-1)/\Delta_K \rceil \geq 2$. By Corollary 1.30, $\min_{c_K \leq k \leq N/2} f_K(k) \geq f_K(2)$. Applying (1.6), we conclude $s(K) \geq f_K(2) = f_K(2) = N - n_1 + N - n_2 - 2$ which is as required for the case $n_k > 1$.

Suppose $n_k = 1$. Then the edges incident with the only vertex in V_k form a spanning tree with edge congestion $N - n_{k-1}$. If there are some more classes V_i of cardinality 1, one can construct similar trees for them. It is clear that all other spanning trees in $K(n_1, \cdots, n_k)$ have weight at least 2.

If $k = 2$, then $n_{k-1} = N - 1$ and the edge congestion $N - n_{k-1} = 1$ is clearly optimal. If $k \geq 3$, then the condition of Corollary 1.30 is satisfied and, as is easy to verify, $f_K(2) \geq N - 1$. Thus for a spanning tree T with weight at least 2, $\mathrm{ec}(K : T) \geq N - 1 \geq N - n_{k-1}$. □

1.6 Random graphs

1.6.1 Spanning tree congestion

In this section, we study the spanning tree congestion of a random graph in $\mathcal{G}(n,p)$ and a random bipartite graph in $\mathcal{G}(n,n;p)$. Recall that $\mathcal{G}(n,p)$ is the probability space of all graphs on $[n]$ such that each edge occurs with probability p independently; and $\mathcal{G}(n,n;p)$ is the space of all bipartite graphs with vertex classes $([n],1),([n],2)$ such that each edge $(i,1),(j,2)$ occurs with probability p independently. We assume in this section that $0 < p < 1$. Let $q = 1 - p > 0$.

A sequence of events $\mathcal{E}_n \subset \mathcal{G}(n,p)$ or $\mathcal{G}(n,n;p)$ is said to hold *asymptotically almost surely (a.a.s.)* if $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$. We first study the isoperimetric properties of a random graph and then show that a spanning tree of weight $\lceil 1/p \rceil$ a.a.s. exists. The main tool is the following Chernoff bound, see for example [3]. Let

$X \sim \mathrm{Bin}(n,p)$. Then for any $\lambda > 0$,

$$\mathbb{P}(X - pn < -\lambda) \;<\; \exp\left(\frac{-\lambda^2}{2pn}\right).$$

For $v \in \mathcal{G}(n,p)$, $\deg(v) \sim \mathrm{Bin}(n-1,p)$. Much sharper results about the degree sequence are known (see for example [10]) but we shall need only the following estimate which holds a.a.s.,

$$pn - O(\sqrt{pn\log n}) = \delta \le \deg(v) \le \Delta = pn + O(\sqrt{pn\log n}).$$

Lemma 1.32. *Let $1 > p = p(n) \ge \frac{\log^2 n}{n}$ and $k = \lceil 1/p \rceil$. Then a random graph $G \in \mathcal{G}(n,p)$ a.a.s. has the property that for any $A \subset [n]$ with $k \le |A| \le n/2$, $|\partial A| \ge knp - 3k\sqrt{pn\log n}$.*

Proof. Note that for $A \subset [n]$ such that $a = |A|$, $|\partial A| \sim \mathrm{Bin}(a(n-a),p)$. Let $\lambda = a\sqrt{6pn\log n}$. Then

$$\mathbb{P}(|\partial A| < a(n-a)p - \lambda) \;<\; \exp\left(\frac{-\lambda^2}{2pa(n-a)}\right) = O(n^{-3a}).$$

For $k \le a \le n/2$, let $\mathcal{E}_a$ be the event that there exists $A \subset [n]$ such that $|A| = a$ and $|\partial A| < a(n-a)p - \lambda$. Since there are $\binom{n}{a}$ sets of a vertices, it follows that

$$\mathbb{P}\left(\bigcup_{a=k}^{\lfloor n/2\rfloor} \mathcal{E}_a\right) \;\le\; \sum_{a=k}^{\lfloor n/2\rfloor} n^a n^{-3a} = o(n^{1-2k}) = O(n^{-1}).$$

Let $g(a,n) = a(n-a)p - a\sqrt{6pn\log n}$. Since a.a.s. $|\partial A| > g(a,n)$ for all A with $k \le |A| \le n/2$, it suffices to show that $g(a,n) \ge knp - 3k\sqrt{pn\log n}$ for $k \le a \le n/2$ where n is large.

Consider $n/2 \ge a \ge 3k$. Then for large n, since $p \gg \log n/n$,

$$g(a,n) \ge \frac{apn}{2} - a\sqrt{6pn\log n} \ge \frac{apn}{3} \ge knp.$$

Note that $g(a,n)$ is increasing on $0 < a < \frac{pn - \sqrt{6pn\log n}}{2p}$. For $k \le a \le 3k$, we have

$$\begin{aligned} g(a,n) &\ge g(k,n) \\ &= k(n-k)p - 2k\sqrt{pn\log n} \\ &> knp - 3k\sqrt{pn\log n}. \end{aligned}$$

This completes the proof. □

We aim to construct a spanning tree such that the branches at a centroid are as balanced as possible. It is constructed with matchings of the neighbourhood of a vertex to the rest of the graph. For an integer $r \geq 1$ and disjoint subsets A, B of $V(G)$ such that $|B| = r|A|$, a subgraph H of G is called an *r-matching*, if $\deg_H(a) = r, \deg_H(b) = 1$ for all $a \in A, b \in B$. Then edges incident with a vertex v and an r-matching from $\Gamma(v)$ to the rest of the graph is a spanning tree. For the existence of matchings, we recall Hall's theorem.

Theorem 1.33 (Hall's theorem)**.** *For a bipartite graph G with vertex classes V_1, V_2, there exists a matching from V_1 to V_2 iff for all $S \subset V_1$, $|\Gamma(S)| \geq |S|$.*

Lemma 1.34. *Let $p = p(n) \geq \frac{\log n}{n^{1/3}}$ and $a = \lfloor pn \rfloor$. Let A, B be disjoint subsets of $[n]$ such that $a = |A| = |B|$. Then in $G \in \mathcal{G}(n,p)$ with probability $1 - o(pn^{-2})$ there exists a perfect matching from A to B.*

Proof. Let $\mathcal{E}_s$ be the event that there exists $S \subset A$ such that $s = |S|$ and $|\Gamma(S) \cap B| < |S|$. By Hall's condition, $\mathbb{P}(\text{there exists no matching from } A \text{ to } B) = \mathbb{P}(\cup_{1 \leq s \leq a} \mathcal{E}_s)$. We estimate this probability by dividing the range of s into three parts. Let $\lambda = 2\sqrt{pa \log \frac{an^2}{p}}$ and

1. $p_1 = \mathbb{P}\left(\cup_{s < pa - \lambda} \mathcal{E}_s\right)$,
2. $p_2 = \mathbb{P}\left(\cup_{s > (1-p)a + \lambda} \mathcal{E}_s\right)$,
3. $p_3 = \mathbb{P}\left(\cup_{pa/2 < s < (1-p/2)a} \mathcal{E}_s\right)$.

For a bound of p_1, we show that one vertex in A a.a.s. has enough neighbours to cover for $pa - \lambda$ vertices. Indeed,

$$\begin{aligned} p_1 &\leq \mathbb{P}(\exists v \in A : |\Gamma(v) \cap B| < pa - \lambda) \\ &\leq a\mathbb{P}(\mathrm{Bin}(a,p) - pa < \lambda) \\ &\leq \exp\left(\log a - \frac{\lambda^2}{2pa}\right) \\ &\leq \exp\left(-\log a - 2\log\frac{n^2}{p}\right) = o(pn^{-2}). \end{aligned}$$

As $p_2 \le \mathbb{P}(\exists v \in B : |\Gamma(v) \cap A| < pa - \lambda)$, $p_2 = o(pn^{-2})$.

For p_3, we show that a set of at least $pa/2$ vertices a.a.s. covers all but at most $pa/2$ vertices of B. Indeed,

$$\begin{aligned} p_3 &\le \mathbb{P}(\exists S \subset A, T \subset B : s = |S| = |T| = pa/2, e(S,T) = 0) \\ &\le \binom{a}{s}^2 (1-p)^{s^2} \\ &\le \left(\frac{ea}{s}\right)^{2s} \exp(-ps^2) \\ &\le \left(\frac{2e}{p}\right)^{p^2 n} \exp\left(\frac{-p^5 n^2}{4}\right) \\ &\le \exp\left(\left(\log 2e - \log p - \frac{p^3 n}{4}\right) p^2 n\right) \end{aligned}$$

which is $o(pn^{-2})$ since $\log \frac{1}{p} \ll p^3 n$ and $p^2 n \to \infty$. Therefore, with probability $1 - (p_1 + p_2 + p_3) = 1 - o(pn^{-2})$ there exists a matching from A to B. □

Lemma 1.35. *For $p(n) \ge \frac{\log n}{n^{1/3}}$, a random graph $G \in \mathcal{G}(n,p)$ a.a.s. has a spanning tree of weight $w = \lceil 1/p \rceil$.*

Proof. Let $a = \lfloor pn \rfloor$ and $w = \lceil 1/p \rceil$. Introduce a set C of vertices such that $n - 1 - \lfloor pn \rfloor + |C|$ is divisible by $\lfloor pn \rfloor$ with $w - 1 = \lceil 1/p \rceil - 1$ being the quotient. Moreover, each edge between C and $[n]$ occurs with probability p independently.

For each vertex $v \in [n]$, expose the neighbourhood $\Gamma(v) \cap [n]$. Let A_v consist of the first $a_v = \min\{a, |\Gamma(v) \cap [n]|\}$ vertices in $\Gamma(v) \cap [n]$ and $a - a_v$ vertices in $[n] \setminus \Gamma(v)$. Thus, if $\deg(v) \ge pn$, $A_v \subset \Gamma(v)$.

Partition $[n] \cup C \setminus (A_v \cup \{v\})$ into $w-1$ classes $B_1, B_2, \cdots, B_{w-1}$ such that $|B_i| = a$ and $C \subset B_{w-1}$. If there is an $(w-1)$-matching from A_v to $\cup B_i$, then by deleting C and putting in $E(v, A_v)$, we obtain a spanning tree T of G with $w(T) = w$. Thus,

$$\begin{aligned} & \mathbb{P}(G \text{ does not have a spanning tree of weight } w \text{ and } \Delta(G) \ge pn) \\ \le\ & \mathbb{P}(\exists v \text{ such that there is no } (w-1)\text{-matching from } A_v \text{ to } \bigcup B_i) \\ \le\ & n\mathbb{P}(\text{for a fixed } v, \text{ there is no } (w-1)\text{-matching from } A_v \text{ to } \bigcup B_i) \\ \le\ & n\mathbb{P}(\text{for a fixed } v, \text{ there exists } i \text{ such that there is no matching from } A_v \text{ to } B_i) \\ \le\ & n(w-1)o(pn^{-2}) = o(n^{-1}), \end{aligned}$$

by Lemma 1.34. Finally,

$$\begin{aligned} & \mathbb{P}(G \text{ does not have a spanning tree of weight } w) \\ \leq\ & \mathbb{P}(G \text{ does not have a spanning tree of weight } w \text{ and } \Delta(G) \geq pn) + \mathbb{P}(\Delta(G) < pn) \\ =\ & o(1). \end{aligned}$$

□

Theorem 1.36. *Let $0 < p < 1$ be fixed. Then for any $G \in \mathcal{G}(n,p)$, we have a.a.s.*

$$\lceil 1/p \rceil pn - O(\sqrt{n \log n}) \leq s(G) \leq \lceil 1/p \rceil pn + O(\sqrt{n \log n}).$$

Proof. By Lemma 1.35, there a.a.s. exists a spanning tree with weight $\lceil 1/p \rceil$ of depth 2. It is easy to check that the edge congestion is at most $\lceil 1/p \rceil \Delta(G)$.

Let T be a spanning tree of G. Since $\Delta(G) = pn + O(\sqrt{n \log n})$ a.a.s., by (1.5),

$$\begin{aligned} w(T) &\geq \frac{n-1}{\Delta(G)} \\ &\geq \left(1 - \frac{1}{n}\right) \frac{1}{p + O\left(\sqrt{\frac{\log n}{n}}\right)} = \frac{1}{p}\left(1 - O\left(\sqrt{\frac{\log n}{n}}\right)\right). \end{aligned}$$

Hence for large n, $w(T) \geq \lceil 1/p \rceil$.

By Lemma 1.6, there exists an edge e such that $T - e$ contains a component A of order $w(T) \geq \lceil 1/p \rceil$. Then by Lemma 1.32,

$$s(G) \geq \mathrm{ec}_G(e,T) \geq |\partial A| \geq \lceil 1/p \rceil np - 3\lceil 1/p \rceil \sqrt{pn \log n}).$$

□

The zigzag nature of $s(G)/n$ as in Figure 1.6.1 reflects the attempt to balance the partition of G according to the typical degree. When $1/p$ is an integer, we have a balanced partition and the spanning tree congestion is smallest.

Next we apply an analogous procedure to study $s(\mathcal{G}(n,n;p))$. Let V_1, V_2 be the two vertex classes of $G \in \mathcal{G}(n,n;p)$. We first note the analogue of Lemma 1.32.

Lemma 1.37. *Let $p \geq \frac{\log^2 n}{n}$ and $k = \lceil 2/p \rceil$. Then a random bipartite graph $G \in \mathcal{G}(n,n;p)$ a.a.s. has the property that for any set $C \subset V_1 \cup V_2$ with $k \leq |C| \leq n$, we have $|\partial C| \geq knp - 4k\sqrt{pn \log n}$.*

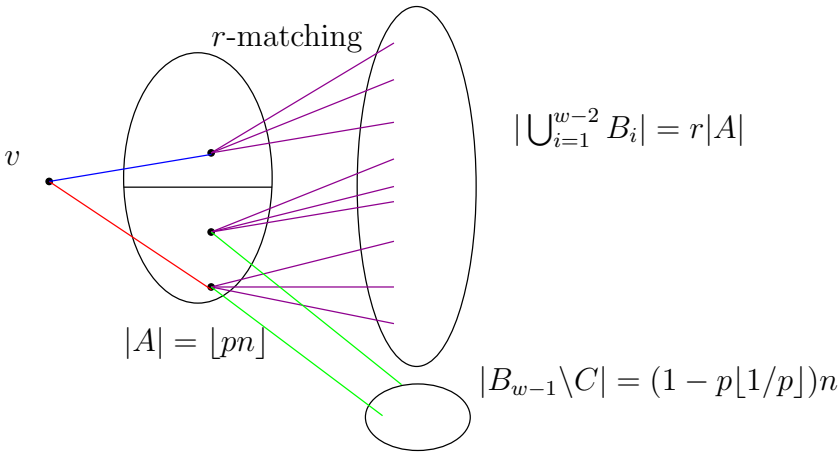


Figure 1.6: A spanning tree of $G \in \mathcal{G}(n,p)$ where $r = \lceil 1/p \rceil - 2$

Proof. For a fixed set C of $k \le c \le n$ vertices, let $c_i = |C \cap V_i|$ for $i = 1, 2$. Then $|\partial C| \sim \mathrm{Bin}(c_1(n-c_2)+c_2(n-c_1), p)$. Let $M = cn - 2c_1c_2 < cn$ and $\lambda = 3c\sqrt{pn \log n}$. By the Chernoff bound,

$$\mathbb{P}(|\partial C| < Mp - \lambda) \;<\; \exp\left(\frac{-\lambda^2}{2pM}\right) < \exp\left(\frac{-9c^2 pn \log n}{2cnp}\right) = o((2n)^{-c-2}).$$

Let $\mathcal{E}_c$ such that there exists $C \subset V_1 \cup V_2$ such that $|C| = c$ and $|\partial C| < Mp - \lambda$. As there are at most $(2n)^c$ sets of c vertices,

$$\mathbb{P}\left(\bigcup_{c=k}^{n} \mathbb{E}_c\right) \le \sum_{c=k}^{n} (2n)^c (2n)^{-c-2} < n^{-1} = o(1).$$

Let $g(c,n) = (cn - c^2/2)p - 3c\sqrt{pn \log n}$. Since $M > cn - c^2/2$, to prove the lemma, it suffices to show that $g(c,n) \ge knp - 4k\sqrt{pn \log n}$. Indeed, for $3k < c \le n$,

$$g(c,n) \ge \frac{cnp}{2} - 3c\sqrt{pn \log n} \ge \frac{cnp}{3} > knp.$$

For $k \le c \le 3k$,

$$g(c,n) \;\ge\; g(k,n) = knp - \frac{k^2 p}{2} - 3k\sqrt{pn \log n} \ge knp - 4k\sqrt{pn \log n}.$$

□

Theorem 1.38. *Let $0 < p < 1$ be fixed. Then for any $G \in \mathcal{G}(n,n;p)$, we have a.a.s.*

$$\lceil 2/p \rceil pn - O(\sqrt{n \log n/p}) \le s(G) \le \lceil 2/p \rceil pn + O(\sqrt{n \log n/p}).$$

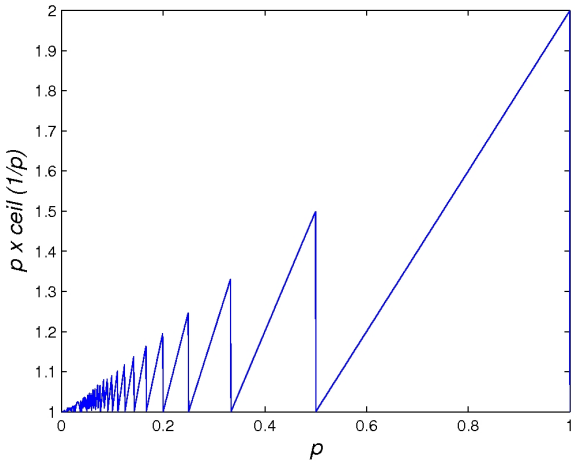


Figure 1.7: A plot of $p\lceil 1/p \rceil$

Proof. Let $r = \lfloor 1/p - 1 \rfloor$. Let $G \in \mathcal{G}(n, n; p)$. Since $p \gg \frac{\log n}{n}$, G a.a.s. has a matching M (see [10, Section 7.3]). We condition on the event that G has a matching. Assume that $x_i \in V_1$ is matched to y_i in V_2.

Fix a vertex x_i. Let $b = \lfloor pn \rfloor$. Let $B = B_{x_i}$ be the first $b_{x_i} = \min\{b, |\Gamma(x_i) \setminus \{y_i\}|\}$ vertices and $b - b_{x_i}$ vertices in $V_2 \setminus \Gamma(x_i)$. Let $A = \Gamma_M(B)$, i.e. the set of neighbours of B under the matching M. Choose $D \subset V_2 \setminus (B \cup \{y_i\})$ such that $|D| = r|A|$, and let $C = \Gamma_M(D)$. Let $U = V_1 \setminus (A \cup C \cup \{x_i\})$, so that $|U| = n(1 - p\lfloor 1/p \rfloor)$ and $W = \Gamma_M(U)$. By Lemma 1.34, with probability $1 - o(n^{-1})$ there exists an $(\lfloor 1/p \rfloor - 1)$-matching R from A to D. To cover the rest of the graph, we consider two cases depending on the fractional part $\{1/p\}$ of $1/p$. Note $\lceil 2/p \rceil = 2\lceil 1/p \rceil - 1$ if $0 < \{1/p\} < 0.5$.

1. $\{1/p\} \geq 1/2$: We match U to a subset S of B. Then $E(x_i, B) \cup R \cup M \cup N$ is a spanning tree T with weight $\lceil 2/p \rceil$.

2. $\{1/p\} < 1/2$: Partition A, B into $A' \cup A''$ and $B' \cup B''$ respectively such that each part has cardinality at least $|U| = |W|$. This is possible since $|U| = p\{1/p\}n < pn/2$. Then we match U to B' and W to A''. Hence $E(x_i, B) \cup R \cup M \setminus E(U, W)$ is a spanning tree with weight $\lceil 2/p \rceil$.

Each of the above matching fails to exist with probability $o(n^{-1})$. Summing over

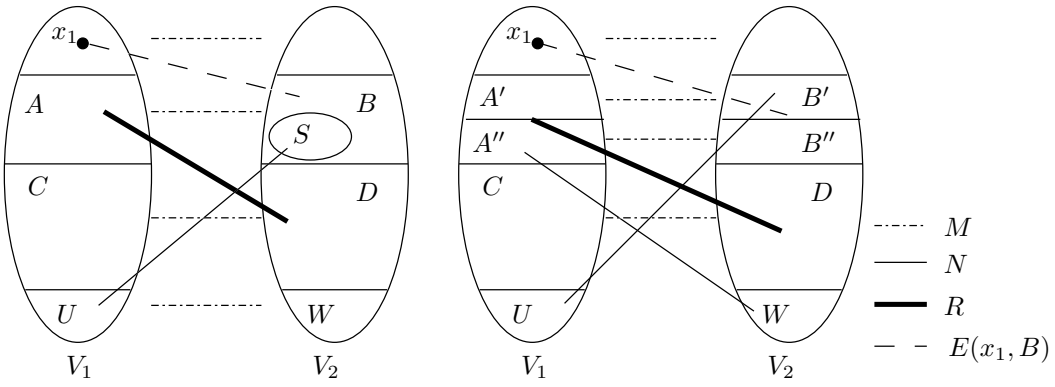


Figure 1.8: A spanning tree of $G \in \mathcal{G}(n, n; p)$

all vertices, we conclude that there a.a.s. exists a spanning tree with weight $\lceil 2/p \rceil$. Note that a.a.s. $\mathrm{ec}(G : T) \leq \lceil 2/p \rceil (pn + O(\sqrt{pn \log n}))$.

For any spanning tree T, we have a.a.s. $w(T) \geq \frac{2n-1}{\Delta}$, where Δ is a.a.s. $pn + O(\sqrt{n \log n})$. By (1.6) and Lemma 1.37, $s(G) \geq w(T)np - O(\sqrt{n \log n})$. □

1.6.2 Tree congestion

We compute the tree congestion of a random graph in $\mathcal{G}(n, p)$ using a characterization of $t(G)$ by Ostrovskii [78]. For $u, v \in V(G)$, let $m(u, v)$ be the maximum number of independent u-v paths. Then $t(G) = \max_{u,v \in V(G)} m(u, v)$.

Theorem 1.39. *Let $p = p(n) \geq \log n / n^{1/3}$. Then for a random graph $G \in \mathcal{G}(n, p)$, $t(G)$ is a.a.s. the second largest degree, which a.a.s. is $pn + c\sqrt{pn \log n}$ for some constant c.*

Proof. By the characterization of $t(G)$, it suffices to study the maximum number of edge-disjoint paths between any two vertices. Let u, w be two fixed vertices of $G \in \mathcal{G}(n, p)$. Let $U = \Gamma(u)$ and $W = \Gamma(w)$. By the Chernoff bound, the common neighbourhood $Y = |U \cap W| \sim \mathrm{Bin}(n - 2, p^2)$ is a.a.s. at most $p^2 n + 3\sqrt{p^2 n \log n}$.

Let $U' = U \setminus W$ and $W' = W \setminus U$. Suppose $|U'| \leq |W'|$. Since a.a.s. we have $|U|, |W| \geq pn - O(\sqrt{pn \log n})$, we may delete some vertices in W' to form W'' such that $|U'| = |W''| \geq (p - p^2)n - O(\sqrt{pn \log n})$. Now, we expose the edges between

U' and W''. The proof of Lemma 1.34 can be adapted to show that with probability $1 - o(n^{-2})$, there exists a matching from U' to W''. Thus, with this probability, there are $\min\{\deg(u), \deg(w)\}$ independent u-w paths. As there are at most $n^2/2$ pairs of vertices, we conclude that a.a.s. the number of independent paths between two vertices is equal to the smaller degree of the two. Hence, $t(G)$ is a.a.s. the second largest degree of G. □

1.7 Maximum spanning tree congestion

So far all graphs that we have considered have spanning tree congestion at most linear in the order n. Are there graphs with larger spanning tree congestion? More precisely, let

$$\mu(n) = \max\{s(G) : G \text{ is a graph of order } n\}.$$

Ostrovskii [78] showed that $\mu(n) \geq cn^{3/2}$ for some constant c with the following graph. For $n = 3k - 2\sqrt{k}$, let $V(G) = V_1 \cup V_2 \cup V_3$ and each V_i induces a clique with $|V_1 \cap V_2| = |V_2 \cap V_3| = \sqrt{k}$, $V_1 \cap V_3 = \emptyset$. On the other hand, it was conjectured there that $\mu(n) = o(n^2)$. It was later shown by Löwenstein, Rautenbach and Regen [70] that $\mu(n) = \Theta(n^{3/2})$ using a theorem obtained independently by Lovász and Győri that a k-connected graph can be partitioned into k connected subgraphs of prescribed orders each containing one of k prescribed vertices.

Here we use Szemerédi's Regularity Lemma to prove that every graph is close to a graph with linear order spanning tree congestion. In particular, this furnishes an alternate proof to the bound $\mu(n) = o(n^2)$. First we construct a spanning tree of a graph based on a regular partition. Then we delete $o(n^2)$ edges such that the edge congestion of this tree is $O(n)$. We now define an operation called *bounded blowup* which is used subsequently in the construction of spanning trees.

Recall the definition of an ϵ-regular pair. For a pair of disjoint set of nonempty vertices (A, B) of a graph G, let $d = d(A, B) = \dfrac{e(A, B)}{|A||B|}$ be its density. Given $0 < \epsilon < 1$, (A, B) is an *ϵ-regular pair* if for any $X \subset A, Y \subset B$, with $|X| \geq \epsilon|A|, |Y| \geq \epsilon|B|$ we have $d - \epsilon < d(X, Y) < d + \epsilon$. The following is a simple consequence of the definition.

Proposition 1.40. *Given a graph G, let (A, B) be an ϵ-regular pair such that $l = |A| = |B|$ and $d = d(A, B) > \epsilon$. Then all but at most ϵl vertices of A have at least $(d - \epsilon)l$ neighbours in B.*

Proof. Let $X = \{v \in A : |\Gamma(v) \cap B| < (d - \epsilon)l\}$, then $d(X, B) < d - \epsilon$. By regularity, $|X| < \epsilon l$. □

A partition of $V(G) = C_0 \cup C_1 \cup \cdots \cup C_k$ is *ϵ-regular* if $|C_0| \leq \epsilon|G|$, $|C_i| = |C_j|$ and all but at most ϵk^2 pairs (C_i, C_j), for $1 \leq i < j \leq k$, are ϵ-regular.

Theorem 1.41 (Szemerédi's Regularity Lemma)**.** *For $0 < \epsilon < 1/2, m \in \mathbb{N}$, there exists constant $M = M(\epsilon, m)$ such that any graph G with $|G| > m$ has an ϵ-regular partition into $(C_i)_{i=0}^k$ where $m \leq k \leq M$.*

The idea of bounded blowup between an ϵ-regular pair (A, B) is that we can require some subset S of A to see only a bounded number of neighbours such that almost all vertices in B are covered. By abuse of notation, we regard a set of edges also as the set of endvertices.

Lemma 1.42. *Let (A, B) be an ϵ-regular pair in a graph G such that $l = |A| = |B|$. Let S be a set of at least $2\epsilon l$ vertices of A. Then there exists a subgraph H on $S \cup B'$ where $B' \subset B$ and $|B'| \geq l - |S|$ such that $d_H(s) \leq \eta = \lfloor \frac{l-|S|}{|S|-\epsilon l} \rfloor + 1, \forall s \in S$ and $d_H(b') = 1, \forall b' \in B'$.*

Proof. We construct H by repeatedly finding matchings from S to B of maximum cardinality, which cannot be too small by regularity. Indeed, let $P_0 = \{s_1b_1, \cdots, s_kb_k\}$ be a maximum matching from S to $B_0 = B$. By maximality, there are no edges between $S_1 = S \setminus P_0$ and $B_1 = B_0 \setminus P_0$. Moreover, it follows that $|S_1| < \epsilon l$. Indeed, if $\epsilon l \leq |S_1| \leq |B_1|$, then by regularity, we have $0 < d - \epsilon < d(S_1, B_1)$ which contradicts $e(S_1, B_1) = 0$. Consequently, $|B_1| < l - (|S| - \epsilon l)$.

For $i \geq 1$, if $|B_i| \geq |S|$, we construct a maximum matching P_i from S to B_i, and let $B_{i+1} = B_i \setminus P_i$; otherwise, we finish our construction of H which consists of the edges $\bigcup_{j=0}^{i} P_j$. Note that for each i, $|B_{i+1}| < l - (i + 1)(|S| - \epsilon l)$. Thus for some $t \leq \eta = \lceil \frac{l-|S|}{|S|-\epsilon l} \rceil$, $|B_t| < |S|$ and we terminate our construction.

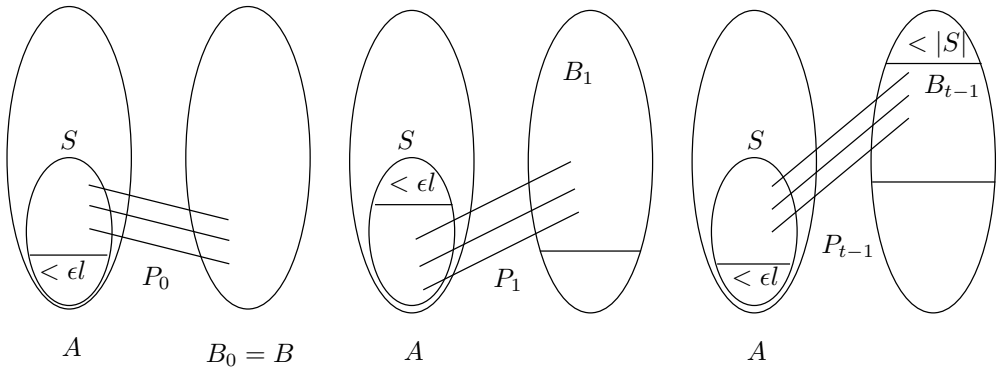


Figure 1.9: Construction of $P(A, B, S)$ in Lemma 1.42

Let $B' = \bigcup_{i=0}^{t}(P_i \cap B)$. Since we construct a matching at each step, $d_H(b') = 1$ for all $b' \in B'$ and $d_H(s) \leq \eta$ for all $s \in S$. □

We shall denote the edges in H constructed in Lemma 1.42 between an ϵ-regular pair (A, B) with a set $S \subset A$ by $P(A, B, S)$. Moreover, we choose $|S| = 2\epsilon l$, η depends only on ϵ. We now prove our main theorem.

Theorem 1.43. *Let $0 < \gamma < 1/2$. There exist $N = N(\gamma), c = c(\gamma)$ such that for any graph G such that $n = |G| > N$, there exists a subgraph G' such that $e(G) - e(G') < \gamma n^2$ and $s(G') \leq cn$.*

Proof. Let $0 < \epsilon < \gamma/5$ and $m \in \mathbb{N}$ such that $m > 1/\epsilon$. Let G be a graph of order $n \geq m$. By Theorem 1.41, there exists $M = M(\epsilon, m)$ such that G admits an ϵ-regular partition $(C_i)_{i=0}^{k}$ where $m \leq k \leq M$. We first clean up G. Indeed, let $d = 3\epsilon$. Delete all edges

1. incident with C_0;
2. between non ϵ-uniform pairs (C_i, C_j);
3. between ϵ-uniform pairs of insufficient densities, i.e. $d(C_i, C_j) < d$;
4. within the vertex classes.

Denote the resulting graph by H. Let R be the reduced graph of H, i.e. $V(R) = \{c_1, \cdots, c_k\}$ and $c_i c_j \in E(R)$ iff (C_i, C_j) is ϵ-regular and $d(C_i, C_j) \geq d$.

Let $\mathcal{F}$ be a spanning forest of R with the least number of components. Root each component at a leaf. Then enumerate the component in a breadth-first-search order. Relabelling if necessary, we consider a component F with vertices $c_1, \cdots, c_{|F|}$.

By Proposition 1.40, there exists $u \in C_1$ such that $|\Gamma(u) \cap C_2| > 2\epsilon l$. Let $S_2 \subset \Gamma(u) \cap C_2$ such that $|S_2| = \lceil 2\epsilon l \rceil$. For the next edge $c_i c_j$ in F where $i < j$, construct $P(C_i, C_j, S_i) \subset E(C_i, C_j)$. This covers all but at most $2\epsilon l$ vertices in C_j. Let $S_j \subset C_j$ be a set of $\lceil 2\epsilon l \rceil$ vertices covered by $P(C_i, C_j, S_i)$.

To cover C_1, construct $P(C_2, C_1, S_2)$. For the rest of C_2 other than S_2, let $S_1 \subset C_1$ be a set of $\lceil 2\epsilon l \rceil$ vertices covered by $P(C_2, C_1, S_2) \cap C_1$. Construct $P(C_1, C_2, S_1)$ and delete, if necessary, some edges to avoid forming cycles with the previously chosen edges between C_1 and C_2. Denote the resulting edge set by $P'(C_1, C_2, S_1)$.

Thus, we have constructed a tree $T(F)$ spanning all but at most $2\epsilon l|F|$ vertices of $H[\bigcup_{c_i \in F} C_i]$.

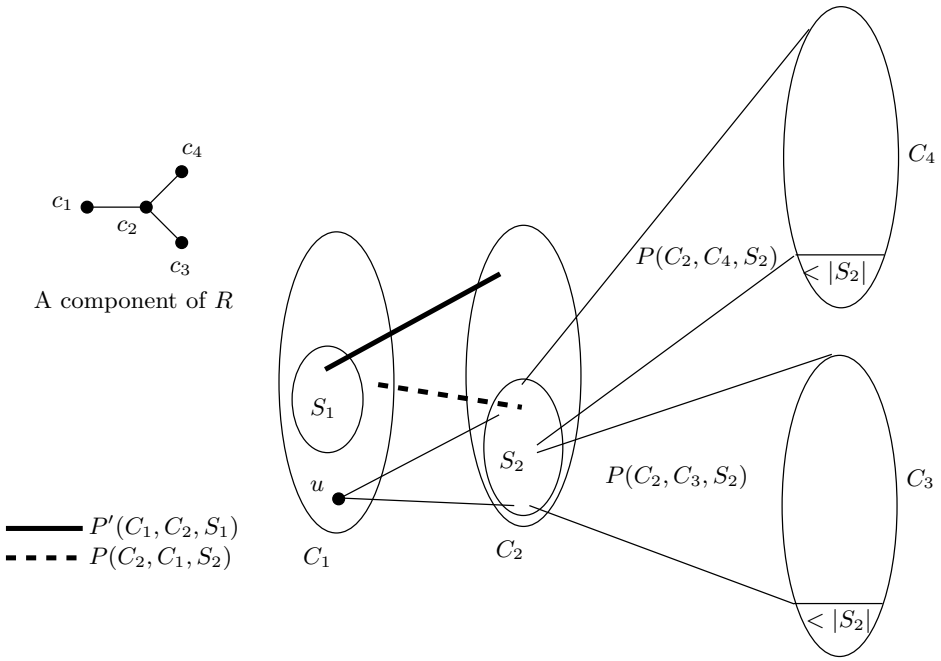


Figure 1.10: Example of the construction of a T_p

Next, we delete all edges incident with vertices not covered by $\bigcup_{F \in \mathcal{F}} T(F)$ to form H'. Choose $J \subset E(G) \setminus \bigcup_{F \in \mathcal{F}} T(F)$ such that $\mathcal{T} = \bigcup_{F \in \mathcal{F}} T(F) \cup J$ is a spanning tree of G' where $E(G') = E(H') \cup J$.

We now estimate the edge congestion $\mathrm{ec}(G' : T)$. As we have deleted all edges between irregular pairs and regular pairs of insufficient densities, there are no edges between any pair $T(F)$ and $T(F')$ in $G' \setminus J$. It follows that the congestion of any edge in J is 1. On the other hand, if $|F| > 1$, $T(F)$ is a tree such that there is a vertex of degree $2\epsilon l$ and the rest has degree at most $\Delta(F)\eta + 1 \leq k\eta + 1 \leq M\eta + 1$. As the depth of F is at most $k \leq M$, any edge of $T(F)$ disconnects a component of order at most $(M\eta + 1)^\eta$ which is a constant. Hence, $\mathrm{ec}(G' : T) = O(n)$.

Finally, we estimate the number of deleted edges $e(G) - e(G')$. Since we have deleted edges between irregular pairs, regular pairs of insufficient densities ($< 3\epsilon$), edges incident with the exceptional set, edges within vertex classes as well as those with the unmatched portion which amounts to at most $2\epsilon l$ in each vertex class, the number of deleted edges is at most

$$\epsilon k^2 l^2 + \frac{1}{2}k^2(3\epsilon)l^2 + \epsilon n^2 + k\frac{l^2}{2} + \frac{k^2}{2}(2\epsilon l)l.$$

Recalling that $kl \leq n$ and $m \leq k$, we conclude that we have deleted at most $(\epsilon + 3\epsilon/2 + \epsilon + \frac{1}{2m} + \epsilon)n^2 < 5\epsilon n^2 < \gamma n^2$ edges.

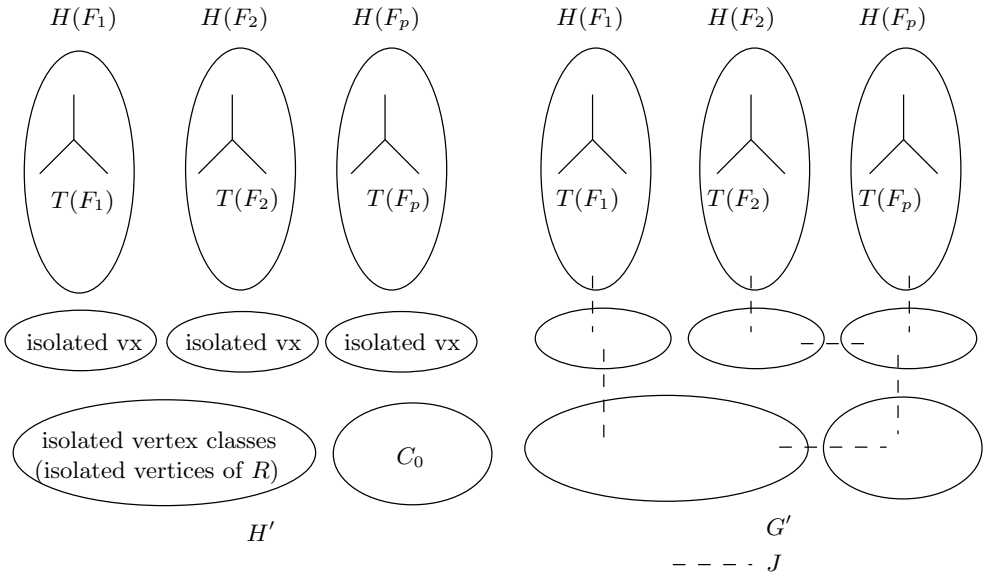


Figure 1.11: Construction of $\mathcal{T}$. Each $T(F_i)$ corresponds to a component of a spanning forest of R

□

Remark 1.44. *The theorem is sharp in the sense that in general there is no subgraph G' of spanning tree congestion $o(n)$ with $e(G) - e(G') = o(n^2)$. Indeed, suppose we delete all but ϵn^2 edges of K_n. Then the resulting graph G' (if connected) satisfies $s(G') = \Omega(n)$. Indeed, as every edge of K_n lies in the edge-boundary of a component of $T - e$, we have $s(G') \geq e(G')/(n-1) = \Omega(n)$.*

Corollary 1.45. $\mu(n) = o(n^2)$.

Proof. Assume on the contrary that there exist a constant $c > 0$ and a sequence of graphs G_i with $n_i = |G_i| \to \infty$ such that $s(G_i) \geq cn_i^2$.

By the theorem, there exists a constant $C > 0$ such that for each G_i with $|n_i| \geq N(c/2)$, we can delete at most $cn_i^2/2$ edges to form G_i' which satisfies $s(G_i') \leq Cn_i$. Putting back these deleted edges increases the edge congestion by at most $cn_i^2/2$. Thus, $s(G_i) \leq Cn_i + cn_i^2/2 < cn_i^2$ which is a contradiction. □

1.8 Maximum spanning tree congestion of planar graphs

The aim of this section is to find sharp estimates of

$$\mu_p(n) = \max\{s(G) : \ G \text{ is a planar graph of order } n\}.$$

Planar graphs G of order n satisfying $s(G) = \mu_p(n)$ can be called the *most congested planar graphs*.

Euler's formula implies that a planar graph G with $n \geq 3$ vertices has at most $3n - 6$ edges. As $n - 1$ of them are in a spanning tree so that they are detours for themselves, it follows that $s(G) \leq 3n - 6 - (n-1) + 1$. Thus $\mu_p(n) \leq 2n - 4$.

This section is joint work with M.I. Ostrovskii.

1.8.1 Estimates from above for planar graphs

We need some standard terminology and notation, see for example [28]. Recall that the *radius* of a connected graph G is defined by

$$\operatorname{rad}(G) = \min_{x \in V(G)} \max_{y \in V(G)} d_G(x, y). \tag{1.7}$$

A vertex x for which the minimum in (1.7) is attained is called *central*.

Let T be a spanning tree of a connected plane graph G and T^* be the dual tree of T in G^*. Recall (1.2) that $\operatorname{cs}(G^* : T^*) = \operatorname{ec}(G : T)$. To find an upper bound of $\mu_p(n)$, it suffices to look at $\operatorname{cs}(G^* : T^*)$.

Lemma 1.46. *Let G be a connected plane graph. Then*

$$\min\{\operatorname{cs}(G^* : T^*) : T \text{ a spanning tree of } G\} \leq 2 \operatorname{rad}(G^*) + 1.$$

Proof. It suffices to observe that any breadth-first-search (BFS) tree T^* in G^* rooted at a central vertex C satisfies $\operatorname{cs}(G^* : T^*) \leq 2 \operatorname{rad}(G^*) + 1$. Indeed, we need only the defining property of a BFS tree in G^* rooted at C: it is a spanning tree in G^* for which the distance between any vertex and C is the same as in G^*, and therefore is at most $\operatorname{rad}(G^*)$. The plus one accounts for the extra edge to close the cycle. □

By adding edges to G such that all faces are triangles, we obtain a *triangulation* G_t of G. It is easy to construct examples showing that G_t is not uniquely determined by G in general. Recall that H is a *minor* of G if H can be obtained from G by a sequence of edge deletions and/or contractions.

Lemma 1.47. $\operatorname{rad}(G_t^*) \geq \operatorname{rad}(G^*)$.

Proof. To see this, it suffices to observe that G^* is a contraction minor of G_t^*, obtained via contracting edges among sets of triangular faces of G_t originated from the same face of G. Since passing to a contraction minor we can only decrease the radius, we get the desired inequality. □

The following result is well known, see [28, Chapter 4] for details.

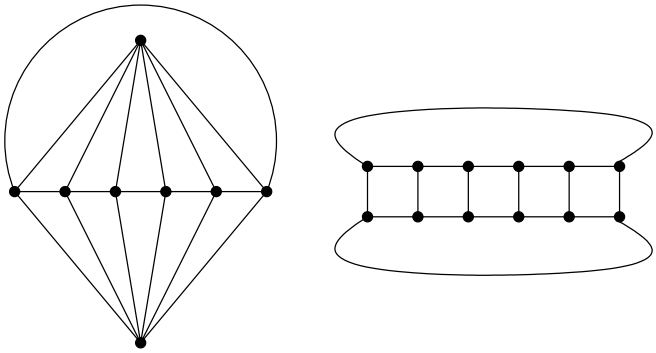


Figure 1.12: Plane triangulation B_n with large spanning tree congestion and its dual on the right

Lemma 1.48. *A triangulation of a planar graph with at least* 4 *vertices is* 3*-connected.*

Finally we need the following tight estimate for a radius of a 3-connected graph obtained by Iida [49]. (See [29], [41], [42], [50], [51] for preceding and related estimates.)

Theorem 1.49 (Iida [49])**.** *Let G be a* 3*-connected graph with n vertices and radius r. Then $4r - 4 \leq n$.*

Corollary 1.50. $\mu_p(n) \leq \begin{cases} n+1 & \text{if } n \text{ is even.} \\ n & \text{if } n \text{ is odd.} \end{cases}$

Proof. By Lemma 1.48, G_t is 3-connected so that G_t^* is also 3-connected with $|G_t^*| = 2n - 4$. By Theorem 1.49, $\mathrm{rad}(G_t^*) \leq \frac{2n}{4} = \frac{n}{2}$. Hence, by Lemma 1.47, $\mathrm{rad}(G^*) \leq \frac{n}{2}$. Combining these inequalities with Lemma 1.46 we get

$$s(G) \leq \begin{cases} 2\left(\frac{n}{2}\right) + 1 = n + 1 & \text{if } n \text{ is even.} \\ 2\left(\frac{n-1}{2}\right) + 1 = n & \text{if } n \text{ is odd.} \end{cases}$$

□

1.8.2 Planar graphs with large spanning tree congestion

For $n \geq 5$, let $B_n = C_{n-2} + 2K_1$ be the graph of a bipyramid on an $(n-2)$-gon base.

Proposition 1.51. $s(B_n) = n$ *for even* $n \geq 6$.

Proof. Observe that the dual of B_n is $B_n^* = C_{n-2} \times K_2$.

By (1.2), it suffices to show that for any spanning tree T^* of B_n^*, there is an edge whose addition creates a cycle of length at least n. This proves $s(B_n) \geq n$. The inequality $s(B_n) \leq n$ is easy to check.

Call an edge joining the two copies of C_{n-2} *vertical* and the rest of $E(B_n^*)$ *horizontal*. Moreover, we say that the two copies of C_{n-2} are of different *levels*. Note that there is at least one vertical edge in T^*. Suppose T^* contains at least two vertical edges. Consider the horizontal edges between two consecutive vertical edges. Then at most one horizontal edge is missing in T^* on each level and at least one level misses a horizontal edge. In fact, there is one interval in which both levels miss a horizontal edge. If we add a missing horizontal edge, then either it completes a cycle of length $n-2$, in which case all horizontal edges on the level are in T^*, or it completes a cycle using at least two vertical edges and hence of length at least n. In the former case, addition of the other horizontal edge completes a cycle that uses at least 2 vertical edges and so has length at least n.

The remaining case is when T^* has exactly one vertical edge e. In this case all but two horizontal edges are in T^* and the missing ones are opposite to each other. Since n is even (and this is the only place an even n is required), the cycle obtained by adding to $E(T^*)$ the vertical edge which is the most distant from e produces a cycle of length at least n. □

Corollary 1.52. $\mu_p(n) \geq \begin{cases} n & \text{if } n \text{ is even.} \\ n-1 & \text{if } n \text{ is odd.} \end{cases}$

It might be possible to use Lemma 1.46 to fill the gap of 1 between Corollaries 1.50 and 1.52. Indeed, Iida [49] gave detailed information about the extremal graph which might allow the construction of a tree with cycle-stretch at most $2\,\mathrm{rad}(G^*)$.

Chapter 2

Enumeration of independent sets

An *independent set* (or a *stable* set) is a set of vertices of a graph G which does not induce any edges. Let $\mathcal{I}(G)$ denote the set of independent sets and $\mathcal{I}_k(G)$ be those of cardinality k. The *independent set polynomial* (or *independence polynomial*) is the generating function of $\mathcal{I}(G)$:

$$I(G,z) = \sum_{I \in \mathcal{I}(G)} z^{|I|} = \sum_{k=0}^{\alpha(G)} i_k(G) z^k,$$

where $i_k(G) = |\mathcal{I}_k(G)|$ and $\alpha(G)$ is the independence number.

The independent set polynomial is of interest to both combinatorialists and physicists. For example, the partition function of the repulsive gas model where two gas particles cannot occur in adjacent sites coincides with the polynomial. See for example [62, 89] and the references therein. From a more graph theoretic point of view, see for example, [38, 40]; for a survey, [64].

Putting $z = 1$, we get the number of independent sets $i(G) := I(G,1)$ first studied by Prodinger and Tichy [85]. Extremal problems of $i(G)$ in different classes of graphs have been studied, for example, trees [85, 67, 98] and regular graphs [53, 97]. For trees in particular, a number of parameters involving independent sets/edges have been studied and in most cases, the stars and the paths are extremal (see [39] for a survey).

As a generating function, $I(G,z)$ is useful to study the average size of independent sets. Indeed, we shall begin by finding the extrema with respect to the average size

in Section 2.2.1. Like other tree parameters, the paths and the stars are extremal.

In mathematical chemistry, $i(G)$ is known as the *Merrifield-Simmons index* [72] or the *σ-index*. There are experimental relations between physical properties (e.g. boiling points) of compounds and the σ-index of their molecular graphs, see Gutman and Polansky [37]. It is also of interest in chemistry ([66]) to know if for an integer k, there exists a compound whose chemical graph has σ-index k. Trees are common chemical graphs, but it remains a conjecture, proposed by Linek [69], that for all sufficiently large integer k, there exists a tree with $i(T) = k$. A weaker version was proposed by Wagner [96] that for all integers $k, m > 0$, there exists a tree such that $i(T) \equiv k \pmod{m}$. We prove this conjecture in Section 2.3.1, see also [58].

More generally, Alon, Haber and Krivelevich [2] considered *F-matchings* in G which are subgraphs of G whose components are copies of F. They asked whether each residue modulo m can be written as the number of F-matchings of a tree. An analogous question about *induced F-matchings*, i.e. an F-matching which does not induce edges outside the copies, was also raised. We answer these two questions positively in Section 2.3.2. In Section 2.3.4, we consider the growth of the number of trees in other residue classes.

2.1 Preliminaries

Let $I(G, z)$ be the independent set polynomial of a graph G. For $v \in G$, let $\Gamma[v] = \Gamma(v) \cup v$ be the closed neighbourhood of v. The fundamental relation of $I(G, z)$ is that for each $v \in G$,

$$I(G, z) = I(G - v, z) + zI(G - \Gamma[v], z), \tag{2.1}$$

which can be proved by considering independent sets containing v or not.

Two operations are useful in our study.

Definition 2.1. *Given rooted graphs $(G_1, r_1), \cdots, (G_m, r_m)$, let $\oplus_{j=1}^{m}(G_j, r_j)$ denote the rooted graph (G, r) obtained by adding a new vertex r joined to each r_j.*

Definition 2.2. *For rooted graphs R and S, the* coalescence $R \circ S$ *is obtained by identifying the roots of R and S.*

The coalescence is also called the *direct sum*.

Applying (2.1), we have

$$I(\oplus_{j=1}^{m}(G_j, r_j), z) = \prod_{j=1}^{m} I(G_j, z) + z \prod_{j=1}^{m} I(G_j - r_j, z). \tag{2.2}$$

As an illustration, consider $i(P_n)$. Putting $z = 1$ in (2.1), $i(P_n) = i(P_{n-1}) + i(P_{n-2})$ for $n \geq 2$ and $i(P_0) = 1, i(P_1) = 2$. Thus, $i(P_n)$ is a Fibonacci number and in fact, this is the name Prodinger and Tichy [85] called $i(G)$. Moreover, they proved that $i(P_n) \leq i(T) \leq i(K_{1,n-1})$ for any tree T of order n. In fact, finer inequalities hold, see also [37].

Proposition 2.3. *For $k \geq 0$ and any tree T of order n, $i_k(P_n) \leq i_k(T) \leq i_k(K_{1,n-1})$.*

Proof. We induct on $n + k$. The statement is clearly true for $n + k \leq 3$. Let uv be an edge of T with u a leaf. Then

$$\begin{aligned} i_k(T) &= i_k(T-u) + i_{k-1}(T-u-v) \\ &\leq i_k(K_{1,n-2}) + i_{k-1}(T-u-v) \\ &\leq \binom{n-2}{k} + \binom{n-2}{k-1} = \binom{n-1}{k} = i_k(K_{1,n}) \end{aligned}$$

since $i_{k-1}(T-u-v) \leq i_{k-1}(K^c_{n-2})$.

On the other hand, since $i_{k-1}(G) \geq i_{k-1}(G')$ if $G \subset G'$. Hence, $i_{k-1}(T-u-v) \geq i_{k-1}(T')$ for some tree T' of order $n-2$. By induction, $i_{k-1}(T') \geq i_{k-1}(P_{n-2})$, hence we have,

$$\begin{aligned} i_k(T) &\geq i_k(P_{n-1}) + i_{k-1}(T-u-v) \\ &\geq i_k(P_{n-1}) + i_{k-1}(P_{n-2}) = i_k(P_n). \end{aligned}$$

□

In fact, $i_k(T) = i_k(K_{1,n-1})$ for some $n-1 \geq k \geq 3$ iff $T = K_{1,n-1}$; $i_k(T) = i_k(P_n)$ for some $\lceil n/2 \rceil \geq k \geq 3$ iff $T = P_n$.

Corollary 2.4. *For $z \geq 0$, $I(P_n, z) \leq I(T, z) \leq I(K_{1,n-1}, z)$.*

We remark that for a connected graph G, passing to a spanning tree does not decrease $i(G)$, therefore, we have the corollary $i_k(G) \leq i_k(K_{1,n-1})$. Another point is that $i_k(T) \leq i_k(K_{1,n-1})$ can also be proved by considering the number of k-cliques in the complement of T. Indeed, if $e(G) = x(x-1)/2$, then it can be proved by induction on k that the number of k-cliques in G is at most $\binom{x}{k}$ where equality holds if and only if x is integral and G is complete.

2.2 Average order of an independent set

Define a probability space on $\mathcal{I}(G)$ by assigning each I a probability proportional to $z^{|I|}$ for some $z > 0$. Hence $\mathbb{P}_z(I \text{ is chosen }) = z^{|I|}/I(G, z)$. This is called the *hard-core model with intensity* z, see for example [53, 89]. Let $X = |I|$. Then

$$\mathbb{E}_z(X) = \frac{\sum_{I \in \mathcal{I}(G)} |I| z^{|I|}}{I(G, z)} = \frac{z \frac{d}{dz} I(G, z)}{I(G, z)}.$$

Write $a(G, z) = \mathbb{E}_z(X)$ and $a(G) = a(G, 1)$ for the expectation on uniform distribution.

By linearity of expectation, $a(G, z)$ is the sum of $a(G_i, z)$ over the components G_i of G. More explicitly, $I(G, z) = \prod_i I(G_i, z)$ so that

$$a(G, z) = \frac{z \sum_i I'(G_i, z) \prod_{j \neq i} I(G_j, z)}{\prod_i I(G_i, z)} = \sum_i a(G_i, z).$$

Differentiating (2.1), we have a recursion for $a(G, z)$, namely, for any $v \in V(G)$,

$$a(G, z) = \frac{I(G - v, z)}{I(G, z)} a(G - v, z) + \frac{z I(G - \Gamma[v], z)}{I(G, z)} (1 + a(G - \Gamma[v], z)). \tag{2.3}$$

Let $i'(G) = \frac{d}{dz} I(G, z)|_{z=1}$. Evaluating (2.3) at $z = 1$, we get

$$a(G) = \frac{i(G - v)}{i(G)} a(G - v) + \frac{i(G - \Gamma[v])}{i(G)} (1 + a(G - \Gamma[v])). \tag{2.4}$$

2.2.1 Trees

We have seen that the paths and the stars are the unique trees attaining the extrema of the number of independent sets, as well as in the weighted version. In fact, it is also true on expectation.

Proposition 2.5. *For any $n \geq 1$, if $T \neq P_n, K_{1,n-1}$, then $a(P_n) < a(T) < a(K_{1,n-1})$.*

To prove Proposition 2.5, we first estimate some parameters of the paths and the stars. Let $h_n = i'(P_n), f_n = i(P_n), p(n) = a(P_n)$ and $\lambda = \frac{-1+\sqrt{5}}{2}, \bar{\lambda} = \frac{-1-\sqrt{5}}{2}$ which are the solutions to $1 - x - x^2 = 0$. We include for completeness a proof for the standard formulae of f_n and h_n.

Lemma 2.6. *The sequences $(f_n)_{n\geq 0}, (h_n)_{n\geq 0}$ satisfy*

1. *$f_{-1} = 1, f_0 = 1, f_1 = 2$ and $h_0 = 0, h_1 = 1$;*
2. *$f_n = f_{n-1} + f_{n-2}$ and $h_n = h_{n-1} + h_{n-2} + f_{n-2}$.*

Explicitly, we have the standard Binet formula

$$f_n = \frac{1}{\sqrt{5}}\frac{1}{\lambda^{n+2}} - \frac{1}{\sqrt{5}}\frac{1}{\bar{\lambda}^{n+2}}$$

and

$$h_n = \frac{n+1}{5\lambda^{n+1}} - \frac{1}{5\sqrt{5}\lambda^{n+1}} + \frac{n+1}{5\bar{\lambda}^{n+1}} + \frac{1}{5\sqrt{5}\bar{\lambda}^{n+1}}.$$

Proof. By (2.2), $I(P_n, z) = I(P_{n-1}, z) + zI(P_{n-2}, z)$. Therefore, $i(P_n) = i(P_{n-1}) + i(P_{n-2})$ and $i'(P_n) = i'(P_{n-1}) + i'(P_{n-2}) + i(P_{n-2})$.

Write $F(z), H(z)$ for the generating functions for $(f_n)_{n\geq 0}, (h_n)_{n\geq 0}$ respectively. We have

$$\begin{aligned} F(z) &= 1 + 2z + z\sum_{n\geq 2} f_{n-1}z^{n-1} + z^2\sum_{n\geq 2} f_{n-2}z^{n-2} \\ &= 1 + 2z + z(F(z) - 1) + z^2F(z); \\ H(z) &= z + z\sum_{n\geq 2} h_{n-1}z^{n-1} + z^2\sum_{n\geq 2} h_{n-2}z^{n-2} + z^2\sum_{n\geq 2} f_{n-2}z^{n-2} \\ &= z + zH(z) + z^2H(z) + z^2F(z). \end{aligned}$$

Hence,

$$F(z) = \frac{1+z}{1-z-z^2} = \frac{1}{\sqrt{5}\lambda^2}\frac{1}{1-z/\lambda} - \frac{1}{\sqrt{5}\bar{\lambda}^2}\frac{1}{1-z/\bar{\lambda}},$$

from which we conclude the formula for f_n. On the other hand,

$$\begin{aligned} H(z) &= \frac{z}{(z^2+z-1)^2} \\ &= \frac{1}{5\lambda}\frac{1}{(1-z/\lambda)^2} + \frac{1}{5\bar{\lambda}}\frac{1}{\left(1-z/\bar{\lambda}\right)^2} - \frac{1}{5\sqrt{5}\lambda}\frac{1}{1-z/\lambda} + \frac{1}{5\sqrt{5}\bar{\lambda}}\frac{1}{1-z/\bar{\lambda}}. \end{aligned}$$

As $[z^n]\frac{1}{(1-az)^2} = (n+1)a^n$, we arrive at the formula for h_n. □

As $|\bar{\lambda}| \approx 1.618$, for large values of n,

$$\begin{aligned} f_n &\approx \frac{1}{\sqrt{5}\lambda^{n+2}}, \\ h_n &\approx \frac{1}{5\lambda^{n+1}}\left(n+1-\frac{1}{\sqrt{5}}\right), \\ p(n) &\approx \frac{\lambda}{\sqrt{5}}\left(n+1-\frac{1}{\sqrt{5}}\right). \end{aligned}$$

Let $s(n) = a(K_{1,n-1}) = \frac{1+(n-1)2^{n-2}}{1+2^{n-1}}$.

Lemma 2.7. *For all integers $i \geq j \geq 1$,*

1. $s(i) + s(j) \leq s(i+j-1) + s(1)$,
2. $p(i) + p(j) \geq p(i+j-1) + p(1) \geq p(i+j-2) + p(2)$;
3. $p(i) < p(i-1) + \mu$ *where* $\mu = 0.28$ *unless* $i \leq 5$;
4. $p(i) + p(j) \geq p(i+j) + 2\epsilon$ *where* $\epsilon = 0.14$ *unless* $i \leq 4$;

Proof. Each of these can be proved by the approximations for large i, j complemented by direct checking of small values. □

Proof of Proposition 2.5. We prove the lower bound by induction on $n = |T|$; the upper bound can be strengthened to Proposition 2.8. By checking the small cases directly, we assume $n \geq 10$ so that we can apply Lemma 2.7 without the exceptional cases.

There exists a vertex v adjacent to an end of two pendant paths of length x and y respectively since $T \neq P_n$. Furthermore, by choosing v with minimal $x+y$, we may assume $x+y \leq 2n/3$. Denote by R the rest of the tree, i.e. T with v and the two paths deleted. Let $r = |R| = n-x-y-1$. Let $A = i(T-v), B = i(T-\Gamma[v])$. Expanding $a(T)$ at v, we get by induction

$$\begin{aligned}
& a(T) - p(n) \\
= \; & \frac{A}{A+B}(a(T-v) - p(n)) + \frac{B}{A+B}(1 + a(T-\Gamma[v]) - p(n)) \\
\geq \; & \frac{A}{A+B}(p(x) + p(y) + p(r) - p(n)) \\
& + \frac{B}{A+B}(1 + p(x-1) + p(y-1) + p(r-1) - p(n)) \\
\geq \; & \frac{A}{A+B}(p(n-1) - p(n) + 2\epsilon) + \frac{B}{A+B}(1 + p(n-4) - p(n) + 2\epsilon) \\
\geq \; & \frac{A}{A+B}(-\mu + 2\epsilon) + \frac{B}{A+B}(1 - 4\mu + 2\epsilon) \\
> \; & 0.
\end{aligned}$$

□

2.2.2 Graphs

In this section, we maximize $a(G)$ over connected graphs G and estimate $a(G)$ for graphs with few edges. Recall that $a(G,z) = zI'(G,z)/I(G,z)$.

Proposition 2.8. *For any $z \geq 1$, integer $n > 0$ and any graph $G \neq K_n^c$, $a(G,z) < a(K_n^c, z) = nz/(1+z)$. Moreover, if $G \ncong K_{1,n-1}$ is connected, then $a(G,z) < a(K_{1,n-1}, z)$.*

Proof. The statement is true for $n = |G| \leq 3$. Suppose G is a disconnected graph with components $G_1, \cdots, G_g$. Then $a(G,z) = \sum a(G_i, z) \leq \sum |G_i| z/(1+z) = nz/(1+z)$ by induction. Moreover the inequality is strict unless $G_i = K_1$ for all i.

Suppose G is connected. Furthermore assume there exists $v \in V(G)$ with $d = \deg(v) \geq 3$. Note that $a(K_{1,n-1}, z) = z\frac{(n-1)(1+z)^{n-2}+1}{(1+z)^{n-1}+z}$. By (2.4) and induction, we

have

$$\begin{aligned}
a(G,z) &= \frac{I(G-v,z)}{I(G,z)}a(G-v,z) + \frac{zI(G-\Gamma[v],z)}{I(G,z)}\left(1+a(G-\Gamma[v],z)\right)\\
&\leq \frac{I(G-v,z)}{I(G,z)}\frac{(n-1)z}{1+z} + \frac{zI(G-\Gamma[v],z)}{I(G,z)}\left(1+\frac{(n-1-d)z}{1+z}\right)\\
&= \frac{(n-1)z}{1+z} + \left(1-\frac{zd}{1+z}\right)\frac{zI(G-\Gamma[v],z)}{I(G,z)}\\
&\leq \frac{(n-1)z}{1+z} + \left(1-\frac{zd}{1+z}\right)\frac{z[1+(n-1-d)z]}{I(K_{1,n-1},z)}\\
&= \frac{z}{I(K_{1,n-1},z)}\left(\frac{z(zd-(2+z))(d-n+1)}{1+z} + (n-1)(1+z)^{n-2}+1\right)\\
&\leq a(K_{1,n-1},z)
\end{aligned}$$

as $\frac{2+z}{z} \leq 3 \leq d \leq n-1$. The second inequality follows from that $I(G,z) \leq I(K_{1,n-1},z)$ and $I(G-\Gamma[v],z) \geq I(K_{n-1-d},z) = 1+(n-1-d)z$. Unless $G-v$ is empty, the first inequality is strict. If $\Delta(G)=2$, then $G-v$ can be empty only when $n=3$. Thus, $a(G,z) \leq a(K_{1,n-1},z)$ with equality if and only if $G = K_{1,n-1}$. □

Proposition 2.9. *Let G be a graph on n vertices and m edges where $m \leq n-1$. Then*

1. *$a(G) \geq \frac{n}{2} - \frac{m}{3}$; equality holds iff each component has order at most 2.*
2. *$a(G) \geq \frac{2(n-1-m)}{3} + p(2m-n+2)$ if further G is a forest and $m > n/2$; equality holds iff $G = P_{2m-n+2} \cup (n-1-m)P_2$.*
3. *$a(G) \leq a(K_{1,m}) + \frac{n-m-1}{2}$; equality holds iff $G = K_{1,m} \cup (n-m-1)P_1$.*

Proof. Let $G_1, G_2, \cdots, G_k$ be the components of G.

1. We prove by induction on $n+m$ which clearly starts. Since the bound is linear in both n and m, we may assume G is connected. If suffices to show that if $\Delta(G) > 1$, then $a(G) > \frac{n}{2} - \frac{m}{3}$. Indeed, let v be a vertex with $d = \deg(v) > 1$ and $e(G-\Gamma[v]) = m-d-c$ for some $c \geq 0$. Let $A = i(G-u)$ and $B = i(G-\Gamma[u])$.

By (2.4) and induction,

$$\begin{aligned} a(G) &\geq \frac{A}{A+B}\left(\frac{n-1}{2}-\frac{m-d}{3}\right)+\frac{B}{A+B}\left(1+\frac{n-1-d}{2}-\frac{m-d-c}{3}\right) \\ &> \left(\frac{n}{2}-\frac{m}{3}\right)+\frac{d}{3}-\frac{1}{2}+\frac{B}{A+B}\left(1+\frac{c}{3}\right) \\ &> \left(\frac{n}{2}-\frac{m}{3}\right). \end{aligned}$$

2. By Proposition 2.5, $a(G_i) \geq p(|G_i|)$. By 2 of Lemma 2.7 having one more P_2 component does not increase a, so we have $a(G) \geq \sum p(|G_i|) \geq p(2m-n+2)+(n-1-m)p(2)$. First inequality is strict if some G_i is not a path; second is if there is more than one component of order at least 3.

3. By Proposition 2.8, $a(G_i) \leq s(|G_i|)$. By 1 of Lemma 2.7, merging two stars does not decrease a, hence we have $a(G) = \sum a(G_i) \leq \sum s(|G_i|) \leq s(m+1)+(n-m-1)s(1)$. First inequality is strict if G_i is not a star; second is strict if there is more than one component of order at least 2.

□

We remark on the problem suggested by John Talbot of maximizing $a(G)$ among cubic graphs. Kahn [53] and Zhao [97] showed that the disjoint union of k copies of $K_{r,r}$ maximizes the number of independent sets over all r-regular graphs of order $2kr$. This suggests that $K_{r,r}$ is extremal with respect to average as well. More precisely, we propose the following.

Conjecture 2.10. *For any $r \geq 1$, given an r-regular graph G,*

$$\frac{a(G)}{|G|} \leq \frac{a(K_{r,r})}{2r}.$$

2.3 Inverse problems

The number of independent sets $i(G)$ is studied as the Fibonacci number [85] in graph theory and as the Merrifield-Simmons index [72] in mathematical chemistry. In certain applications [66], it is of interest to know if an integer can be realized as the index of

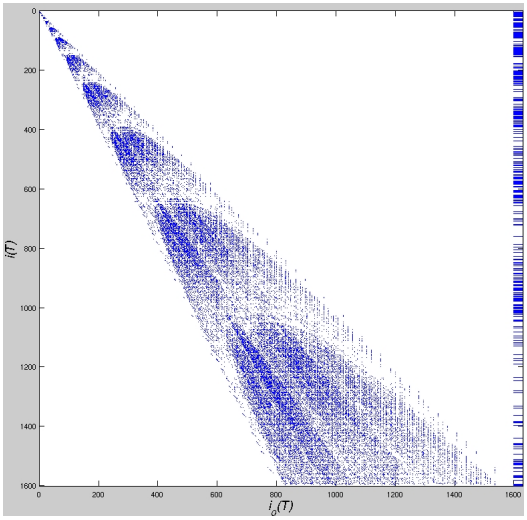


Figure 2.1: A dot indicates the pair $(i(T), i_0(T))$ is constructible. The short segments on the right indicate nonconstructible numbers, i.e. their extensions do not hit a dot

a chemical graph. To be precise, given integer k, the inverse problem asks whether there exists a graph G such that $i(G) = k$. Clearly there does as $i(K_{k-1}) = k$. The question becomes more interesting if we restrict ourselves to certain classes of graphs. For the class of bipartite graphs, Linek [69] answered the question affirmatively. Here we are interested in the class of trees. Note that $i(P_0) = 1, i(P_1) = 2, i(P_3) = 3$ and $i(P_4) = 5$ so that there are no trees corresponding to 4. In the same paper, Linek raised the following conjecture (see also [66]).

Conjecture 2.11 ([69])**.** *Any sufficiently large integer is the number of independent sets of some tree.*

For a rooted tree (T, t), let $i_0(T) = i(T - t)$, which is different from the previous usage as the number of independent sets of size 0. We wrote a program in `Matlab` to compute $(i_0(T), i(T))$ for all rooted trees of order up to 27. A plot of constructible $i_0(T)$ against $i(T)$ shows a curious band feature. From 1 to $i(P_{28}) = 832040$, there are 823 integers without a corresponding tree, the largest of which is 88013. Although there have been extensive studies of $i(G)$ from an extremal perspective, for example [67, 98, 95], some new idea is required for Linek's conjecture.

An interesting paper of Wagner [96] considered the residue of the parameter $i(T)$. Let $\mathcal{T}_n^U$ denote the set of unlabelled trees of order n and $m > 0$ an integer. It was shown via singularity analysis that asymptotically almost every (a.a.e.) tree T sampled uniformly at random (u.a.r.) from $\mathcal{T}_n^U$ has $i(T)$ a multiple of m. If we do not insist on counting precisely the number of trees T with this property, an application of a result of Schwenk [88] immediately proves something slightly stronger, see Section 2.3.3.

A crucial step in Wagner's proof is to show that there is a rooted tree T with both $i_0(T,t)$ and $i(T)$ a multiple of m to begin with, in particular $0 \in \mathbb{Z}_m$ is constructible as the number of independent sets amongst trees. Such a tree was constructed with the Fibonacci sequence and led to the following weaker version of Conjecture 2.11. Let

$$C(m) = \{i(T) \pmod m : T \text{ tree } \}.$$

Conjecture 2.12 ([96]). *For $m \in \mathbb{N}, C(m) = \mathbb{Z}_m$.*

In Section 2.3.1, which is based on [58], we prove Conjecture 2.12. Since the computation of $i(T)$ involves $i_0(T,t) = i(T-t)$ where t is the root, it is natural to prove the following stronger statement. Let

$$D(m) = \{(i_0(T,t), i(T)) \pmod m : (T,t) \text{ rooted tree}\}.$$

Theorem 2.13. *For $m \in \mathbb{N}, D(m) = \mathbb{Z}_m^2$.*

Searching for an intuitive explanation of Wagner's result, Alon, Haber and Krivelevich [2] considered F-matchings which generalize both independent sets and matchings. Let F, G be graphs. An *F-matching* (or *F-packing*) is a subgraph of G consisting of vertex-disjoint copies of F; an *induced F-matching* is such a subgraph which does not induce any edges in G outside the copies of F. Since both versions are multiplicative over components, if there is a component whose number of F-matchings is a multiple of m, then the residue of the whole graph is necessarily zero. Developing this property, the authors of [2, Lemma 8] constructed a limb such that if contained in T, the number of F-matchings in T is a zero residue. Then it was proved a.a.e. tree

sampled u.a.r. from the space of labelled trees $\mathcal{T}_n^L$ of order n contains such a limb, from which it follows that the number of F-matchings in a.a.e. tree is a zero residue. A tree whose number of F-matchings is a multiple of m is required in the proofs and it was constructed using a linear recursion. Similar to Conjecture 2.12, the authors of [2] raised the following question.

Question 2.14. *Given $k, m > 0$ and a tree F, is there a tree T whose number of F-matchings is congruent to k (mod m)?*

The analogous question for induced F-matchings was also raised. While linear recursion seems a natural approach, the divisibility properties may be hard to analyze in general. We answer the two questions positively in a more general setting by constructing trees with no two disjoint copies of F. Indeed, for an integer $r \geq 0$, let an (F, r)-*matching* be a possibly empty collection of copies of F in G such that for any distinct copies F_1, F_2, we have $\min\{\text{dist}_G(x, y) : x \in F_1, y \in F_2\} \geq r$ where dist_G denotes the graph distance in G. Thus an F-matching is an $(F, 1)$-matching and an induced F-matching is an $(F, 2)$-matching. Let $\mathcal{S}_r(F, G)$ be the set of (F, r)-matchings and $s_r(F, G) = |\mathcal{S}_r(F, G)|$.

Theorem 2.15. *For integers $k, m, r \geq 1$ and any tree F, there exist infinitely many trees T such that $s_r(F, T) \equiv k$ (mod m), except $\{s_1(K_1, T)$ (mod m) : T a tree$\} = \mathbb{Z}_m$ iff $m \leq 2$.*

We shall prove Theorem 2.15 and other natural extensions in Section 2.3.2. In Section 2.3.3 we generalize the results in [96] and [2] on the asymptotic behaviour of $s_r(F, T)$ for $r = 1, 2$. While zero is the dominating residue, we will show that the number of trees in other residue classes still grow exponentially in Section 2.3.4. The proof constructs finite trees that control the residue in some limited sense.

An interesting feature appears that the results given for F with diameter at least 3 are easier to prove. In fact for Lemma 2.25, we are unable to construct trees for $F = P_1, P_2$ or P_3 and general r. Perhaps copies of these trees occur rather too easily.

2.3.1 The number of independent sets

Using (2.2) for $(T,t) = \oplus_{i=1}^{k}(T_i,t_i)$, we have [66, 96]

$$i_0(T) = \prod_{k=1}^{j} i(T_k), \tag{2.5}$$

$$i(T) = \prod_{k=1}^{j} i(T_k) + \prod_{k=1}^{j} i_0(T_k,t_k). \tag{2.6}$$

Let $\mu : \mathbb{Z}_m^2 \to \mathbb{Z}_m^2$ be the Fibonacci operator $(a,b) \mapsto (b,a+b)$. The sequence $\left(\mu^k(a,b) : k \in \mathbb{N}\right)$ must contain repeated elements since $\mathbb{Z}_m^2$ is finite. But once it repeats, the sequence becomes periodic. Moreover, as μ is invertible with $\mu^{-1}(a,b) = (b-a,a)$, the sequence is periodic from the start. Denote by $[a,b]$ the orbit of (a,b) under μ.

Let $(a,b)\cdot(c,d) = (ac,bd)$ and $\varphi(T,t) = (i_0(T,t), i(T))$.

Lemma 2.16. *For $m \in \mathbb{N}$ with $C = C(m)$ and $D = D(m)$, we have*

1. *$(a,b) \in D \Rightarrow [a,b] \subset D$;*
2. *$[0,1] \subset D$;*
3. *$(a,b),(c,d) \in D \Rightarrow (ac,bd) \in D$; and*
4. *$c \in C, (a,b) \in D \Rightarrow (ca,cb) \in D$.*

Proof. Let $(T_1,t_1),(T_2,t_2)$ be rooted trees such that $\varphi(T_1,t_1) \equiv (a,b)$ and $\varphi(T_2,t_2) \equiv (c,d) \pmod m$.

1. Consider (T,t) obtained from T_1 by joining a new vertex t to t_1. Then $\varphi(T,t) = (b,b+a)$. Hence, $\mu(a,b) \in D$. Inductively adding a leaf to the root, we see that the whole orbit $[a,b]$ lies in D.
2. Consider the paths: as $i_0(P_1) = 1, i(P_1) = 2$, we have $[0,1] = [1,2] \subset D$.
3. Note that $\varphi((T_1,t_1)\oplus(T_2,t_2),t) = (bd,bd+ac)$. By 1, $[bd,bd+ac] \subset D$ so that $(ac,bd) = \mu^{-1}(bd,bd+ac) \in D$.

4. If $c \in C$, then $(x, c) \in D$ for some x. As $(0, 1) \in D$, it follows by 3 that $(x, c) \cdot (0, 1) = (0, c) \in D$. Then $(c, c) = \mu(0, c) \in D$ and so $(c, c) \cdot (a, b) = (ac, bc) \in D$.

□

Lemma 2.17. *For all $x \in \mathbb{Z}_m$, $(1, x)$ and $(x, 1)$ are in $D(m)$.*

Proof. By Lemma 2.16, $[0, 1] \subset D$. Now, the Fibonacci sequence $[0, 1]$ looks like

$$\cdots, 2, -1, 1, 0, 1, 1, 2, \cdots.$$

Thus, $-1 \in C$ and $(1, 1), (-1, 1), (1, 2), (2, -1) \in D$. Since $-1 \in C$ and $(-1, 1) \in D$, $(1, -1) \in D$.

Suppose $(1, a), (a, 1) \in D$. By Lemma 2.16, each of the following is in D:

$$(1, a) \overset{\mu^{-1}}{\Rightarrow} (a - 1, 1) \overset{\cdot(1,-1)}{\Rightarrow} (a - 1, -1) \overset{\mu}{\Rightarrow} (-1, a - 2) \overset{\cdot(-1,1)}{\Rightarrow} (1, a - 2),$$
$$(a, 1) \overset{\mu}{\Rightarrow} (1, a + 1) \overset{\cdot(-1,1)}{\Rightarrow} (-1, a + 1) \overset{\mu^{-1}}{\Rightarrow} (a + 2, -1) \overset{\cdot(1,-1)}{\Rightarrow} (a + 2, 1).$$

Applying the argument repeatedly to the pairs $(1, 1), (1, 2)$ and $(2, 1)$, we have that $\{(1, 1 - 2b), (1, 2 - 2b), (2 + 2b, 1), (1 + 2b, 1) : b \in \mathbb{Z}_m\} = \{(1, x), (x, 1) : x \in \mathbb{Z}_m\} \subset D$. □

Proof of Theorem 2.13. For $(x, y) \in \mathbb{Z}_m^2$, take $(x, 1)$ and $(1, y)$ in D and multiply. □

We remark that the trees in our construction have maximum degree 3, so that for any integer $m > 0$, $\{i(T) \pmod m : T \text{ a tree}, \Delta(T) \leq 3\} = \mathbb{Z}_m$. This is in contrast to the result of Burr [17] that the Fibonacci numbers (of the paths) form a complete system of residues iff $m = t \cdot 5^k, t = 1, 2, 4, 6, 7, 14$ or 3^j where $k \geq 0, j \geq 1$.

2.3.2 General F-matchings

In this section, we prove Theorem 2.15, i.e. with the exception of $s_1(K_1, T)$, we can represent any residue modulo m as $s_r(F, T)$ for some tree T.

We first introduce some notation. Suppose F is a tree with at least two edges. Let $d = \text{diam}(F) > 1$ be the diameter of F. Take a path $u_0 u_1 \cdots u_d$ of length d in

F. Denote $F - u_d$ by D (for deficit). Also, let λ be the number of leaves adjacent to u_{d-1}.

Given integer $m > 0$, let $M = \Delta(F)!m$. Then for any $l = 1, 2, \cdots, \Delta(F)$, the binomial coefficient $\binom{M}{l}$ is divisible by m. This will allow us to control the formation of copies of F.

Define a tree $W = W(a)$ by identifying u_0 of a copies of F and $M - a$ copies of D. Root W at $w = u_0$.

Since m is fixed, we usually omit '(mod m)' after '$\equiv$'. Recall that a *branch* at v is a maximal subtree of T such that v is a leaf.

Proof of Theorem 2.15. Let $d = \text{diam}(F)$. Suppose $F = K_1$. If $r = 1$, then $s_1(K_1, T) = 2^{|T|}$ and $\{s_1(K_1, T) \pmod m$: T tree $\} = \mathbb{Z}_m$ iff $m \leq 2$. The case for $r = 2$ has been proved in Section 2.3.1. For $r > 2$, since the leaves of $K_{1,n}$ are at distance 2 from each other, at most one vertex can be chosen in any (K_1, r)-matching. Thus, $s_r(K_1, K_{1,n}) = n + 2$. By varying n over $1, 2, \cdots, m$, we are done.

For other F, we construct various trees which do not contain disjoint copies of F.

If $d = 1$, then $F = K_2$. As $s_r(K_2, K_{1,n}) = n + 1$, the result holds.

Suppose $d > 1$, consider $(W, w) = W(a)$ where $a = k - 1$. There are three possibilities for a copy F_0 of F in W.

1. F_0 does not contain w. Then F_0 is contained in a component with strictly fewer vertices than F which is impossible. This case does not occur.

2. F_0 contains exactly one edge incident with w. Then F_0 lies in a branch of W at w. Clearly F_0 cannot be embedded into D, so this branch is isomorphic to F. There are a such copies of F_0.

3. F_0 contains at least two edges incident with w. Then F_0 does not contain any leaves adjacent to u_{d-1} for otherwise there is a path on more than d edges in F_0. Thus the number of such copies f depends only on M but not a. In fact, since the M branches of W within distance $d-1$ of w are isomorphic, f can be written as a sum of terms each of which divisible by $\binom{M}{l}$ for some $1 < l \leq \Delta(F)$.

Indeed, we first choose $l > 1$ branches from M, then we further arrange these branches to form copies of F. By the choice of M, $f \equiv 0$.

As there are no disjoint copies of F in W, we have $s_r(F, T) \equiv 1 + a \equiv k$. □

The above proof motivates a further definition. Let us call a rooted tree (T, t) *symmetric with respect to* F if each isomorphism type of a branch within distance $d - 1$ of t occurs a multiple of $M = \Delta(F)!m$ times. Then as we have just seen, the number of copies of F using at least two edges incident with t is divisible by m. With this notion, it is straightforward to generalize Theorem 2.15 to families of trees without pairwise embeddings.

Corollary 2.18. *Given integer $r > 0$ and trees $F_1, F_2, \cdots, F_l$ such that each F_i is not embeddable into F_j for any $j \neq i$, then for any integers $k_1, k_2, \cdots, k_l > 0$, there exist infinitely many trees T such that $s_r(F_i, T) \equiv k_i \pmod m$.*

Proof. Let $d_i = \mathrm{diam}(F_i)$. Assume that $d_1 \leq d_2 \leq \cdots \leq d_l$. For each F_i, construct $W_i = W_i(a_i)$ by identifying a_i copies of F_i and $M - a_i$ copies of $D_i = F_i - u_{d_i}$ where $M = (\max_{1 \leq i \leq l} \Delta(F_i))!m$ which works for all F_i. Identify the root of each W_i to form T. The nonembeddability condition implies that there are no disjoint copies of F_i in T.

As no copy of F_i contains a leaf u_{d_j} in F_j for $j > i$, by symmetry, $s_r(F_i, T)$ is independent of $a_{i+1}, a_{i+2}, \cdots, a_l$. Therefore, we can choose $a_1 = k_1 - 1, a_2, \cdots, a_l$ sequentially such that $s_r(F_i, T) \equiv k_i$ for all i. □

Theorem 2.15 answers the questions of Alon, Haber and Krivelevich [2] on the number of F-matchings and on induced F-matchings. Note also that by restricting ourselves to trees with no disjoint copies of F in the proof, we avoided multiplication in $\mathbb{Z}_m$ (and hence number theoretic difficulties). However it seems unavoidable if we want to control s_1, s_2 simultaneously. The idea of the following construction is to restrain the formation of a copy of F by supplying a large multiple of m copies of branches that we do not want F to be built on.

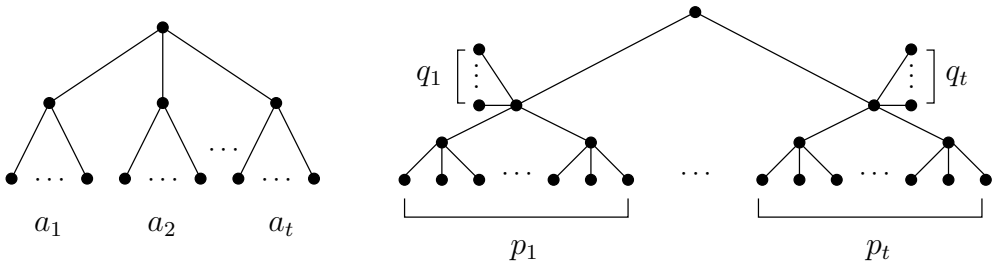


Figure 2.2: T for $F = K_2, K_{1,n}$ respectively in the proof of Theorem 2.19

Theorem 2.19. *For integers $j, k, m > 0$ and a tree F with at least one edge, there exist infinitely many trees T with $s_1(F,T) \equiv j, s_2(F,T) \equiv k \pmod m$.*

Proof. Let $d = \mathrm{diam}(F)$. Suppose $F = K_2$. For integers $a_1, a_2, \cdots, a_t > 0$ with $t > 3$ and $t \equiv k$, let T be a tree obtained by joining a vertex v by an edge to the centre of a copy K_{1,a_i} for each i. Then $s_1(K_2,T) = \prod_{i=1}^{t}(1+a_i) + \sum_{i=1}^{t}\prod_{l\neq i}(1+a_l)$ and $s_2(K_2,T) = \prod_{i=1}^{t}(1+a_i) + t$. For any j, let $a_1 \equiv -1, a_2 \equiv j-1, a_i \equiv 0$ for $i > 2$. Then $s_1(K_2,T) \equiv j, s_2(K_2,T) \equiv k$.

If $d = 2$, then $F = K_{1,n}$ for some $n > 1$. Given $p_1, \cdots p_M, q_1, \cdots, q_M > n$, define $Q(p_i, q_i)$ by identifying the centre of K_{1,q_i} with a leaf of p_i copies of $K_{1,n}$. Let $x_i = s_1(F, Q(p_i,q_i)) = 1 + \binom{p_i+q_i}{n} + p_i$. Then we have

$$\begin{aligned} s_1(F,T) &= \prod_{i=1}^{M} x_i + \sum_{i=1}^{M}\binom{p_i+q_i}{n-1}\prod_{l\neq i} x_l + \sum_{A\subset[n],|A|=M-n}\prod_{i\in A} x_i, \\ s_2(F,T) &= \prod_{i=1}^{M} x_i + \sum_{i=1}^{M}\binom{p_i+q_i}{n-1} + \binom{M}{n}. \end{aligned}$$

Given j, k, let $p_i + q_i \equiv n-1$ for $1 \le i \le k$ and $p_i + q_i$ be some multiple of m such that $\binom{p_i+q_i}{n-1} \equiv 0$ for $i > k$; and $x_i \equiv 1$, for $1 \le i \le M-n-1, x_{M-n} \equiv j, x_j \equiv 0$ for $j > M-n$. This is feasible because after we have chosen a value for $p_i + q_i$, we can adjust that of x_i by changing p_i. Then $s_1(F,T) \equiv j, s_2(F,T) \equiv k$.

Suppose $d > 2$. We construct a tree T whose constituents are X, Y and Z described as follows. For integers $a, b > 0$, let $(X, x) = W(a)$ and $(Y, y) = W(b)$.

To construct Z, we first consider a poset. Let Z' be obtained by identifying $z = u_{d-1}$ of $c > \Delta(F)$ copies of F. Then any copy F_0 of F in Z' contains z. Define a

fragment to be a branch at z of a copy F_0 in Z'. Note that the set $\mathcal{G}$ of isomorphism type of the fragments depends only on F but not c. Define a partial order on $\mathcal{G}$ as follows: for any $G_i, G_l \in \mathcal{G}$, $G_i \leq G_l$ iff G_i can be embedded into G_l as a rooted subtree. Fix a linear extension $(G_1, G_2, \cdots, G_g)$ of $\mathcal{G}$ where $g = |\mathcal{G}|$. Thus $G_1 = K_2$ and G_g is F with all the leaves adjacent to u_{d-1} deleted.

Let $Z_g = Z'$. For $1 < i < g$, form Z_i by attaching copies of G_i to Z_{i+1} such that the number of copies of G_i is a multiple of M. Note that some copies may come from G_l for $l > i$ but not from $l < i$. Define Z by attaching leaves to z in Z_2 such that $\deg_Z(z) = Cm + \lambda$ for some integer C. Identify x with z and add the edge xy to form T (see Figure 2.3). Consider a copy F_0 in T.

1. F_0 lies completely in X or Y. The numbers of such copies are respectively congruent to a and b (mod m).

2. F_0 contains xy and vertices in $X - x$ or $Y - y$. Such a copy does not contain leaves adjacent to u_{d-1} within X or Y. By symmetry, the contribution is 0 (mod m).

3. F_0 contains vertices in both $X - x$ and $Z - z$ but not xy. Again, by symmetry, the contribution is again 0 (mod m).

4. F_0 lies completely in Z. Thus F_0 is formed by choosing fragments in $\mathcal{G}$. If F_0 uses any fragment other than G_1 or G_g, as there are a multiple of M such fragments, the contribution is 0 (mod m). Otherwise F_0 contains exactly one copy G_g so that we need to choose λ leaves at z. Since there are $Cm + \lambda - 1$ available edges (one is used by G_g) incident with z in Z and $\binom{Cm+\lambda-1}{\lambda} \equiv 0$, the contribution is also 0 (mod m).

5. F_0 contains xy with the rest in Z. By the same reasoning as 4, we may assume F_0 chooses a copy of G_g. To complete F_0, we need to choose only $\lambda - 1$ edges incident with z because xy accounts for one leaf already. Since $\binom{Cm+\lambda-1}{\lambda-1} \equiv 1$ and there are c copies of G_g to choose from, the number of copies is congruent to c (mod m).

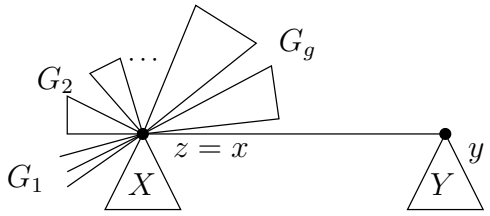


Figure 2.3: T for F with diameter $d > 2$ in the proof of Theorem 2.19

Since any copy uses either x or y, $s_1(F,T) \equiv (1+a)(1+b)+c$ and $s_2(F,T) \equiv 1+a+b+c$. By choosing $a \equiv -1, b \equiv k-j, c \equiv j$, we have $s_1(F,T) \equiv j, s_2(F,T) \equiv k$. □

We conjecture that the theorem is true for general r.

Conjecture 2.20. *Given integers $r, m \geq 1$ and a tree F with at least one edge, for any $k_1, \cdots, k_r \in \mathbb{Z}_m$, there exists a tree T such that $s_i(F,T) \equiv k_i \pmod m$ for all $1 \leq i \leq r$.*

2.3.3 Asymptotic behaviour of divisibility properties

We next generalize the results in Wagner [96] and Alon, Haber and Krivelevich [2] to general r and to a polynomial version. In fact, the proof of the generalization is based on [2] and consists of two parts as well. Given trees F, T, recall that $\mathcal{S}_r(F,T)$ denotes the set of (F,r)-matchings in T, define the (F,r)-*matching polynomial* as

$$S_r(F,T,z) = \sum_{S \in \mathcal{S}_r(F,T)} z^{|S|}$$

where $|S|$ denotes the number of copies of F in T. Thus $S_2(K_1,T,z)$ is the *independent set polynomial* and $S_1(K_2,T,z)$ is the *matching polynomial*.

Recall that a *limb* L at v consists of a number of branches at v. For a subset $X \subset T$, let $\Gamma_{r-1}(X)$ denote the set of vertices within distance $r-1$ in T. The following is analogous to Lemma 8 in [2]; we include a proof for completeness.

Lemma 2.21. *Given trees F, H and integer $r > 0$, there exists a rooted tree (L,l) such that if a tree T contains L as a limb at l, then $S_r(F,H,z)$ divides $S_r(F,T,z)$.*

Proof. Let L be defined by joining a vertex v to a copy of H by an edge and $\Delta(F)+1$ copies by a path of length r. Suppose T contains L as a limb at v. Consider (F,r)-matchings not containing v. Since there is a component of H in $T-v$, these matchings contribute a multiple of $S_r(F,H,z)$. On the other hand, deleting a copy F_0 of F containing v along with $\Gamma_{r-1}(F_0)$ always leaves a forest with an H component. As there are finitely many possibilities for $L-F_0-\Gamma_{r-1}(F_0)$, the contribution is also a multiple of $S_r(F,H,z)$. Thus, the sum $S_r(F,T,z)$ is divisible by $S_r(F,H,z)$. □

The probabilistic part is the following classical result of Schwenk's,

Theorem 2.22 ([88])**.** *Given any rooted tree (L,l), a.a.e. unlabelled tree in $\mathcal{T}_n^U$ contains L as a limb at l.*

The following analogue of Theorem 2.22 for labelled trees was proved in [2] by concentration inequalities. For the sake of variation, we sketch a proof using generating functions (for background, we refer to [32]). A rooted tree L occurs as a limb in another rooted tree T if L consists of branches of descendants at some vertex; but if $L=T$, we do not consider L as a limb. For a power series $A(z)=\sum_{n\geq 0}a_n z^n$, let $\mathrm{rad}(A)$ be its radius of convergence.

Theorem 2.23. *Given rooted tree L, a.a.e. labelled tree in $\mathcal{T}_n^L$ contains L as a limb.*

Proof. Let $l=|L|>1$ and ρ be the size of the automorphism group of L as a rooted tree. Let $T(z,u)$ be the EGF of rooted trees with u counting the occurrence of L as a limb. Then $T(z,u)=z\exp\left(T(z,u)+(u-1)\frac{z^l}{\rho}\right)$. Let $G(z)=T(z,0)$, i.e. the EGF for rooted trees forbidding L as a limb. Then

$$G(z)=z\exp\left(G(z)-\frac{z^l}{\rho}\right).$$

Let $g(z)=z\exp\left(-\frac{z^l}{\rho}\right)$. As $T(g(z),1)$ is a function analytic around $z=0$ and satisfies the equation for G with $T(g(0),1)=G(0)$, it is the local solution for $G(z)$. It is well known (by Cayley's and Stirling's formulæ) that $\mathrm{rad}(T(z,1))=1/e$. Hence

$$\mathrm{rad}(G)=\mathrm{rad}(T(g(z),1))=\sup\{z>0:g(x)<1/e,\forall 0\leq x<z\}>1/e.$$

This implies that there are exponentially fewer rooted trees without L as a limb than the total number of rooted trees. As we can root a tree in n ways, the probability of a random unrooted tree without L as a limb is at most n times the corresponding probability for a rooted tree sampled uniformly, which tends to 0 exponentially fast. □

By Theorem 2.15, for any $m > 0$, there exists a tree H such that m divides $s_r(F, H) = S_r(F, H, 1)$. Thus the following generalizes the results in [96] and [2].

Corollary 2.24. *Given trees F, H and integer $r > 0$, a.a.e. tree $T \in \mathcal{T}_n^U$ has the property that $S_r(F, H, z)$ divides $S_r(F, T, z)$. The same is true when T is sampled from $\mathcal{T}_n^L$.*

Proof. It follows directly from Lemma 2.21, Theorems 2.22 and 2.23. □

2.3.4 Trees in other residue classes

Returning to the modulo m case, it was shown in [96] that the number of trees with an odd number of independent set is small compared to those with an even number but it still grows exponentially fast. While the use of 'nullifying trees' in [2] gives a quick proof that the number of F-matchings of a random tree is a.a.s. a zero residue, it is not clear whether the other residue classes still grow exponentially. We show that this is the case for trees F with diameter at least 3.

Lemma 2.25. *Given a tree F with $|F| > 3$ and integers $r, k, m > 0$, then for any sufficiently large n, there exists a tree T of order n such that $s_r(F, T) \equiv k \pmod m$.*

Proof. Given F, m and r, for each k, construct $W_k = W(k-1)$ as in Section 2.3.2 such that $s_r(F, W_k) \equiv k$. Let $N = \max_{1 \le k \le m} |W_k|$. Consider an integer $n > N$.

If $\mathrm{diam}(F) = 2$, since $|F| > 3$, $F = K_{1,f}$ for some $f > 2$. By adding a path on $n - |W_k|$ vertices to any leaf of W_k, we obtain a tree T of order n such that $s_r(F, T) \equiv k$.

If $\mathrm{diam}(F) > 2$, then F is not a star. Therefore attaching $n - |W_k|$ leaves to the root of W_k does not change the residue of $s_r(F, W_k)$. Hence we can span all residues of m with trees of order n. □

In the following, we restrict ourselves to $r = 1$ and F with diameter at least 3. Let $\mathcal{A}(k)$ denote the family of (unlabelled) rooted trees T such that $s_1(F, T) \equiv k$ and $\mathcal{A}_n(k)$ the set of such trees of order n. Let $a_n(k) = |\mathcal{A}_n(k)|$ and $A(k, z) = \sum_{n\geq 0} a_n(k) z^n$. We shall prove that $\mathrm{rad}(A(k, z)) < 1$ which implies that $a_n(k)$ grows exponentially fast. Let $\mathcal{R}^U$ be the class of unlabelled rooted trees and $R^U(z)$ be its OGF. It is well known (see for example [32, Chapter 1] for a derivation) that

$$R^U(z) = z \exp\left(\sum_{i\geq 1} \frac{R^U(z^i)}{i}\right). \tag{2.7}$$

For a rooted tree (T, t), let $s_1(F, T^+) = s_1(F, T) - s_1(F, T - t)$.

Theorem 2.26. *For a tree F of diameter at least 3 and integer $m \geq 1$, there exists $C > 1$ such that for any integer k and sufficiently large integer n, the number of trees T of order n such that $s_1(F, T) \equiv k \pmod m$ is at least C^n.*

Proof. We first reduce the general case to $k = 1$. Let $H \in \mathcal{A}(1)$ and $W_k \in \mathcal{A}(k)$ such that $|W_k| = n - M|H|$ as constructed in Lemma 2.25. Join the root of W_k to M copies of H to form T. It is easy to check that $T \in \mathcal{A}(k)$. In other words, given a tree $H \in \mathcal{A}(1)$, we can construct a tree of order $M|H| + |W_k|$ in $\mathcal{A}(k)$. Hence, $a_n(k) \geq a_{(n-|W_k|)/M}(1)$. Since M and $|W_k|$ are constant, to prove the theorem, it suffices to show that the number of choices for H grows exponentially fast.

To this end, we construct a rooted tree (V, v) such that if v is joined to any collection of trees in $\mathcal{A}(1)$, the resulting tree remains in $\mathcal{A}(1)$. We require (V, v) to satisfy:

1. $s_1(F, V - v) \equiv 0$;
2. $s_1(F, V^+) \equiv 1$;
3. V is symmetric.

Let $(W, w) = W(m-1) \in \mathcal{A}(0)$ and D be constructed as in Section 2.3.2. Identify u_{d-1} of D with w to form (C, c) (for complete). Delete vertices in W which are at distance d of w; and identify w with u_{d-1} of D to form (I, i) (for incomplete). Define V by adding a new vertex v joined to the root of a copy of C and $M-1$ copies of I (see Figure 2.4).

Clearly V is symmetric. Since $W \in \mathcal{A}(0)$ and D does not contain a copy of F, $C \in \mathcal{A}(0)$. As $V - v$ contains C as a component, $s_1(F, V - v) \equiv 0$. Consider a copy containing v. By symmetry of V and the fact that $C \in \mathcal{A}(0)$, we may assume that the copy contains e as the only edge incident with v. Since W is symmetric as well, we may further assume that no edge in W is involved. Thus there is exactly one such copy, namely, $D \cup e$. Hence, $s_1(F, V^+) \equiv 1$.

We claim that if we join any collection of trees in $\mathcal{A}(1)$ to the root v of V by an edge, the resulting tree remains in $\mathcal{A}(1)$. Indeed, for $H_1, \cdots, H_p \in \mathcal{A}(1)$, let $(H, h) = \oplus_{i=1}^{p} H_i$. Define T by identifying h and v. Then

$$s_1(F, T) \equiv s_1(F, V) \prod_{i=1}^{p} s_1(F, H_i) + s_1(F, V - v) s_1(F, H^+) \equiv 1,$$

where the first term counts F-matchings which do not contain any edge in H incident with v and the second counts the rest so that v is not used in F-matchings in V. To repeat, the tree defined by joining a collection of trees in $\mathcal{A}(1)$ to the 'large root' V is in $\mathcal{A}(1)$. Hence, the coefficients of $A(1, z)$ dominate those of $B(z)$ where

$$B(z) = z^{|V|} \exp\left(\sum_{i \geq 1} \frac{B(z^i)}{i} \right).$$

Suppose $R^U(z)$ satisfies (2.7), then the solution around $z = 0$ is $B(z) = R^U(z^{|V|})$. Thus

$$\mathrm{rad}(A(1, z)) \leq \mathrm{rad}(B) = \mathrm{rad}(R^U)^{1/|V|}.$$

It was proved by Pólya [84] that $\mathrm{rad}(R^U) < 1$. Hence, $\mathrm{rad}(A(1, z)) < 1$. It follows that $a_n(1)$ grows exponentially fast and so does $a_n(k)$ for any k. □

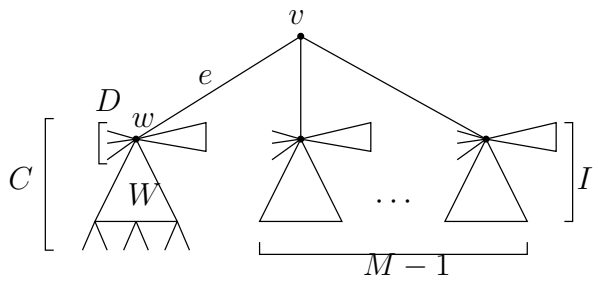


Figure 2.4: Construction of V in Theorem 2.26

2.3.5 Maximum number of (F, r)-matchings

It was proved in [85] that the star and the path maximize the number of independent sets and matchings respectively among trees of order n, with both $s_2(K_1, T)$ and $s_1(K_2, T) < 2^n$. It would be interesting to study the extremal trees for other pairs (F, r). Here we start in this direction by showing the following. Let $\mu(F, r, n) = \max\{s_r(F, T) : T \text{ is a tree of order } n\}$.

Proposition 2.27. *For any $r \geq 1$ and tree F, $\lambda(F, r) = \lim_{n\to\infty} \mu(F, r, n)^{1/n}$ exists and $1 < \lambda(F, r) \leq 2$.*

Proof. Let $\mu_n = \mu(F, r, n)$. We may assume $f = |F| > 2$. Since joining two trees by a path of length r preserves (F, r)-matchings, it follows that $\mu_{n+k+r-1} \geq \mu_n \mu_k$ for all integers $n, k \geq 0$. By an adaptation of Fekete's Lemma, $\lambda = \lim_{n\to\infty} \mu_n^{1/n} \leq \infty$ exists. By taking disjoint copies of F and joining them to a vertex with a path of length r, we obtain a tree T with $s_r(F, T) > c^n$ for some constant $c > 1$. Hence $\lambda > 1$. On the other hand, as an (F, r)-matching corresponds to an edge subset, $\mu_n \leq 2^{n-1}$. Hence, $\lambda \leq 2$. □

It would be interesting to study extremal questions on λ. But in the spirit of inverse problems, we ask the following.

Question 2.28. *Let $\Lambda = \{\lambda(F, r) : F \text{ tree}, r \geq 1\}$. Is Λ dense in $[1, 2]$?*

For example, we show that 1 is an accumulation point. Indeed, consider $(K_1, 2p)$-matchings in a tree T where $n = (q + 1)p = |T|$ for some $p, q \geq 1$. Let r_k be the

number of $(K_1, 2p)$-matchings of size k. It is easy to show by induction on k that for any $(K_1, 2p)$-matching H of size $k \geq 1$, there are at least $(k+1)p$ vertices in T within distance $2p-1$ of H. Then $r_{k+1}(k+1) \leq r_k(n-(k+1)p)$ since each $(K_1, 2p)$-matching of size k blocks at least $(k+1)p$ choices for a vertex addition to the matching. Therefore, with $r_1 = n = (q+1)p$,

$$\begin{aligned} r_{k+1} &\leq \frac{n-(k+1)p}{k+1}\frac{n-kp}{k}\cdots\frac{n-2p}{2}r_1 \\ &\leq \frac{p^{k+1}(q-k)(q-(k-1))\cdots(q-1)(q+1)}{(k+1)!} \\ &\leq \binom{q+1}{k+1}p^{k+1}. \end{aligned}$$

Hence,

$$s_{2p}(K_1, T) \leq \sum_{k\geq 0,\ (k+1)p<n} \binom{q+1}{k+1}p^{k+1} \leq (1+p)^{q+1}.$$

Then $\lambda(K_1, 2p) \leq (1+p)^{1/p}$ which tends to 1 as $p \to \infty$.

Chapter 3

Graph reconstruction with polynomials

When a graph polynomial is considered as a graph invariant, there are essentially two questions. Firstly, what can be deduced from the given polynomial; and secondly, whether two polynomials carry the same information. There is a great variety of articles on either question, for example [34, 38, 71, 75, 76, 77, 87].

Given the wide spectrum of graph polynomials, we have to restrict ourselves to only two types. The first one is subgraph counts which enumerate subgraphs in a family according to size. Examples include the independent set polynomial which enumerates independent sets in a graph; and the path sequence which counts the number of paths of various lengths.

The second type originates from partition functions in statistical physics which record the weight of each homomorphism between weighted graphs. For example, the Ising polynomial counts homomorphisms from a graph to a looped K_2. See the survey [14] for a wealth of results about weighted graph homomorphisms.

Notable omissions in our study include polynomials that satisfy a contraction-deletion relation, the banner of which is the Tutte polynomial. The Tutte polynomial manifests in a wide range of areas and has various specializations and generalizations. As an invariant, there accompanies a large body of literature on graphs determined by their Tutte polynomial or its relatives, for example, [34, 73]. In this direction,

Bollobás, Pebody and Riordan [12] studied general contraction-deletion invariants and conjectured that a.a.e. graph is Tutte-unique.

After introducing the definitions and giving a method to generate p-equivalent graphs in Sections 3.1 and 3.2 for some polynomials p, we study the independent set polynomial (Section 3.3) and the Ising polynomial (Section 3.4). In particular, we prove that in both cases, there are no local modifications that produce equivalent graphs in general.

Section 3.4 is joint work with Alex Scott.

3.1 Definitions

Let G, H be graphs. A *graph homomorphism* is a map $\phi : V(G) \to V(H)$ that preserves adjacency. We also write G for $V(G)$. A *weight* on H is a function from $V(H) \cup E(H)$ to some polynomial ring. For our purposes, as a nonedge corresponds to an edge with zero weight, we may assume H is complete with loops at each vertex and ϕ is any map. Let the *homomorphism count* from G to H be

$$\hom(G, H) = \sum_{\phi: G \to H} \prod_{u \in G} w_{\phi(u)} \prod_{uv \in E(G)} w_{\phi(u)\phi(v)}.$$

For a partial map $\psi : A \subset G \to H$, let

$$\hom(G, H; \psi) = \sum_{\phi: G \to H,\ \phi|_A = \psi} \prod_{u \in G} w_{\phi(u)} \prod_{uv \in E(G)} w_{\phi(u)\phi(v)}.$$

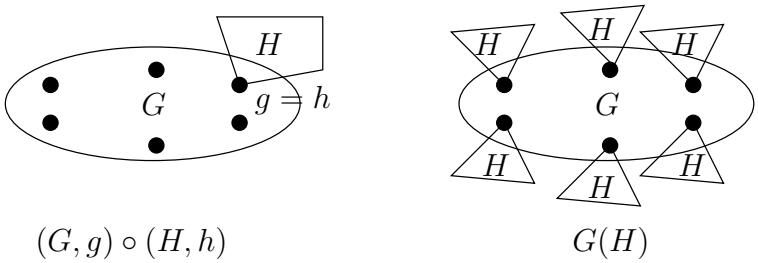
If $A = \{g\}$ is understood, we write $\hom(G, H)_i$ for $\hom(G, H; \psi(g) = i)$.

Let $\mathcal{F}(G)$ be a family of subgraphs of G. A *subgraph count* associated to $\mathcal{F}(G)$ is defined as

$$p(G, z) = \sum_{F \in \mathcal{F}(G)} z^{|F|}.$$

Note that a subgraph count may also be a homomorphism count. Indeed, let K_n° be the complete graph of order n with a loop at each vertex. Then the independent set polynomial is equal to $\hom(\cdot, K_2^\circ)$ with $w_1 = z, w_2 = w_{22} = w_{12} = 1, w_{11} = 0$.

For a graph polynomial p, two graphs G and H are called *p-equivalent* if $p(G) = p(H)$. In general it appears difficult to show that a polynomial p is equivalent to a



$(G, g) \circ (H, h)$ $G(H)$

Figure 3.1: Coalescence and decorations

homomorphism count $\hom(\cdot, G)$, by which we mean, F and F' are p-equivalent iff $\hom(F, G) = \hom(F', G)$. (There is a characterization when a graph parameter is equal to the number of homomorphisms to some graph G, see the references in [14].) The question is to find a fitting G: for trees, the path sequence is determined by the Ising polynomial which is a much finer invariant.

For a graph polynomial p, two graphs G and H are called *p-equivalent* if $p(G) \equiv p(H)$.

3.2 Decorations

Recall that for rooted graphs $(G, g), (H, h)$, $(G, g) \circ (H, h)$ denotes the coalescence obtained by identifying g and h. Define the *decoration $G(H)$ of G by H* by identifying each vertex of G by a copy of (H, h).

Let K be a graph with weight w and $V(K) = [\kappa]$. Note that $\hom(G, K)$ is a polynomial in indeterminate w_i, w_{ij} for $1 \leq i, j \leq \kappa$. Consider a homomorphism $\psi : G \to K$ such that $\psi(v) = i$. Any extension of ψ to $G(H)$ maps the root h of the copy of H attached at v to i. Thus, in effect, we have mapped a vertex of weight $\hom(H, K; \psi(h) = i)$ to i. Hence, $\hom(G(H), K)$ is $\hom(G, K)$ evaluated at $w_i = \hom(H, K)_i$. This implies the following.

Proposition 3.1. *Suppose G_1, G_2, K satisfy $\hom(G_1, K) = \hom(G_2, K)$. Then $\hom(G_1(H), K) = \hom(G_2(H), K)$ for any rooted graph H.*

Next, we consider decorations on a class of subgraph counts p. Suppose p counts

the subgraphs in $\mathcal{F}(G)$. Let (H, h) be a rooted graph. Assume that

$$\text{each } F \in \mathcal{F}(G) \text{ extends to subgraphs in } \mathcal{F}(G(H)) \text{ which are enumerated by a rule dependent on } |F| \text{ and } H \text{ but not specific } F. \tag{3.1}$$

For example, if $\mathcal{F}(G)$ is the family of independent sets, then an independent set in $G(H)$ is an independent set I in G such that each $v \in I$ is extended to $I(H - \Gamma_H[h], z)$ and each $u \notin I$ to $I(H - h, z)$ which depends only on $|I|$ and H.

Proposition 3.2. *Suppose $p(G, z)$ counts the subgraphs in $\mathcal{F}(G)$ which satisfies (3.1). Then for any rooted graph H, $p(G_1, z) = p(G_2, z)$ implies $p(G_1(H), z) = p(G_2(H), z)$.*

We remark that the condition on $\mathcal{F}(G)$ is satisfied by all the level one polynomials to be introduced in Section 4.1.

3.3 Independent set polynomial

Recall that the *independent set polynomial* is $I(G, z) = \sum_{k=0}^{\alpha(G)} i_k(G) z^k$ where $i_k(G)$ is the number of independent sets of order k. Through the first terms of the independent set polynomial $I(G, z)$, we can infer the order, size, the number of triangles in the complement. Yet, there are many properties that are not determined. For example, the number of triangles (C_3 with a leaf attached to one vertex and P_4) and connectivity ($C_4 \cup K_1$ and the *fork*, i.e. the tree on 5 vertices with 3 leaves) are not.

Information contained in the first few coefficients of $I(G, z)$ can be summarized as follows.

Proposition 3.3 ([21])**.** *Let G be graph of order n, size m having $t(G)$ triangles. Then $[z^0]I(G, z) = 1$, $[z^1]I(G, z) = n$, $[z^2]I(G, z) = e(K_n) - m$ and $[z^3]I(G, z) = \binom{n}{3} - m(n-2) + \sum_{v \in V(G)} \binom{\deg(v)}{2} - t(G)$.*

One way to construct p-equivalent graphs is by local switching. We shall see more about switching in trees in the next chapter. But here is a negative result of switching with respect to $I(\cdot, z)$.

Theorem 3.4. *Suppose H_1, H_2 are graphs on the same vertex set with the property that: for any graph G containing H_1 as an induced subgraph, $I(G,z) = I(G',z)$ where G' is obtained by switching H_1 to H_2. Then $H_1 = H_2$.*

Proof. It suffices to show that H is determined by $I(G,z)$ of graphs G containing H as an induced subgraph. Let $u, v \in V(H)$. Attach a leaf to H at u to form a graph H_1. Then $I(H_1,z) = I(H,z) + zI(H-u,z)$. Thus, $I(H-u,z)$ can be deduced which in turn determines $\deg_H(u)$. Similarly, $\deg_H(v)$ can be deduced. Add a vertex w to H adjacent to only u, v to form H_2. Then $I(H_2,z) = I(H,z) + zI(H-u-v,z)$. Thus, $\deg_H(u) + \deg_H(v) - A(u,v)$ is determined, where $A(u,v) = 1$ if $uv \in E(G)$, zero otherwise. Hence, H is determined uniquely. □

For a typical random graph $\mathcal{G}(n,1/2)$, Noy [77] showed that there are simply not enough independent set polynomials for each graph so that a.a.e. graph is not determined by the polynomial. However, it can be checked that at the two ends, when either the graph or its complement is a forest with small components, the graph is determined by the polynomial.

3.4 Ising polynomial

Given a graph, the Ising model assigns two states, customarily positive and negative spins, to each vertex. An assignment of states is given a Hamiltonian which depends on the number of positive vertices and the number of edges whose ends have different states. The partition function of the model sums the Hamiltonian over all assignments and is given by the *Ising polynomial* defined by Lee and Yang [62] as

$$P(G,a,z) = \sum_{\substack{(A,B) \text{ partitions } V(G) \\ |A|=k, e(A,B)=b}} a^b z^k.$$

They proved interesting results on the location of the zeros which are critical in the study of phase transition of the model.

Recently, Andrén and Markström [4] considered a version of the polynomial as a Laurent polynomial and studied it from a graph theoretic point of view. They defined

the Ising polynomial as

$$Z(G,x,y) = \sum_{\sigma:V(G)\to\{-1,1\}} \prod_{uv\in E(G)} x^{\sigma(u)\sigma(v)} \prod_{v\in V(G)} y^{\sigma(v)}.$$

That is, $Z(G,x,y) = \hom(G, K_2^\circ)$ with $w_1 = y, w_2 = y^{-1}, w_{11} = w_{22} = x, w_{12} = x^{-1}$. Note that $Z(G,x,y) = x^m y^n P(G, x^{-2}, y^{-2})$ where $n = |G|$ and $m = e(G)$.

The Ising polynomial of a graph captures much combinatorial information [4] but there exists a pair of graphs of order 7 with the same Ising polynomial. Call two graphs *isomagnetic* if they have the same Ising polynomial; and that a graph is *magnetically unique* if any isomagnetic graph is necessarily isomorphic. By a rotor construction and a result on the distribution of a given triangulation in a random triangulation, Andrén and Markström [4] showed that a random triangulation is a.a.s. not magnetically unique. They also conjectured the same for $G \in \mathcal{G}(n, p(n))$ for $1/n \le p(n) \le 1/2$. In view of this conjecture, it is of interest to consider the following question.

Question 3.5. *Does there exist a finite weighted graph K by indeterminates such that* $\hom(\cdot, K)$ *determines a.a.e. graph?*

We may assume that each vertex, loop and edge of K receives a different variable, so that $\hom(\cdot, K)$ encodes the order, size of each colour class and the number of edges between each pair of classes for every vertex colouring with $|K|$ colours. Let

$$\Phi(n) = \min\{k : \exists G \ncong H, |G| = |H| = k+1, \hom(G, K_n^\circ) = \hom(H, K_n^\circ)\}. \tag{3.2}$$

Clearly, $\Phi(n) \ge n$ because we can colour every vertex with a distinct colour. It seems likely that $\Phi(n)$ is finite for each n.

Conjecture 3.6. *For each n, there exist nonisomorphic G, H such that* $\hom(G, K_n^\circ) = \hom(H, K_n^\circ)$.

If the conjecture is true, it would be interesting to study the growth of $\Phi(n)$.

In Section 3.4.1, we first exhibit families of magnetically unique graphs. Then we study two ways of constructing isomagnetic graphs in Section 3.4.2 and prove that no switching is possible to construct isomagnetic graphs in general.

3.4.1 Magnetically unique graphs

A graph G is called *magnetically unique* if it is determined by $P(G, a, z)$. Our aim is to exhibit some families of magnetically unique graphs. We first reproduce the information encoded by the Ising polynomial.

Theorem 3.7 ([4]). *The following can be deduced from the Ising polynomial of a graph G:*

1. *the order n and size m;*
2. *the degree sequence;*
3. *the size of each component;*
4. *the girth;*
5. *the number of subgraphs with i edges and j odd vertices; hence, the number of Eulerian subgraphs with i edges;*
6. *the matching polynomial; in case G is a tree, this is the same as characteristic polynomial;*
7. *the independent set polynomial if G is regular.*

Furthermore, by 5 in Theorem 3.7, we can deduce the number of copies of C_3, C_4 and C_5.

By the theorem, it follows immediately that $K_n, K_{p,q}$, 2-regular graphs, trees which are determined by their spectra, regular graphs which are determined by the independent set polynomials are magnetically unique.

Regular complete multipartite graphs are also magnetically unique. Indeed, if G is $(k-1)p$ regular and there are k bipartitions of the vertices into $V_0 \cup V_1$ such that $|V_0| = p$ and $e(V_0, V_1) = p^2(k-1)$, then G has k independent sets of size p so that G is a complete multipartite graph. Similarly, a complete multipartite graph in which each class has a distinct size is also magnetically unique.

As a further example, we prove that the hypercubes are magnetically unique. The proof is analogous to the result of de Mier and Noy [73] that hypercubes are Tutte-unique. A characterization of the hypercube Q_d is given by Mulder [74].

Lemma 3.8. *A connected d-regular graph is isomorphic to Q_d iff it has 2^d vertices and every pair of vertices at distance 2 have precisely two common neighbours.*

Theorem 3.9. *The hypercube Q_d is magnetically unique for all d.*

Proof. Suppose G is isomagnetic to Q_d. Then G is connected and d-regular. By Lemma 3.8, it suffices to show that every pair of vertices at distance 2 have precisely two common neighbours.

Since the girth of G is the same as that of Q_d, G is triangle-free. Note that $K_{2,3}$ is the unique triangle-free graph of order 5 and size 6. Since $K_{2,3} \not\subset Q_d$, $[a^{5d-12}z^5]P(G,a,z) = 0$. In other words, any two vertices in G at distance 2 have at most two common neighbours.

Counting the number of unordered paths on 3 vertices, we have

$$2^d\binom{d}{2} = \sum_{\{x,y\}:d(x,y)=2} |\Gamma(x)\cap\Gamma(y)|.$$

As there are $2^{d-2}\binom{d}{2}$ copies of C_4 in Q_d and hence in G, there are at least four times as many P_3. This accounts for all copies of P_3. Therefore, every pair of vertices at distance 2 have two common neighbours. □

We remark that one can also apply the proof in [73] to show that a wheel on at least 7 vertices is magnetically unique. For further related results, we refer to Garijo, Goodall and Nešetřil [34].

3.4.2 Isomagnetic graphs

We describe two ways of constructing isomagnetic graphs. The first one is the sum of two graphs. Recall that the *sum* $G+H$ of two graphs is obtained by adding all edges between disjoint copies of G and H. It can be easily checked (or using [4, Theorem 3.2]) that if G_1, G_2 are isomagnetic, then so are G_1+H, G_2+H for any graph H.



Figure 3.2: A smallest pair of strongly isomagnetic graphs where the root is the outlined vertex

The second way uses strongly isomagnetic graphs. For a rooted graph (G, g), let the *rooted Ising polynomial* be

$$P^+(G, a, z) = \sum_{(A,B) \text{ partitions } G \text{ with } g \in A} a^{e(A,B)} z^{|A|}.$$

Write $P^-(G, a, z)$ for the corresponding sum with $g \notin A$. Two rooted graphs are called *strongly isomagnetic* if their rooted Ising polynomials are equal. As $P^-(G, a, z) = P^+(G, a, 1/z)z^{|G|}$, strongly isomagnetic graphs are isomagnetic. A pair of strongly isomagnetic graphs is given in Figure 3.2. Recall that for rooted graphs G and H, $G \circ H$ denotes coalescence.

Lemma 3.10. *The following statements are equivalent:*

1. *G_1 and G_2 are isomagnetic and so are $(G_1, r_1) \circ K_2$ and $(G_2, r_2) \circ K_2$ where K_2 is the rooted edge graph.*

2. *$(G_1, r_1), (G_2, r_2)$ are strongly isomagnetic.*

3. *$(G_1, r_1) \circ H$ and $(G_2, r_2) \circ H$ are isomagnetic for any rooted graph H.*

Proof. 1. $\Rightarrow$ 2. : It suffices to show that $P^+(G, a, z)$ is determined by $P(G, a, z)$ and $P(G \circ K_2, a, z)$. Indeed, let

$$P^+ = P^+(G, a, z) = zf_1(a) + z^2 f_2(a) + \cdots + z^{n-1} f_{n-1}(a) + z^n$$

and $f_0 = 0, f_n = 1$. Then $P^- = P^-(G, a, z) = 1 + zf_{n-1} + z^2 f_{n-2} + \cdots f_1 z^{n-1}$. As $P(G \circ K_2, a, z) = (a + z)P^+ + (az + 1)P^-$, we have

$$[z^k]P(G \circ K_2, a, z) = af_k + f_{k-1} + af_{n-k+1} + f_{n-k}$$

for all $k = 1, 2, \cdots, n$.

Note that for $k = 1$, we can solve for f_1 and f_{n-1} from $af_1 + f_{n-1} + 1$ and $[z^1]P(G, a, z) = f_1 + f_{n-1}$. Proceed to $k = 2$ and solve for f_2 and f_{n-2}; and so on. Inductively, we can compute $P^+(G, a, z)$.

2. $\Rightarrow$ 3. : there is clearly a bijection of the states.

3. $\Rightarrow$ 1. : trivial. $\square$

For some graph polynomial p, it is possible to produce p-equivalent graphs by local modification. For example, flipping the rotor graph R given in [4] preserves the Ising polynomial of a planar graph containing R. More generally, we look for a pair of graphs H_1, H_2 on the same vertex set H such that H_1 is an induced subgraph of G, the graph G' obtained by deleting the edges of H_1 and adding those of H_2 is isomagnetic to G. As a.a.e. graph in $G \in \mathcal{G}(n, 1/2)$ contains H_1 as an induced subgraph, we would then be able to conclude that a.a.e. graph is not magnetically unique. However, such a pair does not exist. See [12] for analogous results for the Tutte polynomial and the chromatic polynomial.

Theorem 3.11. *Suppose H_1, H_2 are two graphs on the same vertex set H such that whenever G contains H_1 as an induced subgraph, switching H_1 with H_2 gives a graph isomagnetic to G. Then $H_1 = H_2$.*

Proof. It suffices to show that H is determined by $P(G, a, z)$ for graphs G containing H as an induced subgraph. Let $u, v \in V(H)$. Let

$$P_1 = z^2 f_2 + z^3 f_3 + \cdots + z^{n-1} f_{n-1} + z^n,$$

where each $f_i \in \mathbb{Z}[a]$, count the bipartitions with both $\sigma(u) = \sigma(v) = 1$. Then

$$P_2 = 1 + z f_{n-1} + z^2 f_{n-2} + \cdots + z^{n-2} f_2$$

counts those with $\sigma(u) = \sigma(v) = -1$.

Moreover, let

$$\begin{aligned} P_3 &= z g_1 + z^2 g_2 + \cdots + z^{n-1} g_{n-1}, \\ P_4 &= z g_{n-1} + z^2 g_{n-2} + \cdots + z^{n-1} g_1 \end{aligned}$$

count the bipartitions with $\sigma(u) = -1, \sigma(v) = 1$ and $\sigma(u) = 1, \sigma(v) = -1$ respectively, with $g_i \in \mathbb{Z}[a]$.

Let $A(u, v) = 1$ if u and v are adjacent, 0 otherwise. It is easy to check that $g_1 = a^{\deg(v)}, g_{n-1} = a^{\deg(u)}$ and $f_2 = a^{\deg(u)+\deg(v)-2A(u,v)}$. We now show that we can compute g_1, g_{n-1} and f_2, so that we can compute $A(u, v)$.

Consider adding a leaf to H at u. By Lemma 3.10, $P_1 + P_4$ is determined uniquely. Similarly $P_1 + P_3$ is also determined. In particular, we can deduce $g_1, g_{n-1}, f_2 + g_{n-2}$ and $f_{n-2} + g_2$.

Next consider adding k independent u-v paths on two edges to H to form G. Then

$$P(G, a, z) = (a^2 + z)^k P_1 + (a^2 z + 1)^k P_2 + (a + az)^k (P_3 + P_4).$$

Note that

$$[z^2]P(G, a, z) = a^{2k} f_2 + f_{n-2} + ka^2 f_{n-1} + \binom{k}{2} a^4 + a^k (g_2 + g_{n-2} + k(g_1 + g_{n-1})).$$

Running k over $1, 2$ and 3, we have the following system of linear equations:

$$\begin{pmatrix} a^6 - a^3 & 1 - a^3 & 3a^2 \\ a^4 - a^2 & 1 - a^2 & 2a^2 \\ a^2 - a & 1 - a & a^2 \end{pmatrix} \begin{pmatrix} f_2 \\ f_{n-2} \\ f_{n-1} \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix},$$

where each q_i can be computed from known information.

As the determinant is not the zero polynomial, $f_2, f_{n-2}, f_{n-1} \in \mathbb{Z}[a] \subset \mathbb{Q}(a)$ are uniquely determined. Thus, $A(u, v)$ is uniquely determined and hence so is H. □

Note that the above is independent from Theorem 3.4 since the independent set polynomial is generally not contained in the Ising polynomial.

Chapter 4

Tree reconstruction with polynomials

We divide some polynomials into three levels according to their tree distinguishing power. Motivated by Schwenk's result [88] that asymptotically almost every tree shares the characteristic polynomial with another tree, we define level one to comprise of polynomials that fail to distinguish rooted trees. After giving some examples, we show that the typical size of equivalence classes defined by a level one polynomial grows exponentially fast.

Level two consists of polynomials whose rooted versions distinguish rooted trees; the first member is the greedoid Tutte polynomial [35]. By proving that the Ising polynomial belongs to this level, we strengthen a result of Bollobás and Riordan [13]. Moreover, we show that the Negami polynomial determines the greedoid Tutte polynomial improving a result of Martin, Morin and Wagner [71].

Level three consists of polynomials which encode information about all partitions of the vertex set. Essentially there is only one member: the chromatic symmetric function introduced by Stanley [92]. It remains a conjecture that this is a complete invariant for trees. We will show that caterpillars are distinguished by the chromatic symmetric function.

A summary of the polynomials is given in Figure 4.1.

Section 4.3 is joint work with Alex Scott.

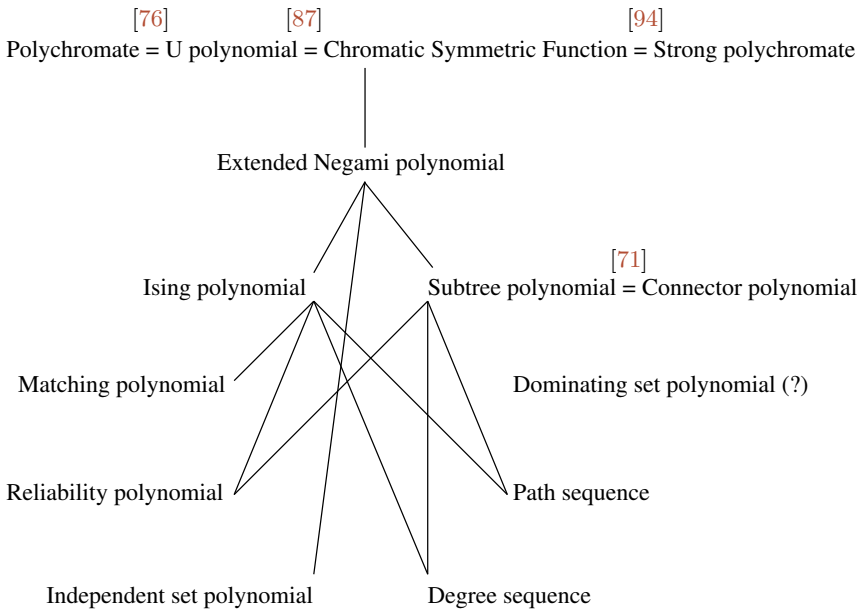


Figure 4.1: Hierarchy of some polynomials in tree distinguishing power: a line indicates deducibility, absence of a line indicates incomparability. The dominating set polynomial is incomparable with every other except possibly the polychromate

4.1 Level one polynomials

Let T be a tree rooted at t. Suppose $T_1, T_2, \cdots, T_k$ are the rooted subtrees of $T - t$, then we write $T = \oplus_{i=1}^{k} T_i$. A polynomial p is of level one if there exist rooted trees L_1, L_2 such that if a tree T_1 has L_1 as a limb and T_2 is obtained from T_1 by swapping L_1 with L_2 then $p(T_1) = p(T_2)$. We give examples of level one polynomial in the following. Even though the polynomials can be defined for general graphs, we restrict our attention to trees.

4.1.1 Independent set polynomial

Let $\mathcal{I}(T)$ be the set of independent sets. The independent set polynomial is $I(T, z) = \sum_{I \in \mathcal{I}(T)} z^{|I|}$. Let $I_0(T, z) = I(T - t, z)$. A recursion is

$$I(T, z) = \prod_{i=1}^{k} I(T_i, z) + z \prod_{i=1}^{k} I_0(T_i, z).$$

We found by computer that the smallest pair of rooted trees with equal I and I_0 has order 8.

4.1.2 Matching polynomial

Let $\mathcal{M}(T)$ be the set of sets of independent edges. The matching polynomial is $M(T,z) = \sum_{M\in\mathcal{M}(T)} z^{|M|}$. Let $M_0(T,z) = M(T-t,z)$. A recursion is

$$M(T,z) = \prod_{i=1}^{k} M(T_i,z) + \sum_{i=1}^{k} M_0(T_i,z) \prod_{j\neq i} M(T_j,z).$$

The smallest pair of rooted trees with equal M and M_0 has order 9, given by Schwenk [88].

4.1.3 (F,r)-matching polynomials

Recall that the (F,r)-matching polynomial $S_r(F,\cdot,z)$ enumerates (F,r)-matchings with weight the number of copies in the matching. Thus $I(\cdot,z) = S_2(K_1,\cdot,z)$ and $M(\cdot,z) = S_1(K_2,\cdot,z)$. As I and M are of level one, it seems likely that so is $S_r(F,T,z)$.

Conjecture 4.1. *For any integer $r \geq 1$ and tree F, there exist nonisomorphic rooted trees R_1, R_2 such that if a tree T_1 has R_1 as a limb, and that T_2 is obtained from T_1 by swapping R_1 with R_2, then $S_r(F,T_1,z) = S_r(F,T_2,z)$.*

4.1.4 Reliability polynomial

The *reliability polynomial* of a rooted tree computes the expected size of components containing the root if each edge is allowed to fail independently with probability p. Aivaliotis, Gordon and Graveman [5] considered an unrooted version of the polynomial. More precisely, for an unrooted tree T with n edges, let $\mathcal{F}$ be the collection of complement of subtrees. For $S \subset E$, define the rank $r(S) = \max_{A\subset S}\{|A| : A \in \mathcal{F}\}$. Let

$$R(T,p) = \sum_{S\subset E} r(S) p^{r(S)} (1-p)^{|E|-r(S)}$$

be the corresponding rank-generating polynomial. It was proved in [5] that

$$R(T,p) = -np^n + \sum_{e \in E(T)} (p^{|A_e|+1} + p^{|B_e|+1})$$

where A_e, B_e are the components of $T - e$. In other words, $R(T,p)$ encodes the order of the classes defined by cuts of size one. Therefore, $R(T,p)$ is equivalent to $\rho(T,z) := [a]P(T,a,z)$. Write $\rho^+(T,z)$ for $[a]P^+(T,a,z)$. Then $P^+(T,a,z) = z^{|T|} + a\rho^+(T,z) + g(a,z)$, for some $g(a,z)$ such that $a^2|g(a,z)$. Thus for $T = \oplus_{i=1}^k T_i$,

$$\rho^+(T,z) = [a]\frac{\prod_{i=1}^k P^+(T_i,a,z)}{z^{k-1}} = z^{|T|}\sum_{i=1}^k \frac{\rho^+(T_i,z)}{z^{|T_i|}}.$$

Therefore, if we switch rooted trees T_1 with T_1' with equal ρ^+ then $R(T,p)$ remains unchanged. Note also that since $[a]P^-(T,a,z) = [a]z^n P^+(T,a,1/z)$, equality in ρ^+ is equivalent to that in ρ. The smallest pair of rooted trees with equal ρ has order 6, see Figure 4.2.

4.1.5 Path sequences

Denote the set of paths on at least one vertex in T by $\mathcal{P}(T)$. Let the *path sequence* be $\mathrm{PS}(T,z) = \sum_{P \in \mathcal{P}(T)} z^{|P|}$. Thus $\mathrm{PS}(P_1,z) = z$ and $\mathrm{PS}(P_2,z) = 2z + z^2$. Define the rooted version as

$$\mathrm{PS}_0(T,z) = \sum_{k>0} |\{P \subset V(T) : P \text{ is a path on } k \text{ vertices with } t \text{ as an end}\}| z^k.$$

Let $\mathrm{PS}_r(T,z) = \mathrm{PS}(T,z) - \mathrm{PS}_0(T,z)$ count the rest of the paths.

For $T = \oplus_{i=1}^k T_i$, a recursion for PS is given by

$$\begin{aligned}
\mathrm{PS}_0(T,z) &= z + z\sum_{i=1}^k \mathrm{PS}_0(T_i,z), \\
\mathrm{PS}_r(T,z) &= \sum_{i=1}^k \mathrm{PS}(T_i,z) + z\sum_{1 \le i < j \le k} \mathrm{PS}_0(T_i,z)\mathrm{PS}_0(T_j,z).
\end{aligned}$$

Indeed, any path starting at t is either t itself or continued by a path in T_i starting at t_i for some i. Any other path does not contain t or has t as an internal vertex in which case the two flanks of t are paths starting at t_i, t_j for some $i \neq j$.

We found three pairs of rooted trees of order 10 with equal PS and PS_0 with computer, see Figure 4.2 for a pair.

Since the only type of subgraphs in a tree with two odd vertices is precisely the paths, by 5 of Theorem 3.7, we can deduce $\mathrm{PS}(T,z)$ from $Z(G,x,y)$.

4.1.6 Dominating set polynomial

Recall that $A \subset V(T)$ is a *dominating set* if $V(T) = A \cup \Gamma(A)$. Let $\mathcal{D}(T)$ be the collection of dominating sets. The *dominating set* (or *domination*) *polynomial* is $D(T,z) = \sum_{D\in\mathcal{D}(T)} z^{|D|}$, see [1]. The rooted version $D_0(T,z)$ counts only dominating sets containing t. Let $[[X]] = 1$ if X is true, zero otherwise. For $T = \oplus_{i=1}^{k} T_i$, we have the recursion:

$$\begin{aligned}
D_0(T,z) &= z\prod_{i=1}^{k}(\ (1+z)(D(T_i - t_i, z) + z[[\ |T_i| = 2\ ]]\ ).\\
D(T-t,z) &= \prod_{i=1}^{k} D(T_i,z).\\
(D-D_0)(T,z) &= D(T-t,z) - \prod_{i=1}^{k}(D-D_0)(T_i,z).\\
D(T,z) &= D_0(T,z) + (D-D_0)(T,z).
\end{aligned}$$

Indeed, a dominating set not using the root consists of a dominating set in $T-t$ and contains at least one root t_i, hence the third equation.

The smallest pair of rooted trees with identical domination data has order 7, see Figure 4.2.

It would be interesting to know if a finite collection of level one polynomial remains 'level one'. That is to say, is there a pair of rooted trees swapping of which as limbs produces p-equivalent trees for every p in the collection. For example, there are two rooted trees on 16 vertices with identical I, I_0, M, M_0, see Figure 4.2. On the other hand, if we allow an infinite collection, then the answer is negative. Indeed, consider all the $(F,1)$-matching polynomials. Suppose R_1, R_2 are two rooted trees such that $S_1(F,R_1) = S_1(F,R_2)$ for all trees F. As $S_1(K_1,R_1,z) = S_1(K_1,R_2,z)$, it

follows that $|R_1| = |R_2|$. Attach $|R_1|$ leaves to the root of each R_i to form F_i. Since $S_1(F_1, F_1, z) = S_1(F_1, F_2, z)$, F_1 is isomorphic to F_2 as unrooted trees. Since any isomorphism maps the root of F_1 to that of F_2, it induces an isomorphism between R_1 and R_2 as rooted trees.

4.2 Switching

Let $\mathcal{R}^U, \mathcal{T}^U$ be the classes of unlabelled rooted trees and unrooted trees respectively. Let $R^U(z), T^U(z)$ be the generating function of $\mathcal{R}^U, \mathcal{T}^U$ respectively. By nicely relating the number of edge-rootings and vertex-rootings, Otter [81] proved that $\mathrm{rad}(R^U) = \mathrm{rad}(T^U)$ and computed it to be approximately 0.338219, refining the results of Pólya [84] who first studied $R^U(z)$ and showed that $0 < \mathrm{rad}(R^U) < 1$.

Schwenk [88] proved that for a.a.e. tree, there is another tree with the same characteristic polynomial. He proved the following theorem and used a specific pair of limbs (A, B) such that switching A to B does not change the characteristic polynomial. For a rooted tree T, a *limb* A at v consists of one or more branches of v not containing the parent of v and that $T \neq A$.

Theorem 4.2 ([88]). *Let A be a rooted tree. The OGF $T_A(z)$ of rooted trees without A as a limb satisfies $T_A(z) = z(1 - z^{|A|}) \exp\left(\sum_{k\geq 1} \frac{T_A(z^k)}{k}\right)$. Moreover, the* $\mathrm{rad}(T_A) > \mathrm{rad}(R^U)$.

Furthermore, Schwenk considered C a coalescence of k copies of A. As a.a.e. tree contains C as a limb, by switching copies of A to B one by one, he concluded that there exists arbitrary large family of cospectral trees. However, the order of trees in such a family may be large compared with k. It is our aim in this section to refine this result by showing that there are exponentially large families of p-equivalent trees for a level one polynomial p.

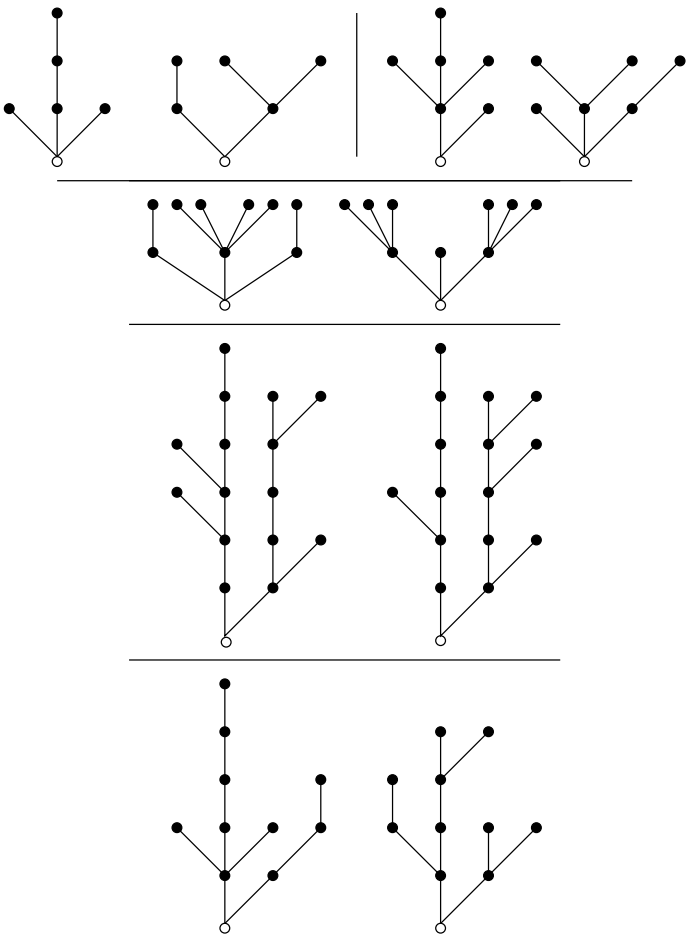


Figure 4.2: Pairs of trees (L, R) with the same level one polynomial of both rooted and unrooted versions with the outlined circles as the root. Top to bottom, left to right: $\rho^+(L) = \rho^+(R)$; $D(L) = D(R), D_0(L) = D_0(R), D(L-l) = D(R-r)$; $\mathrm{PS}(L) = \mathrm{PS}(R), \mathrm{PS}_0(L) = \mathrm{PS}_0(R)$;$M(L) = M(R), M(L-l) = M(R-r), I(L) = I(R), I(L-l) = I(R-r)$; $S_3(K_1, L) = S_3(K_1, R), S^i(L) = S^i(R), i = 1, 2, 3$

4.2.1 Unlabelled trees

Let p be a level one polynomial. By definition, there exist rooted trees A, B such that replacing a copy of a limb A by that of B produces an p-equivalent tree. Furthermore, by attaching an edge to the roots of A or B if necessary, we may assume that the roots have degree 1.

We refine p-equivalence using A and B. An (A, B)-*switch* replaces a copy of a limb A by a copy of B or vice versa. Two rooted trees S, T are called (A, B)-*equivalent* if S can be obtained from T by a sequence of (A, B)-switches. Two unrooted trees are (A, B)-*equivalent* if they can be rooted to be (A, B)-equivalent rooted trees. Then (A, B)-equivalence refines p-equivalence. We aim to show that the number of (A, B)-equivalence classes of unrooted trees is small compared to the total number of trees. We consider a further refinement.

We call a vertex of T to which exactly m copies of A or B are attached as limb an m-*attachment vertex*; we repeat that we do not count A, B as a limb if $T = A$ or B. By labelling the attachment vertices, we circumvent the difficulty of counting similar vertices in T, i.e. vertices in the same orbit under $\mathrm{Aut}(T)$. To make it clear, by changing the number of copies at a k-attachment vertex and a j-attachment vertex, we obtain a total of $(k+1)(j+1)$ distinct equivalent trees, regardless whether the vertices are similar or not. Note that if two rooted unlabelled (more precisely, partially labelled at the attachment vertices) trees are equivalent, then the unrooted versions are (A, B)-equivalent.

We use generating functions, about which we refer to Flajolet and Sedgewick [32]. Denote the multiset construction by MSet. Singling out A and B, we have the specification

$$\mathcal{R}^U = \mathcal{Z} \times \mathrm{MSet}(A) \times \mathrm{MSet}(B) \times \mathrm{MSet}(\mathcal{R}^U - A - B).$$

We count the occurrence of A or B by the number of copies with $u_1, u_2, \cdots$. In other words, a tree counted with u_m^k has k m-attachment vertices. Denote $(u_1, u_2, \cdots)$

by $\mathbf{u}$ and $(u_1^k, u_2^k, \cdots)$ by $\mathbf{u}^k$. With $n = |A| = |B|$, we have

$$R^U(z, \mathbf{u}) = z(1 + 2z^n u_1 + 3z^{2n} u_2 + 4z^{3n} u_3 + \cdots) \exp\left(\sum_{k\geq 1} \frac{(R^U - 2z^n)(z^k, \mathbf{u}^k)}{k}\right).$$

Indeed, the term $(m+1)z^{mn}u_m$ counts the $(m+1)$ ways of choosing m limbs from A or B; and since we do not count the occurrence of A or B in the exponential term, we have the correction $-2z^n$.

As switching at each m-attachment vertex gives rise to $m+1$ equivalent trees, if we put $u_m = \frac{1}{m+1}$, we obtain the OGF $E(z)$ of the number of equivalence classes,

$$\begin{aligned} E(z) &= z(1 + z^n + z^{2n} + z^{3n} + \cdots) \exp\left(\sum_{k\geq 1} \frac{E(z^k)}{k}\right) \exp\left(-\sum_{k\geq 1} \frac{2z^{kn}}{k}\right) \\ &= z(1 - z^n) \exp\left(\sum_{k\geq 1} \frac{E(z^k)}{k}\right), \end{aligned}$$

since $-\log(1-z) = \sum_{k\geq 1} \frac{z^k}{k}$. Note that E is the same as the generating function of trees forbidden a limb of order n in Theorem 4.2. I would like to thank Colin McDiarmid for pointing out the reason is that forbidding A is equivalent to having all copies of B.

Theorem 4.3. *Let A, B be rooted trees of the same order with root degree 1. Then there exists constant $C > 1$ such that a.a.e. tree of sufficiently large order n lies in an (A, B)-equivalence class of size at least C^n.*

Proof. Let E be the OGF of the number of equivalence classes and $t_n = [z^n]T^U(z)$. By Theorem 4.2, $\mathrm{rad}(E) > \mathrm{rad}(R^U)$. Since $\mathrm{rad}(R^U) = \mathrm{rad}(T^U)$, there exists $c > 1$ such that $\frac{[z^n]T^U(z)}{[z^n]E(z)} > c^n$. Let $d < 1$ such that $1 < dc$. Then the number of trees in equivalence classes of size at most $(dc)^n$ is at most $(dc)^n \frac{t_n}{c^n} < d^n t_n$. This proportion tends to 0 as $n \to \infty$. □

We have now proved the main result of the section.

Theorem 4.4. *Let p be a level one polynomial. There exists a constant $C > 1$ such that a.a.e. tree of sufficiently large order n lies in an p-equivalence class of size at least C^n.*

4.2.2 Labelled trees

Labelled trees exhibit similar behaviour and are easier to deal with. An analogous study is included here. The equivalence relations are defined as in the unlabelled case. Let A, B be two rooted trees of order $n > 2$. We count the occurrence of A or B as limbs by a single counter u, that is we have the specification for rooted labelled trees

$$\mathcal{R}^L = \mathcal{Z} \star \mathrm{Set}(uA) \star \mathrm{Set}(uB) \star \mathrm{Set}(\mathcal{R}^L - A - B).$$

Let α, β be the order of the automorphism groups of A and B as rooted trees respectively. For example, if $A = P_4$ rooted at an end and $B = K_{1,3}$ rooted at a leaf, then $\alpha = 1, \beta = 2$. A translation into EGF is

$$\begin{aligned} R^L(z,u) &= z \exp\left(\frac{uz^n}{\alpha} + \frac{uz^n}{\beta}\right) \exp\left(R^L(z,u) - \frac{z^n}{\alpha} - \frac{z^n}{\beta}\right) \\ &= z \exp\left(z^n(u-1)\left(\frac{1}{\alpha} + \frac{1}{\beta}\right)\right) \exp(R^L(z,u)). \end{aligned}$$

Each occurrence of A can be replaced by $n!/\alpha$ copies of A or $n!/\beta$ copies of B to give an (A, B)-equivalent tree. Therefore, if we put $u = \frac{1}{n!/\alpha + n!/\beta}$, we obtain the EGF $E(z)$ of the number of equivalence classes which satisfies

$$E(z) = z \exp\left(\left(\frac{1}{n!} - \frac{1}{\alpha} - \frac{1}{\beta}\right) z^n + E(z)\right).$$

Let $f(z) = z \exp(-\gamma z^n)$ where $\gamma = -\frac{1}{n!} + \frac{1}{\alpha} + \frac{1}{\beta}$ and satisfies $0 < \gamma < 2$.

Let $R(z) = R^L(z, 1)$. The well-known Cayley formula tells us that $[z^k]R(z) = k^{k-1}$ so that $R(z) = \sum_{k \geq 1} \frac{k^{k-1}}{k!} z^k$. It follows that $\mathrm{rad}(R) = 1/e$.

We next check that $\mathrm{rad}(E) > 1/e$. Indeed, since $R \circ f$ is a solution to $y(z) = f(z) \exp(y(z))$ and has a positive radius of convergence, it is the analytic solution around $z = 0$. Hence,

$$E(z) = R \circ f(z) = \sum_{n \geq 1} \frac{n^{n-1}}{n!} f(z)^n$$

converges as long as $z \geq 0$ and $f(z)$ has not reached $\mathrm{rad}(R) = 1/e$. In other words, $\mathrm{rad}(E) = \sup_{z>0}\{f(x) < 1/e, \forall 0 \leq x < z\}$. Moreover, note that $\max_{z>0} f(z) > 1/e$ for all $m > 2$ and $0 < \gamma < 2$ so that $\mathrm{rad}(E) < \infty$. Thus, we have $1/e < \mathrm{rad}(E) < \infty$.

By an argument similar to the unlabelled case, Theorem 4.3, we have the following.

Theorem 4.5. *Let A, B be rooted trees of the same order. There exists $C > 1$ such that a.a.e. tree on n vertices lies in an (A, B)-equivalence class of size at least C^n.*

We remark that the theorem can also be proved by considering the distribution of the number of copies of A as a limb in a random tree. It can be shown that there is a.a.s. a linear number of copies, so that switching gives rise to exponentially large families of (A, B)-equivalent trees.

4.3 Level two polynomials

There are three level two polynomials, namely, the greedoid Tutte polynomial (or equivalently the subtree polynomial), the Ising polynomial and the Negami polynomial. The main results in this section are that the rooted Ising polynomial distinguishes rooted trees and that the Negami polynomial determines the subtree polynomial.

The usual Tutte polynomial of a tree depends only on its order. Using a branching greedoid rank instead, Gordon and McMahon [35] defined a rooted Tutte polynomial and proved that it distinguishes rooted trees. Later, Chaudhary and Gordon [22] proved that the polynomial can be equivalently defined using a pruning rank. More precisely, a rooted subtree of a rooted tree (T, t) has root t. For $A \subset E = E(T)$, let $R(A) = \max\{|F| : E - F \text{ is a rooted subtree and } F \subset A\}$. The *greedoid Tutte polynomial* is defined as

$$f_0(T, x, y) = \sum_{A \subset E} x^{R(E)-R(A)} y^{|A|-R(A)}.$$

For an unrooted tree T and any $A \subset E$, let $r(A) = \max\{|F| : E-F \text{ is a subtree, } F \subset A\}$, i.e. the size of the largest subtree complement in A. The unrooted version of f_0 is defined by

$$f(T, x, y) = \sum_{A \subset E} x^{r(E)-r(A)} y^{|A|-r(A)}.$$

Moreover, it was proved in [22] that f_0, f count the subtrees by the order and the number of leaves. For a rooted subtree (S, t) of (T, t), let $l(S)$ be the number of edges

incident with a vertex without descendants. For a subtree S of T, let $l(S)$ be the number of leaves. Let $g(T) = \sum_{r \in T} f_0(T, r)$.

Lemma 4.6 ([22]). *Let T be a tree of order $n+1$. For any $r \in T$,*

$$f(T, x, y) = \sum_{S\text{: subtree}} x^{|S|}(y+1)^{|S|-l(S)}$$

and

$$f_0((T, r), x, y) = \sum_{S\text{: rooted subtree}} x^{|S|}(y+1)^{|S|-l(S)}.$$

Moreover, $g(T) = f(T) + x(y+1)\dfrac{\partial f}{\partial x} - (y^2+y)\dfrac{\partial f}{\partial y} + n(1-xy)$.

Therefore, the unrooted greedoid Tutte polynomial is equivalent to the *subtree polynomial* defined as

$$S(T, x, y) = \sum_{S\text{: subtree}} x^{|S|}y^{l(S)}.$$

Negami and Ota [75] defined the extended Negami polynomial for general graphs. For our purposes, the *Negami polynomial* for trees is

$$N(T, x, y, z, w) = \sum_{(A,B)\text{ partitions }T} x^{|A|}y^{e(A)}z^{e(B)}w^{e(A,B)}.$$

Or equivalently, $N(T, x, y, z, w) = \hom(T, K_2^\circ)$ with $w_1 = x, w_2 = 1, w_{11} = y, w_{22} = z, w_{1,2} = w$. As $e(A, B) = e(G) - e(A) - e(B)$, we may omit one of y, z or w.

The Negami polynomial determines the Ising polynomial, to wit, $P(T, a, z) = N(T, a, 1, 1, z)$. In fact, we will prove that the Negami polynomial also determines the subtree polynomial in Section 4.3.2. As the chromatic symmetric function is a stronger invariant for trees than the Negami polynomial, this extends a result in Martin, Morin and Wagner [71].

We start by showing that the rooted Ising polynomial distinguishes rooted trees in Section 4.3.1.

We remark that a rooted polynomial defined by Contia, Contuccic and Falcolinia [24] also distinguishes rooted trees. But there is no natural unrooted version so we do not consider it here.

4.3.1 Ising polynomial and trees

Andrén and Markström [4] computed the number of isomagnetic trees, i.e. trees that are not magnetically unique, of order up to 18 and noted that the proportion does not seem to increase. Moreover, for each tree which is not magnetically unique, there is exactly one other isomagnetic tree. We extend the computation several orders onwards and find that the same phenomena persist. The data lead us to suspect that the Ising polynomial distinguishes most trees and the following question.

Question 4.7. *Is there a triple of mutually isomagnetic trees?*

We remark that there are no known triples of trees with the same greedoid Tutte polynomial either. It is conceivable that if the rooted version of a polynomial distinguishes rooted trees, the unrooted version defines equivalence classes of finite size.

n	$\lvert\mathcal{T}_n^U\rvert$	No. of isomagnetic trees	No. of Negami-equiv. trees
11	235	2	2
12	551	4	2
13	1301	2	2
14	3159	14	10
15	7741	8	2
16	19320	36	14
17	48629	52	38
18	123867	92	30
19	317955	106	46
20	823065	260	112
21	2144505	156	44
22	5623756	570	180
23	14828074	582	320
24	39299897	1024	244
25	104636890	954	336

Recall that one way to construct isomagnetic graphs uses strongly isomagnetic graphs. Emboldened by our success with the application of Schwenk's theorem in the case

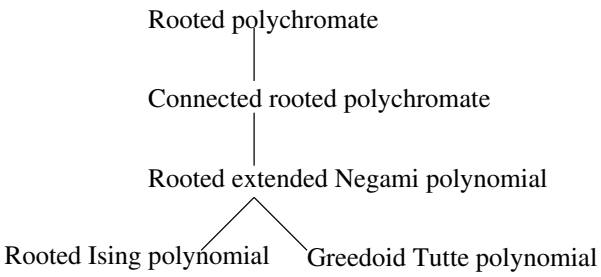


Figure 4.3: Hierarchy of some graph polynomials of rooted trees

of the independent set polynomial, we looked for strongly isomagnetic trees. After a long search, it dawned on us that such a pair does not exist.

Theorem 4.8. *The rooted Ising polynomial distinguishes rooted trees.*

The greedoid Tutte polynomial defined by Gordon and McMahon [35] was the first known polynomial invariant that distinguishes rooted trees. Later, Bollobás and Riordan [13] considered a rooted version of the polychromate (a generalization of the chromatic polynomial introduced by Brylawski [16]). They found that by restricting to the connected rooted polychromate, there is a nice multiplicative formula from which they showed that the rooted polychromate distinguishes rooted trees. As we shall see in Section 4.3.2, the rooted polychromate determines the rooted Negami polynomial which in turn determines the rooted greedoid Tutte polynomial. Here, we concern ourselves with the task of showing that the rooted Ising polynomial distinguishes rooted trees.

We remark that while both the Ising polynomial and the greedoid Tutte polynomial are determined by the polychromate, they are incomparable in terms of tree distinguishing power.

Recall that for a rooted tree (T, t), the rooted Ising polynomial is

$$P^+(T, z, a) = \sum_{\substack{(A,B) \text{ partitions } T \\ \text{with } t \in A}} z^{|A|} a^{e(A,B)}.$$

To prove Theorem 4.8, we first rephrase Eisenstein's irreducibility criterion for which we give a proof for completeness.

Lemma 4.9. *Let* $F(z,a) = a + f_1(a)z + f_2(a)z^2 + \cdots f_{n-1}(a)z^{n-1} + z^n$ *such that* $a|f_i(a)$ *for all* $1 \le i \le n-1$. *Then* F *is irreducible over* $\mathbb{Z}[a][z] = \mathbb{Z}[a,z]$.

Proof. Suppose F is reducible. Then $F = BC$ where

$$\begin{aligned} B(z,a) &= (\pm 1 + b_1(a)z + b_2(a)z^2 + \cdots b_{k-1}(a)z^{k-1} + z^k), \\ C(z,a) &= (\pm a + c_1(a)z + c_2(a)z^2 + \cdots c_{l-1}(a)z^{l-1} + z^l), \end{aligned}$$

and $k > 0$. For simplicity, we assume both signs in the constant z term are positive. Since $f_1 = c_1 + ab_1$ is divisible by a, so is c_1. Assume that $c_1, c_2, \cdots, c_{i-1}$ are divisible by a. For each $i < n$, as $f_i = c_i + b_1c_{i-1} + b_2c_{i-2} \cdots + b_ia$ (where the coefficients are 0 as appropriate), c_i is divisible by a. By induction, this holds for all coefficients in C which contradicts $c_l = 1$. □

Proof of Theorem 4.8. We prove by induction which clearly starts.

Let T be a rooted tree defined by identifying the roots of $T_1, \cdots, T_k$ where the root degree of T_i is 1 for all i. Let $P_i = P^+(T_i, z, a)$. Then $P^+(T, z, a) = z\prod_{i=1}^{k}(P_i/z)$.

Suppose $P^+(T,z,a) = P^+(T',z,a)$. Then we have $\prod_{i=1}^{k}(P_i/z) = \prod_{i=1}^{k'}(P_i'/z)$. It suffices to show that each P_i/z is irreducible. For then by unique factorization in $\mathbb{Z}[a,z]$, we conclude that $k = k'$ and (by relabelling if necessary) $P_i = P_i'$ for all $i = 1, 2, \cdots, k$.

We show that each P_i/z satisfies the conditions in Lemma 4.9. As the degree of the root of T_i is 1, $[z^1]P_i = a$. Moreover, $[z^j]P_i$ is divisible by a for all $1 < j < |T_i|$ as T_i is connected. By definition, $[z^{|T_i|}]P_i = 1$. Thus, by Lemma 4.9, P_i/z is irreducible.

By induction, each P_i determines a unique rooted tree and so T is uniquely determined. □

4.3.2 Negami and subtree polynomials

We now use the second part of Lemma 4.6 to prove that for trees, N determines g. We follow a slightly tortuous path by first looking at the rooted case. Define the *rooted Negami polynomial*, which counts partitions where the root is always in A, as

$$N_0(T,x,y,w) = \sum_{(A,B) \text{ partitions } T \text{ with } t \in A} x^{|A|}y^{e(A)}w^{e(A,B)}.$$

We shall show that N_0 determines g_0. Then we recover the unrooted version from the rooted one. Indeed, this is possible when we do not consider any particular root but the sum of all of them.

Theorem 4.10. *The Negami polynomial of a tree determines its greedoid Tutte polynomial.*

Proof. The flow of the argument is

$$N \xrightarrow{1.} \sum_r N_0(T,r) \xrightarrow{2.} \sum_r f_0(T,r) \xrightarrow{3.} f(T).$$

1. From N, we can deduce the sum of the rooted versions. Indeed, as the root is always in A, we have

$$\sum_{r\in T} N_0(T,r) \;=\; \sum_{(A,B) \text{ partitions } T} |A| x^{|A|} y^{e(A)} w^{e(A,B)} = x\frac{\partial N}{\partial x}.$$

2. Next, note that $f_0(T,r)$ is determined by $N_0(T,r)$. Indeed, a monomial of $N_0(T,r)$ in which the power of y is one less than that of x corresponds to a rooted subtree S at r and the power of w is precisely the order of the vertex-boundary $\Gamma(S)\setminus S$. Observe that $\Gamma(S)\setminus S$ is the set of leaves of $S\cup\Gamma(S)$. Therefore by Lemma 4.6, $N_0(T,r)$ determines $f_0(T,r)$. Since coefficient extraction is linear, $\sum_r f_0(T,r)$ is determined by $\sum_r N_0(T,r)$.

3. Thus, it suffices to show that f can be computed from $g = \sum_r f_0(T,r)$. Let $g_{ij} = [x^i y^j]g$. By Lemma 4.6,

$$g_{ij} = f_{ij}(1+i+j) + (i-j+1)f_{i,j-1} + n[x^i y^j](1-xy).$$

   Then we can recursively compute f_{ij} given the boundary conditions. So it remains to show that they are also determined by N. Firstly, $f_{00} = 1$ and $f_{0j} = 0$ for $j > 0$. Moreover, note that

$$\begin{aligned} f(T,x,0) &= \sum_{A\subset E, |A|=r(A)} x^{|E|-|A|} \\ &= \sum_{S\text{: subtree}} x^{|S|}, \end{aligned}$$

since $|A| = r(A)$ if and only if A^c is a subtree. By considering bipartitions induced by a subtree in one class, $f(T, x, 0)$ can be read off from N. Thus, f_{i0} is also determined by N. Hence, f is uniquely determined by g.

□

4.4 Level three polynomials

Level three polynomials encode information about all partitions of a graph and include the following. Let $K^\circ_{\mathbb{N}}$ be the complete graph on $\mathbb{N}$ with a loop at each vertex.

Let $\{x_1, x_2, x_3, \cdots\}$ be distinct commuting indeterminates. Stanley [92] generalized the chromatic polynomial to the *chromatic symmetric function* as

$$X(G) = \sum_{\kappa} \prod_{v \in G} x_{\kappa(v)}$$

where the sum is over proper colourings $\kappa : V(G) \to \mathbb{N}$. Even though the chromatic polynomial is not a homomorphism count except at nonnegative integers (see [14]), its 'symmetrization' can be expressed as $X(G) = \hom(G, K^\circ_{\mathbb{N}})$ with $w_i = x_i$ and $w_{ii} = 0$ and $w_{ij} = 1$ for all $j \neq i$. Although $X(G)$ is not a polynomial in the usual sense, for a graph of order n, we may restrict the codomain of κ to $[n]$, so that the abuse of 'polynomial' can be justified. Note that $X(G)$ is not a complete invariant for graphs (there are graphs on 5 vertices having the same $X(G)$). However, it is conjectured to distinguish all trees.

Conjecture 4.11 (Stanley [92])**.** *The chromatic symmetric function is a complete invariant for trees.*

The *Tutte symmetric function* can be defined as $\hom(G, K^\circ_{\mathbb{N}})$ with $w_{ii} = (1 + t)$ and $w_i \neq w_j, w_{ij} = 1$ for $i \neq j$. The essential information captured is for each vertex colouring, the order of each colour class and the total number of monochromatic edges. This is exactly what the *polychromate* defined by Brylawski [16] encodes. In other words, the Tutte symmetric function is a symmetrization of the polychromate.

The *U-polynomial* introduced by Noble and Welsh [76] is in fact equivalent to the polychromate as shown by Sarmineto [87].

The *strong polychromate* introduced by Bollobás and Riordan [13] encodes, for each vertex colouring, not only the order, but the size as well, of each colour class. In other words, it symmetrizes into $\hom(\cdot, K^{\circ}_{\mathbb{N}})$ with distinct w_i, w_{ii} for each i and $w_{ij} = 1$ for $i \neq j$.

Thus, the chromatic symmetric function is contained in the Tutte symmetric function which is equivalent to the polychromate which in turn is contained in the strong polychromate. However, they are equivalent for trees.

Theorem 4.12 (Thatte [94])**.** *For trees, the strong polychromate can be deduced from the polychromate.*

Hence, Stanley's conjecture that $X(T)$ defines uniquely the tree T is equivalent to versions where X is replaced by any other level three polynomials.

Computationally, we verify Stanley's conjecture for trees up to 25 vertices. Let the *tripartite polynomial* be $\hom(\cdot, K^{\circ}_3)$ with $w_i \neq w_j, w_{ii} \neq w_{jj}, w_{ij} = 1$ for $i \neq j$. By evaluating the three-part polynomial at a certain point for each pair of Negami-equivalent trees, we found that each pair can be differentiated.

Some families of trees have been shown to be determined by the level three polynomial. In particular, as the polychromate determines the matching polynomial of a tree, and no two spiders (subdivided stars) are cospectral [63], spiders are determined by the polychromate. In fact, Martin, Morin and Wagner [71] proved this directly from the chromatic symmetric function. Moreover, they demonstrated that certain caterpillars are reconstructible. Indeed, let $v_1 v_2 \cdots v_{d-1}$ be a path and $(l_1, l_2, \cdots, l_{d-1})$ be a sequence of nonnegative integers such that $l_1, l_{d-1} > 0$. Define a caterpillar $C = C(l_1, l_2, \cdots, l_{d-1})$ be identifying the centre of K_{1,l_i} with v_i for each i. It was proved in [71] that among caterpillars with positive and distinct $l_1, \cdots, l_{d-1}$, the chromatic symmetric function is a complete invariant. In fact this can be extended to all caterpillars.

Theorem 4.13. *The chromatic symmetric function distinguishes caterpillars.*

Proof. It was proved in [71] whether C is a caterpillar or not can be recognized from $X(C)$. Suppose $C = C(l_1, l_2, \cdots, l_{d-1})$ is a caterpillar, our task is to reconstruct the sequence $(l_1, l_2, \cdots, l_{d-1})$ from $X(C)$. (In the following, it may be helpful to think in terms of the strong polychromate.) There is nothing to prove if $d \leq 2$ so we assume $d \geq 3$ and the result has been proved up to $d-1$.

Let $n = |C|$. By considering cuts of size 1, we see that the multiset

$$P = [(|A_e|, |B_e|) : e \in E(C), A_e, B_e \text{ are the components of } C - e, |A_e| \leq n/2]$$

is reconstructible. Write (a_i, b_i) for a pair in P and assume that $a_1 \leq a_2 \leq \cdots \leq a_{n-1}$. Since there are $L = \sum_{i=1}^{d-1} l_i$ leaves in C, $a_L = 1$ and $a_{L+1} > 1$. In fact, (a_{L+1}, b_{L+1}) is a cut corresponding to either the edge v_1v_2 or $v_{d-2}v_{d-1}$. Assume it is v_1v_2. Hence $l_1 = a_{L+1} - 1$.

Since the path sequence is determined by $X(C)$ (see either [71, Corollary 5] or [4, Theorem 4.2]), the number of paths of length d is reconstructible. As this quantity is $l_1 l_{d-1}$ and l_1 is known, we can compute l_{d-1}.

Suppose we have reconstructed the head $H = C(l_1, l_2, \cdots, l_h)$ and the tail $T = C(l_t, l_{t+1}, \cdots, l_{d-1})$ by identifying the edges corresponding to the cuts $(a_i, b_i) \in P$ for $L + 1 \leq i \leq j - 1 < n - 1$. We show that the edge corresponding to (a_j, b_j) can be identified.

Let $H' = C(l_1, l_2, \cdots, l_h, x)$ and $T' = C(y, l_t, l_{t+1}, \cdots, l_{d-1})$ where x, y are non-negative integers such that $|H'| = |T'| = a_j$. Since there are at most two edges giving the cut (a_j, b_j), either H' is an extension of H or T' is an extension of T in C.

If $a_{j+1} = a_j$, both extensions are present.

If $a_{j+1} > a_j$, then there is a unique edge e such that $C - e = A_e \cup B_e$ with $|A_e| = a_j$. We claim that we can recognize $X(C[A_e])$ from $X(C)$. Then either $X(C[A_e]) = X(H')$ or $X(C[A_e]) = X(T')$. If $H' \cong T'$, then we can choose arbitrarily. If $H' \ncong T'$, then by induction $X(H') \neq X(T')$. This identifies the extension, and hence the edge for the cut (a_j, b_j). We proceed onwards to a larger j.

It remains to prove the claim. We need to account for partitions with a component of order b_j which is not B_e. Suppose then $S \neq B_e$ is a subtree of order $b_j \geq n/2$

whose spine is $v_k v_{k+1} \cdots v_m$. Then $k \leq h+1$ and $t-1 \leq m$. Hence S is formed by removing, possibly zero, leaves adjacent to vertices in $\{v_k, v_{k+1}, \cdots, v_m\}$. Since we have reconstructed H and T and the degree sequence is reconstructible, we know the collection $[l_i : k \leq i \leq m]$. Thus we can detect the formation of each S and so determine the isomorphism type of $C - S$. Finally, we can find out the possibilities of k and m by checking if

$$m - k + 1 + \sum_{i=k}^{m} l_i \geq b_j.$$

This accounts for all subtrees of order b_j other than the cut (A_e, B_e) and completes the proof of the claim. □

Chapter 5

Independent set polynomial of trees

This chapter is devoted to tree reconstruction with the independent set polynomial. We will first show that no two spiders have the same independent set polynomial in Section 5.1. On the other hand, we construct in Section 5.2 well-covered trees and non-well-covered trees with the same independent set polynomial to disprove a conjecture of Levit and Mandrescu [65].

There is a polynomial version of Ulam's Reconstruction Conjecture. Instead of vertex-deleted subgraphs, the *polynomial deck* of some graph polynomial F, i.e. the multiset $[F(G-v) : v \in V(G)]$ is given. Is $F(G)$ is reconstructible from the polynomial deck? It remains a conjecture [36] that the characteristic polynomial is reconstructible from the polynomial deck. For the independent set polynomial, [15], since it counts non-spanning subgraphs (if the graph is not trivial), we may apply Kelly's Lemma to conclude that for any $0 \leq k < |G|$, $(|G| - k)i_k(G) = \sum_{v \in V(G)} i_k(G - v)$. We shall study information encoded in the polynomial deck in Section 5.3, where we also give constructions of graphs with the same polynomial deck.

5.1 Spiders

Recall that a spider (or starlike tree) is a subdivided star. As an analogue of the result of Lepović and Gutman [63] that no spiders are cospectral, we show that the independent set polynomial distinguishes among spiders. This extends a result by Levit and Mandrescu [65] who showed that the independent set polynomial distinguishes well-covered spiders among well-covered trees.

Theorem 5.1. *No two spiders have the same independent set polynomial.*

For integers $k \geq 3$ and $n_1 \geq n_2 \cdots \geq n_k \geq 1$, let $S = S(n_1, n_2, \cdots, n_k)$ be the spider defined by identifying an endvertex of P_{n_i+1} for $i = 1, 2, \cdots, k$. To compute $I(S, z)$ we use the characteristic polynomial of paths. Recall that the characteristic polynomial of G is $\chi(G, z) = \det(zI - A)$ where A is the adjacency matrix. By a standard result of Sachs (see [26]), for a tree T, $[z^{N-2k}]\chi(T, z) = (-1)^k m_k(T)$, where $m_k(T) = [z^k]M(T, z)$. Since $M(P_{n+1}, z) = I(P_n, z)$, it follows that

$$z^{n+1} I(P_n, -z^{-2}) = \chi(P_{n+1}, z).$$

Write $\mathbf{n} = (n_1, n_2, \cdots, n_k)$ and $N = \sum_{i=1}^{k} n_i$. Expanding at the central vertex,

$$\begin{aligned} I(S, -z^{-2}) &= \prod_{i=1}^{k} I(P_{n_i}, -z^{-2}) - z^{-2} \prod_{i=1}^{k} I(P_{n_i-1}, -z^{-2}) \\ &= \prod_{i=1}^{k} \frac{\chi(P_{n_i+1}, z)}{z^{n_i+1}} - z^{-2} \prod_{i=1}^{k} \frac{\chi(P_{n_i}, z)}{z^{n_i}} \\ &= z^{-N-k} \left(\prod_{i=1}^{k} \chi(P_{n_i+1}, z) - z^{k-2} \prod_{i=1}^{k} \chi(P_{n_i}, z) \right). \end{aligned} \tag{5.1}$$

By Cvetković, Doobs and Sachs [26], $\chi(P_n, 2\cos\theta) = \sin(n+1)\theta / \sin\theta$. With $t^{1/2} = \exp(\sqrt{-1}\theta)$, we have

$$\chi(P_n, t^{1/2} + t^{-1/2}) = \frac{t^{(n+1)/2} - t^{-(n+1)/2}}{t^{1/2} - t^{-1/2}}.$$

Substituting this into (5.1), we have

$$\begin{aligned} & z^{N+k} I(S, -z^{-2}) \\ = & \prod_{i=1}^{k} \chi(P_{n_i+1}, t^{1/2}+t^{-1/2}) - (t^{1/2}+t^{-1/2})^{k-2} \prod_{i=1}^{k} \chi(P_{n_i}, t^{1/2}+t^{-1/2}) \\ = & \prod_{i=1}^{k} \frac{t^{(n_i+2)/2} - t^{-(n_i+2)/2}}{t^{1/2}-t^{-1/2}} - (t^{1/2}+t^{-1/2})^{k-2} \prod_{i=1}^{k} \frac{t^{(n_i+1)/2} - t^{-(n_i+1)/2}}{t^{1/2}-t^{-1/2}} \\ = & \frac{t^{-n/2-k}}{(t^{1/2}-t^{-1/2})^k} A(\mathbf{n}, t) \end{aligned}$$

where

$$A(\mathbf{n}, t) = \prod_{i=1}^{k} (t^{n_i+2} - 1) - t(t+1)^{k-2} \prod_{i=1}^{k} (t^{n_i+1} - 1).$$

Proof of Theorem 5.1. For $n_1 \geq \cdots \geq n_k \geq 1$, let $S = S(n_1, n_2, \cdots, n_k)$ be a spider. If $k \leq 2$, then S is a path and is determined by $i(S)$. So we assume $k \geq 3$. By Lemma 3.3, $\sum_{v \in S} \deg^2(v) = 4|S| - 4 + k(k-3)$ and hence k can be computed from $I(S, z)$.

To prove that S is determined by $I(S, z)$, it suffices to show that we can deduce $n_1, n_2, \cdots, n_k$ from $A(\mathbf{n}, t)$. Since $k \geq 3$, the term with lowest degree depending on the n_i's in $A(\mathbf{n}, t)$ is $(-1)^k(k-2)t^{n_k+3}$. Therefore n_k can be deduced.

Suppose $n_{j+1}, n_{j+2}, \cdots, n_k$ can all be deduced from $A(\mathbf{n}, t)$. Then the lowest *yet unaccounted for* monomial is contained in

$$(-1)^{k-1} \left(t^{n_j+2} - t(t+1)^{k-2} t^{n_j+1}\right) + (-1)^{k-2} \left(t^{n_j+n_k+4} - t(t+1)^{k-2} t^{n_j+n_k+2}\right).$$

Since $n_j + n_k + 3 > n_j + 3$, the lowest term is in fact $(-1)^{k-1}(k-2)t^{n_j+3}$. Hence, n_j is determined by $I(S, z)$. By induction, we can find $n_1, \cdots, n_k$ from $I(S, z)$. This shows that S is determined by $I(S, z)$. □

5.2 Well-covered trees

A graph is *well-covered* if every maximal independent set has the same order. For a graph G, the *corona* $G^* = G(K_2)$ is obtained by adding a leaf to each vertex of G. Clearly, $G(K_2)$ is well-covered. In fact, Finbow, Hartnell and Nowakowski [31] showed that a connected well-covered graph, which is not K_1, C_7, with girth at least 6 is the

corona of some graph. Equipped with this characterization, Levit and Mandrescu [65] improved certain results on the zeros of the independent set polynomial for almost all well-covered graphs. Besides, they proved that if a connected graph G has the same independent set polynomial as a well-covered spider, then G is a well-covered spider. More generally, they proposed the following.

Conjecture 5.2 ([65]). *If G is connected and has the same independent set polynomial as a well-covered tree, then G is a well-covered tree.*

Our purpose here is to disprove Conjecture 5.2; a weaker version will be proved in Section 5.3.

Characterizations of well-covered trees predate general graphs and are given as follows. The *internal join* of two trees is obtained by adding an edge to join two trees at vertices of degree at least 2.

Lemma 5.3 ([86, 30, 31, 65]). *For any tree T, the following are equivalent*

1. *T is well-covered;*
2. *$T = K_1$ or T has a perfect matching consisting of pendant edges;*
3. *T is a well-covered spider or the internal join of a number of well-covered spiders.*

Here is an infinite family of counterexamples to Conjecture 5.2. Start with a path $v_1v_2\cdots v_n$. Join a leaf u_i to each v_i. For a vector of nonnegative integers $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, let $H(\mathbf{x})$ be the tree obtained by joining v_i to an endvertex of x_i edges. By Lemma 5.3, $H(\mathbf{x})$ is well-covered.

For $\mathbf{y} = (1, 0, y_3, \cdots, y_n)$, let $K(\mathbf{y}) = H(\mathbf{y}) - u_1 - u_2$. Note that $K(\mathbf{y})$ is not well-covered for any $y_3, \cdots, y_n$.

Well-covered trees sharing the same independent set polynomial need not be isomorphic. Indeed, $I(H(2,2,2,0),z) = I(H(1,0,3,2),z) = (1+z)^3(1+2z)^2(1+13z+61z^2+132z^3+135z^4+53z^5)$. In fact, a tree need not be a well-covered tree to share the independent set polynomial with one.

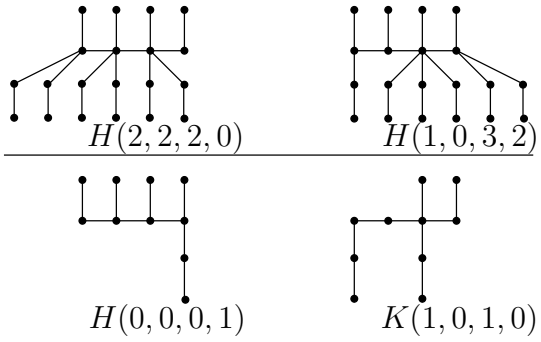


Figure 5.1: Above: well-covered trees with the same independent set polynomial; below: the unique smallest pair of counterexamples to Conjecture 5.2

Proposition 5.4. *For $k \geq 0$, $\mathbf{x} = (0, 0, k, k+1)$ and $\mathbf{y} = (1, 0, k+1, k)$, we have $I(H(\mathbf{x}), z) = I(K(\mathbf{y}), z)$.*

Proof. It is straightforward to check that

$$\begin{aligned} I(H(\mathbf{x}), z) &= (1+4z+3z^2)(1+z)^2(1+2z)^{2k+1} \\ &\quad +z(1+4z+3z^2)(1+z)(1+2z)^k(1+z)^{k+1} \\ &\quad +(z+4z^2+5z^3+2z^4)(1+z)^k(1+2z)^{k+1}, \\ I(K(\mathbf{y}), z) &= (1+4z+3z^2)(1+z)^2(1+2z)^{2k+1} \\ &\quad +z(1+4z+3z^2)(1+z)(1+2z)^{k+1}(1+z)^k \\ &\quad +z(1+z)(1+3z+z^2)(1+z)^{k+1}(1+2z)^k. \end{aligned}$$

Moreover,

$$\begin{aligned} & (1+z)^{-k}(1+2z)^{-k}(I(H(\mathbf{x}), z) - I(K(\mathbf{y}), z)) \\ = \; & z(1+4z+3z^2)(1+z)(1+z) + (z+4z^2+5z^3+2z^4)(1+2z) \\ & -z(1+4z+3z^2)(1+z)(1+2z) - z(1+z)(1+3z+z^2)(1+z) \\ = \; & 0. \end{aligned}$$

Hence, $I(H(\mathbf{x}), z) = I(K(\mathbf{y}), z)$. □

5.3 Independent set polynomial deck

The *independent set polynomial deck*, or the *I-deck* is the multiset $[I(G-v,z) : v \in V(G)]$. The I-deck is a stronger invariant than the independent set polynomial, indeed, the smallest pair of trees having the same I-deck has order 17 as opposed to order 8. We first look at the combinatorial information and then give two methods to construct graphs with the same I-deck.

5.3.1 Information encoded

Lemma 5.5. *Given the I-deck of a graph G, the degree sequence is reconstructible. Moreover, if G is triangle-free, then $[\deg(v), \sum_{u\in\Gamma(v)} \deg(u) : v \in V(G)]$ is reconstructible.*

Proof. For any vertex v, by Lemma 3.3, $i_2(G-v)-i_2(G) = e(G)-e(G-v)-(n-1) = \deg(v)-(n-1)$. Moreover, we have

$$\begin{aligned} i_3(G-v)-i_3(G) &= \binom{n-1}{3} - e(G-v)(n-3) + \sum_{u\in\Gamma(v)} \binom{\deg(u)-1}{2} \\ &\quad - \left(\binom{n}{3} - e(G)(n-2) + \sum_{u\in\Gamma(v)} \binom{\deg(u)}{2} \right) \\ &= K + \frac{1}{2} \sum_{u\in\Gamma(v)} \deg^2(u) - 3\deg(u) + 2 - \deg^2(u) + \deg(u) \\ &= K + \deg(v) - \sum_{u\in\Gamma(v)} \deg(u), \end{aligned}$$

where K is some reconstructible quantity. Hence, we can deduce $\sum_{u\in\Gamma(v)} \deg(u)$. □

Since a spider can be recognized from trees by its degree sequence, coupled with Theorem 5.1, a spider is determined by its I-deck.

Proposition 5.6. *If G is connected and has the same I-deck as a spider S, then $G \cong S$.*

Proof. By Proposition 3.3, $|G| = |S|, e(G) = e(S)$. Hence G is a tree. Furthermore, by Lemma 5.5, G has the same degree sequence as S. Since G has a unique vertex of

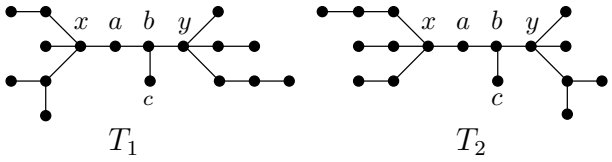


Figure 5.2: Unique smallest pair of trees sharing the same independent set polynomial deck

degree at least 3, G is a spider (if S is a path, then $i(S) = i(G)$ implies $G \cong S$). By Theorem 5.1, $G \cong S$. □

As promised, we prove a weaker version of Conjecture 5.2.

Proposition 5.7. *If G is a connected graph and has the same I-deck as a well-covered tree, then G is a well-covered tree.*

Proof. In fact, we only require that $|G| = n > 1$ is even, $e(G) = n - 1, \deg(I(G, z)) = n/2$ and that G has $n/2$ leaves. Since $I(G, z)$ [15] and the degree sequence are reconstructible from the I-deck, G satisfies all these conditions.

By Lemma 5.3, it suffices to show that every non-leaf vertex is adjacent to a leaf. If v is not a leaf and not adjacent to any, then taking v and the $n/2$ leaves yields an independent set larger than $n/2$. □

5.3.2 Graphs with the same I-deck

Suppose two graphs G_1 and G_2 have the same I-deck, then there is a bijection $f : V(G_1) \to V(G_2)$ such that $I(G_1 - v, z) = I(G_2 - f(v), z)$. We call such a map an *I-hypomorphism*. By Lemma 5.5, v and $f(v)$ have the same degree and, if G is triangle-free, the sum of degree of neighbours as well. Consider the trees in Figure 5.2 which we found by computer to be the smallest pair with the same I-deck. It can be easily checked that an I-hypomorphism can be defined by mapping the middle five vertices (a, b, c, x, y) in the linkage to the corresponding vertices and the rest according to their isomorphism classes in $T - \{a, b, c, x, y\}$ as a rooted tree.

We describe two ways to generate graphs with the same I-deck. It is helpful to extend the definition of $I(G,z)$ to the equivalent bivariate version

$$J(G,z,w)=\sum_{I\in\mathcal{I}(G)} z^{|I|}w^{n-|I|},$$

where $n=|G|$. For any rooted graph (G,u), let $I_0(G,z)=I(G-u,z)$ and $I_u(G,z)=I(G,z)-I(G-u,z)$. Similarly, let $J_u(G,z,w)=J(G,z,w)-J(G-u,z,w)$. Then for any rooted graph (H,h), we have

$$I(G(H),z)=J(G,I_h(H),I(H-h,z)).$$

Note that if $H=K_2$, then this reduces to the following formula in [65] for the corona G^*:

$$z^nI\left(G^*,\frac{1}{z}\right)=(1+z)^nI\left(G,\frac{1}{1+z}\right).$$

Lemma 5.8. *Suppose G_1,G_2 have the same I-deck, then so do $G_1(H)$ and $G_2(H)$ for any rooted graph (H,h).*

Proof. Let $u\in V(G)$. Consider the copy of H attached at u. Let $v\neq u$ be a vertex of this copy. It suffices to show that $I(G(H)-u,z)$ and $I(G(H)-v,z)$ are determined by $I(G-u,z), I(G,z)$ and H. Indeed, by considering extensions of an independent set of G to one of $G(H)$, we have

$$\begin{aligned} I(G(H)-u,z) &= I_0(H,z)\sum_{I\in\mathcal{I}(G-u)} I_h(H,z)^{|I|}I_0(H,z)^{n-1-|I|} \\ &= I_0(H,z)J(G-u,I_h(H,z),I_0(H,z)) \end{aligned}$$

and

$$\begin{aligned} I(G(H)-v,z) &= \sum_{I\in\mathcal{I}(G-u)} I_h(H,z)^{|I|}I_0(H,z)^{n-1-|I|}I(H-h-v,z) \\ &\quad+\sum_{u\in I\in\mathcal{I}(G)} I_h(H,z)^{|I|}I_0(H,z)^{n-|I|}\frac{I_h(H-v,z)}{I_h(H,z)} \\ &= I(H-h-v,z)J(G-u,I_h(H,z),I_0(H,z)) \\ &\quad+J_u(G,I_h(H,z),I_0(H,z)\frac{I_h(H-v,z)}{I_h(H,z)}. \end{aligned}$$

□

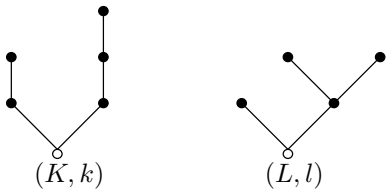


Figure 5.3: Two trees with different parities such that $I(K, z)I(L - l, z) = I(K - k, z)I(L, z)$

The second way is to attach graphs to T_1, T_2 in Figure 5.2.

Lemma 5.9. *Let $(S, s), (R, r)$ be rooted graphs such that $I(R, z)I(S - s, z) = I(R - r, z)I(S, z)$. For $i = 1, 2$, define G_i by identifying s with x and r with y in T_i in Figure 5.2. Then G_1 and G_2 have the same I-deck.*

Proof. Let $f : V(T_1) \to V(T_2)$ be a map such that $I(T_1 - v, z) = I(T_2 - f(v), z)$. It is straightforward to check that $F : V(G_1) \to V(G_2)$ such that $F|T_1 = f, F|_R = id_R$ and $F|_S = id_S$ is an I-hypomorphism. □

To construct trees with the same I-deck, apply Lemma 5.9 with $R = S$ a tree.

5.3.3 Large family of trees sharing the same I-deck

Schwenk [88] proved that there exist arbitrarily large families of cospectral trees. The same method can be applied to the independent set polynomial. In fact, there is an exponentially large family of trees with the same I-deck. Indeed, define an equivalence relation on the set of unlabelled rooted trees $\mathcal{R}_n^U$ on n vertices such that R and S are equivalent iff $I(R, z) = I(S, z)$ and $I(R - r, z) = I(S - s, z)$. Theorem 4.3 implies that a typical tree sampled $\mathcal{R}_n^U$ lies in an exponentially large equivalence class.

Proposition 5.10. *There exists $C > 1$ such that for sufficiently large n, there exist C^n trees of order n with the same I-deck.*

Proof. By Theorem 4.3, there exists $D > 1$ such that for sufficiently large m, there is a set A of at least D^m rooted trees of order m with identical independent set polynomial

and the rooted version. By Lemma 5.9, coalescing every pair of trees from A to x and y of T_1, T_2 produces a pair of trees on $n = 17 + 2m$ vertices with the same I-deck. For even order n, we first coalesce K and L in Figure 5.3 to a pair of trees from A. This method generates at least $kD^{2m} = C^n$ for some constants $k > 0, C > 1$. □

Even though there are exponentially large families of trees with the same I-deck, it remains a question whether a.a.e. tree is distinguished by the I-deck. Perhaps a first step is to consider whether I-deck is 'level one'.

Question 5.11. *Do there exist nonisomorphic rooted trees R, S such that $R \circ T$ and $S \circ T$ have the same independent set polynomial deck for all rooted trees T?*

Summary and future work

In the first part of the thesis, we studied spanning tree congestion of various families of graphs. The main results are :

1. $s(Q_d) = \Theta\left(\frac{\log_2 d}{d} 2^d\right)$;
2. $s(\mathcal{G}(n,p)) = \lceil 1/p \rceil pn + O(\sqrt{pn \log n})$ a.a.s. for $p > \frac{\log n}{n^{1/3}}$;
3. any graph differs by $o(n^2)$ from a graph with linear spanning tree congestion.

Next, we enumerate independent sets in trees. In particular, we prove that

1. $a(P_n) < a(T) < a(K_{1,n-1})$ for any tree $T \neq P_n, K_{1,n-1}$ of order n;
2. any residue is invertible as $i(T)$, as well as $s_r(F,T)$ for $|F| > 1$;
3. the number of trees T with $s_1(F,T)$ in any residue classes grows exponentially fast.

We then consider reconstruction with polynomials and show that

1. there is no switching in general for the independent set polynomial or the Ising polynomial to produce equivalent graphs;
2. typical equivalence class defined by a level one polynomial is exponentially large;
3. the rooted Ising polynomial distinguishes rooted trees;
4. the Negami polynomial of a tree determines the subtree polynomial;
5. the chromatic symmetric function distinguishes caterpillars.

Here are some further questions.

1. A lower bound for $s(P_n^d)$ was derived with the edge-isoperimetry of the grid. Since spanning tree congestion concerns partitioning the whole solid cube rather finding an optimal piece, is it possible to improve the bound? How about using general rectilinear bodies?

2. Is it possible to close the gap in $\mu_p(n)$ in Corollaries 1.50 and 1.52? What are the most congested planar graphs?

3. Conjecture 2.10 on the average size of an independent set in a regular graph.

4. Question 2.28 on inverting the growth rate of (F, r)-matchings.

5. Given (F, r), how do the extremal trees with respect to $s_r(F, T)$ look? What about with respect to the average size of an (F, r)-matching?

6. Is there a natural polynomial invariant which distinguishes a large class of graphs? Question 3.6 on reconstruction with finite homomorphism counts and Stanley's Conjecture 4.11.

7. Does a finite collection of level one polynomials remain level one? Similarly, does the I-deck of trees distinguish rooted trees (Question 5.11)?

8. Questions 4.7 on triple of isomagnetic trees.

Bibliography

[1] S. Akbari, S. Alikhani and Y.-h. Peng, Characterization of graphs using domination polynomials, *European J. Combin.* **31** (2010), 1714–1724.

[2] N. Alon, S. Haber and M. Krivelevich, The number of F-matchings in almost every tree is a zero residue, *Electron. J. Combin.* **18** (2011), #P30.

[3] N. Alon and J.H. Spencer, *The probabilistic method, second edition*, Wiley-Interscience, New York (2000).

[4] D. Andrén and K. Markström, The bivariate Ising polynomial of a graph, *Discrete Appl. Math.* **157** (2009), 2515–2524.

[5] M. Avivaliotis, G. Gordon and W. Graveman, When bad things happen to good trees, *J. Graph Theory* **37** (2001), 79–99.

[6] A.J. Bernstein, Maximally connected arrays on the n-cube, *SIAM J. Appl. Math.* **15** (1967), 1485–1489.

[7] S.L. Bezrukov, Edge isoperimetric problems on graphs, in: *Graph theory and combinatorial biology* (Balatonlelle, 1996), 157–197, Bolyai Soc. Math. Stud. 7, János Bolyai Math. Soc., Budapest (1999).

[8] S.L. Bezrukov, J.D. Chavez, L.H. Harper, M. Röttger and U.-P. Schroeder, The congestion of n-cube layout on a rectangular grid, *Discrete Math.* **213** (2000), 13–19.

[9] N.L. Biggs, *Algebraic Graph Theory, second edition*, Cambridge University Press, Cambridge, England (1993).

[10] B. Bollobás, *Random graphs, second edition,* Cambridge University Press, Cambridge, England (2001).

[11] B. Bollobás and I. Leader, Edge-isoperimetric inequalities in the grid, *Combinatorica* **11** (1991), 299–314.

[12] B. Bollobás, L. Pebody and O. Riordan, Contraction-deletion invariants for graphs, *J. Combin. Theory Ser. B* **80** (2000), 320-345.

[13] B. Bollobás and O. Riordan, Polychromatic polynomials, *Discrete Math.* **219** (2000), 1–7.

[14] C. Borgs, J. Chayes, L. Lovász, V. Sos and K. Vesztergombi, Counting graph homomorphisms, *Algorithms Combin.* **26** (2006), 315-371.

[15] B. Brešar, W. Imrich and S. Klavžar, Reconstructing subgraph-counting graph polynomials of increasing families of graphs, *Discrete Math.* **297** (2005), 159–166.

[16] T. Brylawski, Intersection theory for graphs, *J. Combin. Theory Ser. B* **30** (1981), 233–246.

[17] S.A. Burr, On moduli for which the Fibonacci sequence contains a complete system of residues, *Fibonacci Quart.* **9** (1971), 497–504.

[18] D. Carr, The tree congestion of graphs, REU project, San Bernardino (2005). Available at: `http://www.math.csusb.edu/reu/student_work.html`

[19] A. Castejón, E. Corbacho and R. Vidal, Optimum trees in grids (Spanish), Advances in discrete mathematics in Andalusia, *Serv. Publ. Univ. Cádiz, Cádiz* (2007), 79–86.

[20] A. Castejón and M.I. Ostrovskii, Minimum congestion spanning trees of grids and discrete toruses, *Discuss. Math. Graph Theory* **29** (2009), 511-519.

[21] D. Charles, O. Favaron and D. Rautenbach, Closed formulas for the numbers of small independent sets and matchings and an extremal problem for trees, *Discrete Appl. Math.* **130** (2003), 503–512.

[22] S. Chaudhary and G. Gordon, Tutte polynomials for trees, *J. Graph Theory* **15** (1991), 317–331.

[23] F.R.K. Chung, Z. Füredi, R.L. Graham and P.D. Seymour, On induced subgraphs of the cube, *J. Combin. Theory Ser. A* **49** (1988), 180–187.

[24] R. Conti P. Contucci and C. Falcolini, Polynomial invariants for trees: a statistical mechanics approach, *Discrete Appl. Math.* **81** (1998), 225–237.

[25] A. Cox, Tree congestion for complete n-partite graphs, REU project, San Bernardino (2007).
Available at: `http://www.math.csusb.edu/reu/student_work.html`

[26] D. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs – Theory and Application, third edition*, Barth, Heidelberg (1995).

[27] J. Díaz, J. Petit and M. Serna, A Survey of Graph Layout Problems, *ACM Computing Surveys* **34** (2002), 313–356.

[28] R. Diestel. *Graph Theory, third edition*, Springer-Verlag, Heidelberg (2005).

[29] Y. Egawa and K. Inoue, Radius of $(2k-1)$-connected graphs, *Ars Combin.* **51** (1999), 89–95.

[30] O. Favaron, Very well-covered graphs, *Discrete Math.* **42** (1982), 177–187.

[31] A. Finbow, B. Hartnell and R.J. Nowakowski, A characterization of well covered graphs of girth 5 or greater, *J. Combin. Theory Ser. B* **57** (1993), 44-68.

[32] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press (2009).

[33] H.N. Gabow, A matroid approach to finding edge connectivity and packing arborescences, *ACM* (1991), 112–122.

[34] D. Garijo, A. Goodall and J. Nešetřil, Graph homomorphisms, the Tutte polynomial and "q-state Potts uniqueness", *Electron. Notes Discrete Math.* **34** (2009), 231–236.

[35] G. Gordon and E. McMahon, A greedoid polynomial which distinguishes rooted arborescences, *Proc. Amer. Math. Soc.* **107** (1989), 287–298.

[36] I. Gutman and D.M. Cvetković, The reconstruction problem for the characteristic polynomial of graphs, *Publ. Elektrotehn, Fak. Ser. Fiz. No 498-541, Univ. Beograd* (1975), 45–48.

[37] I. Gutman and O.E. Polansky, *Mathematical concepts in organic chemistry*, Springer-Verlag, Berlin; London (1986).

[38] I. Gutman and F. Harary, Generalizations of the matching polynomial, *Utilitas Math.* **24** (1983), 97–106.

[39] I. Gutman and S. Wagner, Maxima and minima of the Hosoya index and the Merrifield-Simmons index: A survey of results and techniques, *Acta Appl. Math.* **112** (2010), 323–346.

[40] Y.O. Hamidoune, On the numbers of independent k-sets in a claw free graph, *J. Combin. Theory Ser. B* **50** (1990), 241–244.

[41] J. Harant, An upper bound for the radius of a 3-connected graph, *Discrete Math.* **122** (1993), 335–341.

[42] J. Harant and H. Walther, On the radius of graphs, *J. Combin. Theory Ser. B* **30** (1981), 113–117.

[43] L.H. Harper, Optimal assignments of numbers to vertices, *SIAM J. Appl. Math.* **12** (1964), 131–135.

[44] L.H. Harper, *Global methods for combinatorial isoperimetric problems*, Cambridge University Press, Cambridge (2004).

[45] S. Hart, A note on the edges of the n-cube, *Discrete Math.* **14** (1976), 157–163.

[46] M. Hovey, A-cordial graphs, *Discrete Math.* **93** (1991) 183–194.

[47] S. W. Hruska, On tree congestion of graphs, *Discrete Math.* **308** (2008), 1801–1809.

[48] R. Hunter, On spanning tree congestion of product graphs, REU project, San Bernardino (2007).
Available at: `http://www.math.csusb.edu/reu/student_work.html`

[49] T. Iida, Order and radius of 3-connected graphs, *AKCE Int. J. Graphs Comb.* **4** (2007), 21–49.

[50] T. Iida and K. Kobayashi, Order and radius of $(2k-1)$-connected graphs, *SUT J. Math.* **42** (2006), 335–356.

[51] K. Inoue, Radius of 3-connected graphs, *SUT J. Math.* **32** (1996), 83–90.

[52] C. Jordan, Sur les assemblages de lignes, *J. Reine Angew. Math.* **70** (1869), 185–190.

[53] J. Kahn, An entropy approach to the hard-core model on bipartite graphs, *Combin. Probab. Comput.* **10** (2001), 219–237.

[54] S. Khuller, B. Raghavachari and N. Young, Designing multi-commodity flow trees, *Information Processing Letters* **50** (1994), 49–55.

[55] K. Kozawa, Y. Otachi and K. Yamazaki, On spanning tree congestion of graphs, *Discrete Math.* **309** (2009), 4215–4224.

[56] K. Kozawa and Y. Otachi, On spanning tree congestion of Hamming graphs, *preprint*, 2009.

[57] H.-F. Law, Spanning tree congestion of the hypercube, *Discrete Math.* **309** (2009), 6644–6648.

[58] H.-F. Law, On the number of independent sets in a tree, *Electron. J. Combin.* **17** (2010), # N18.

[59] H.-F. Law and M.I. Ostrovskii, Spanning tree congestion: duality and isoperimetry; with an application to multipartite graphs, *Graph Theory Notes N. Y.* **58** (2010), 18-26.

[60] H.-F. Law and M.I. Ostrovskii, Spanning tree congestion of some product graphs, *Indian J. Math.*, to appear.

[61] I. Leader, Discrete isoperimetric inequalities, In: *Probabilistic combinatorics and its applications* (San Francisco, CA, 1991), 57–80, Proc. Sympos. Appl. Math. **44**, Amer. Math. Soc., Providence, RI, (1991).

[62] T.D. Lee and C.N. Yang, Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model, *Physical Rev.* **87** (1952), 410–419.

[63] M. Lepović and I. Gutman, No starlike trees are cospectral, *Discrete. Math.* **242** (2002), 291–295.

[64] V.E. Levit and E. Mandrescu, The independence polynomial of a graph–a survey, *Proceedings of the 1st International Conference on Algebraic Informatics Aristotle Univ. Thessaloniki, Thessaloniki*, (2005), 233–254.

[65] V.E. Levit and E. Mandrescu, On the roots of independence polynomials of almost all very well-covered graphs, *Discrete Appl. Math.* **156** (2008), 478–491.

[66] X. Li, Z. Li, and L. Wang, The inverse problems for some topological indices in combinatorial chemistry, *J. Comput. Biol.* **10** (2003), 47–55.

[67] S.B. Lin and C. Lin, Trees and forests with large and small independent indices, *Chinese J. Math.* **23** (1995), 199-210.

[68] J.H. Lindsey, Assignment of numbers to vertices, *Amer. Math. Monthly* **71** (1964), 508–516.

[69] V. Linek, Bipartite graphs can have any number of independent sets, *Discrete Math.* **76** (1989), 131–136.

[70] C. Löwenstein, D. Rautenbach and F. Regen, On spanning tree congestion, *Discrete Math.* **309** (2009), 4653–4655.

[71] J. Martin, M. Morin and J.D. Wagner, On distinguishing trees by their chromatic symmetric functions, *J. Combin. Theory Ser. A* **115** (2008), 237-253.

[72] R.E. Merrifield and H.E. Simmons, *Topological Methods In Chemistry*, John Wiley & Sons, Inc., New York, (1989).

[73] A. de Mier and M. Noy, On graphs determined by their Tutte polynomials, *Graphs and Combin.* **20** (2004), 105–119.

[74] H.M. Mulder, $(0,\lambda)$-graphs and n-cubes, *Discrete Math.* **28** (1979), 179–188.

[75] S. Negami and K. Ota, Polynomial invariants of graphs. II. *Graphs Combin.* **12** (1996), 189–198.

[76] S.D. Noble and D.J.A. Welsh, A weighted graph polynomial from chromatic invariants of knots, *Ann. Inst. Fourier* **49** (1999), 1057–1087.

[77] M. Noy, Graphs determined by polynomial invariants, *Theoret. Comput. Sci.* **307** (2003), 365–384.

[78] M.I. Ostrovskii, Minimal congestion trees, *Discrete Math.* **285** (2004), 219–226.

[79] M.I. Ostrovskii, Minimum congestion spanning trees in planar graphs, *Discrete Math.* **310** (2010), 1204–1209.

[80] Y. Otachi, H. L. Bodlaender and E. J. van Leeuwen, Complexity results for the spanning tree congestion problem, *Technical Report*, Department of Information and Computing Sciences Utrecht University, Utrecht, The Netherlands (2010).

[81] R. Otter, The Number of Trees, *Ann. Math.* **49** (1948), 583–599.

[82] T. E. Piliposyan and D. O. Muradyan, The problem of finding the length and width of the complete p-partite graph (in Russian), *Uchen. Zapiski Erevan. Gos. Univer.* **144** (1980), 18–25.

[83] D. Peleg, *Distributed computing: a locality-sensitive approach*, SIAM, Philadelphia (2000).

[84] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68** (1937), 145-254.

[85] H. Prodinger and R.F. Tichy, Fibonacci numbers of graphs, *Fibonacci Quart.* **20** (1982), 16–21.

[86] G. Ravindra, Well-covered graphs, *J. Combin. Inform. System Sci.* **2** (1977) 20–21.

[87] I. Sarmiento, The polychromate and a chord diagram polynomial, *Ann. Comb.* **4** (2000), 227–236.

[88] A. Schwenk, Almost all trees are cospectral, *New Directions in the Theory of Graphs, (F. Harary, ed.) Academic Press, New York* (1973), 155–163.

[89] A.D. Scott and A.D. Sokal, The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma, *J. Stat. Phys.* **118** (2005), 1151–1261.

[90] S. Simonson, A variation on the min cut linear arrangement problem, *Math. Systems Theory* **20** (1987), 235–252.

[91] R. Squier, B. Torrence and A. Vogt, The number of edges in a subgraph of a Hamming graph, *Appl. Math. Lett.* **14** (2001), 701–705.

[92] R.P. Stanley, A symmetric function generalization of the chromatic polynomial of a graph, *Adv. Math.* **111** (1995), 166–194.

[93] D. Tanner, Spanning tree congestion critical graphs, REU project, San Bernardino (2007).
Available at: `http://www.math.csusb.edu/reu/student_work.html`

[94] B.D. Thatte, Subgraph posets, partition lattices, graph polynomials and reconstruction (2006).
Available at: `http://arxiv.org/abs/math/0609574`

[95] S. Wagner, Extremal trees with respect to Hosoya index and Merrifield-Simmons index, *MATCH Commun. Math. Comput. Chem.* **57** (2007), 221–233.

[96] S.G. Wagner, Almost all trees have an even number of independent sets, *Electron. J. Combin.* **16** (2009), # R93.

[97] Y. Zhao, The number of independent sets in a regular graph, *Combin. Probab. Comput.* **19** (2010), 315–320.

[98] H. Zhao and X. Li, On the Fibonacci numbers of trees, *Fibonacci Quart.* **44** (2006), 32–38.