

Semi-Inducibility of some small graphs

József Balogh* Bernard Lidický† Dhruv Mubayi‡ Florian Pfender§ Jan Volec¶

January 8, 2026

Abstract

Let H be a fixed graph whose edges are colored red and blue and let $\beta \in [0, 1]$. Let $I(H, \beta)$ be the (asymptotically normalized) maximum number of copies of H in a large red/blue edge-colored complete graph G , where the density of red edges in G is β . This refines the problem of determining the semi-inducibility of H , which is itself a generalization of the classical question of determining the inducibility of H . The function $I(H, \beta)$ for $\beta \in [0, 1]$ was not known for any graph H on more than three vertices, except when H is a monochromatic clique (Kruskal-Katona) or a monochromatic star (Reiher-Wagner). We obtain sharp results for some four and five vertex graphs, addressing several recent questions posed by various authors. We also obtain some general results for trees and stars. Many open problems remain.

1 Introduction

Counting subgraphs inside a host graph is a fundamental problem in extremal combinatorics. A classical question is the following: given a fixed graph F , what is the maximum number of induced copies of F in an n -vertex host graph? This is the inducibility problem introduced by Pippenger and Golumbic [9]. Here we focus on a red-blue (2-edge-colored) refinement of this classical problem. Let H be a graph whose edges are colored red or blue, and let G be a red-blue colored complete graph on n vertices.

*Department of Mathematics, University of Illinois Urbana-Champaign, Urbana, IL, USA, and Extremal Combinatorics and Probability Group (ECOPRO), Institute for Basic Science (IBS), Daejeon, South Korea. Email: jobal@illinois.edu. Supported by NSF grants RTG DMS-1937241, FRG DMS-2152488, the Arnold O. Beckman Research Award (UIUC Campus Research Board RB 24012), the Simons Fellowship and the Institute for Basic Science (IBS-R029-C4).

†Department of Mathematics, Iowa State University, Ames, IA. E-mail: lidicky@iastate.edu. Research of this author is supported in part by NSF FRG DMS-2152490, the Simons Foundation TSM-00013439 and Scott Hanna Professorship.

‡Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, IL 60607. E-mail: mubayi@uic.edu. Research partially supported by NSF Award DMS-2153576.

§Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO. E-mail: Florian.Pfender@ucdenver.edu. Research of this author is supported in part by NSF grant DMS-2152498.

¶Department of Theoretical Computer Science, Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, Prague, 160 00, Czech Republic. E-mail: jan@ucw.cz. Research partially supported by the grant 23-06815M of the Grant Agency of the Czech Republic.

Definition 1. Write $(\#H, G)$ for the number of injections $f : V(H) \rightarrow V(G)$ such that $uv \in E(H)$ is red if only if $f(u)f(v) \in E(G)$ is red. Furthermore, if H has h vertices,

$$\max(H, n) := \max_{G \text{ a red-blue } K_n} (\#H, G) \quad \text{and} \quad \max(H) := \lim_{n \rightarrow \infty} \frac{\max(H, n)}{n^h}.$$

To clarify the normalizing factor in Definition 1, if H is a K_3 with all three edges red, then $(\#H, G)$ is six times the number of red K_3 in G . For $\max(H)$, we chose to use the normalizing factor n^h instead of $(n)_h$ in order to be consistent with the definition of $\max(H)$ defined earlier in [1]. This will not matter, since our results and questions are asymptotic in nature.

The problem of determining $\max(H, n)$ and $\max(H)$ is referred to as the *semi-inducibility* problem. In the classical inducibility setting one demands that, on a chosen set of $|V(H)|$ vertices of G , specified edges are present and all other edges are absent. In contrast, in the semi-inducibility setting some edges of H must appear in the copy (say the red-ones) and some edges must be absent (the blue-ones), but all other pairs are allowed to be either present or absent: thus one allows a “semi-induced” structure.

The semi-inducibility problem was introduced by Basit, Granet, Horsley, Kündgen, and Staden [1], where they also determined $\max(H)$ for alternating colored paths and walks and also for several graphs H on four vertices. Subsequently, $\max(H)$ was determined for the alternating colored six cycle by Chen and Noel [5] and for several more four vertex graphs by Bodnár and Pikhurko [2]. Both [5] and [2] almost exclusively used Flag Algebras. The specific case of the alternating path was also investigated by Chen, Clemen and Noel [4] using entropy.

Our goal is to initiate a systematic attack on a finer version of this problem: we determine or bound $\max(H, n)$ (and hence $\max(H)$) for various small red-blue graphs H , when the underlying graph G has a given density β . More precisely, given $\beta \in [0, 1]$ and red/blue graph H with h vertices, we consider the parameter $I(H, \beta)$ formally defined as follows.

Definition 2. A sequence of red/blue colored cliques $(G_n) = (G_n)_{n=1}^\infty$ has density β if $|V(G_n)| = n$ and the red density of edges in G_n tends to β . The sequence (G_n) is H -good if $\rho(H, (G_n)) := \lim_{n \rightarrow \infty} (\#H, G_n)/n^h$ exists. Then

$$I(H, \beta) = \sup \rho(H, (G_n)),$$

where the supremum is taken over all H -good sequences with density β .

Clearly, $\max(H) = \sup_\beta I(H, \beta)$ over $\beta \in [0, 1]$, but determining the function $I(H, \beta)$ for all β is more challenging and gives a more complete picture of the number of copies of H that can appear in a large graph G . In order to standardize our notation, if a pair is colored red, then we call it an edge and if it is colored blue then we call it a non-edge. This allows us to dispense with red/blue colorings and instead we just consider graphs with certain non-edges specified. In figures, we either use colors red and dotted blue¹ to indicate that red and blue edges are fixed and the colors of the

¹We also refer to the blue edges as non-edges.

remaining non-edges are not specified or we use black and non-edges to describe induced graphs where all pairs are specified.

The function $I(H, \beta)$ is known for all monochromatic cliques and monochromatic stars. The former follows from the Kruskal-Katona theorem and the latter is a result of Reiher and Wagner [11]. In addition, a more recent observation of Liu, Mubayi and Reiher [7], determines $I(H, \beta)$ for the remaining nonmonochromatic 3-vertex graphs H ; when H is a nonmonochromatic triangle, $I(H, \beta)$ is closely related to the problem of minimizing the number of triangles in a graph with a given density, which is answered by a classical result of Razborov [10].

Consequently, the smallest graphs H for which $I(H, \beta)$ is not known are graphs on four vertices. Perhaps the first interesting case is the four vertex 3-edge alternating path AP_4 , which consists of vertices u, v, w, x where uv and wx are red and vw is blue (see Figure 1). More generally, let AP_k be the red/blue colored $(k + 1)$ -vertex path with k edges where every two incident edges have different colors; if k is even the end edges will have distinct colors, while if k is odd, we further stipulate that the first edge and last edge are both red.



Figure 1: Alternating path AP_4 and Alternating 4-cycle AC_4 .

Basit, Granet, Horsley, Kündgen and Staden [1, Corollary 1.4] determined $\max(AP_k)$ when k is even. Chen, Clemen and Noel [4, Theorem 1.2] determined $\max(AP_k)$ for all $k \geq 1$, and Bodnár and Pikhurko [3, Theorem 6.2] also determined $\max(AP_4)$. Despite these results about $\max(AP_k)$, the function $I(AP_k, \beta)$ was not known for any $k \geq 4$.

Our first result completely determines $I(AP_4, \beta)$, answering a question of Basit, Granet, Horsley, Kündgen, and Staden [1, Problem 9.2]. The short proof we give in Section 2 uses only double counting of various quantities related to the vertex degrees in a graph.

Theorem 1.1. *For all $\beta \in [0, 1]$, we have $I(AP_4, \beta) = \beta^2(1 - \beta)$. Equality is achieved only for (asymptotically) regular graphs.*

We generalize part of Theorem 1.1 to the family of double stars defined as follows. For an integer s , we define the alternating double-star $D(s)$ as the following labeled (and ordered) graph on $2s + 2$ vertices: $V(D(s)) = \{v_1, \dots, v_s, v, u, u_1, \dots, u_s\}$, and $E(D(s)) = \{v_1v, \dots, v_sv, uu_1, \dots, uu_s\}$ and $uv \notin E(D(s))$ (see Figure 2).

Clearly $D(1) = AP_4$. The following result determines $I(D(s), \beta)$ for a range of β and also determines $\max(D(s))$.

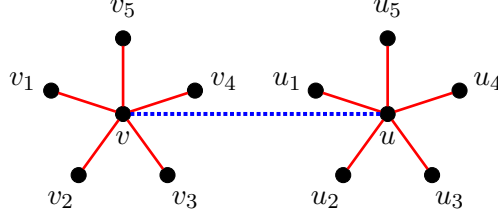


Figure 2: The graph $D(5)$.

Theorem 1.2. Fix $s \geq 1$ and $1 - 1/2s \leq \beta \leq 1$. Then $I(D(s), \beta) = \beta^{2s}(1 - \beta)$. The equality holds for (asymptotically) regular graphs. Moreover, $\max(D(s)) = (2s)^{2s} \cdot (2s + 1)^{-2s-1}$.

Our next result concerns the alternating 4-cycle, which is the graph obtained from AP_4 by adding a non-edge between its endpoints. Let us call this colored graph AC_4 , which has vertices u, v, w, x , red edges uv and wx and blue edges vw and xu (see Figure 1). The number of AC_4 in a graph G is the same as the number of AC_4 in the complement of G . Therefore $I(AC_4, \beta) = I(AC_4, 1 - \beta)$.

In [1], $\max(AC_4)$ was determined, and the question of determining $I(AC_4, \beta)$ for all β was reiterated in Problem 9.3 of [1] (earlier the third author also had posed this question). Our result below answers this question whenever $\beta \in \{1/k, 1 - 1/k\}$ for every positive integer k .

Theorem 1.3. We have $I(AC_4, \beta) \leq \beta^2(1 - \beta)$, with equality if $\beta \in \bigcup_{k \in \mathbb{N}} \{1/k, 1 - 1/k\}$.

Assume $\beta \leq 1/2$. The proof of Theorem 1.3 shows that equality holds only if the underlying graph is almost regular and has only a few 3-edge paths whose endpoints are not adjacent. The only way to make such a construction is to take vertex-disjoint cliques of size $(\beta + o(1))n$. This forces $1/\beta$ to be an integer. We suspect that when $1/\beta$ is not an integer, roughly the same construction is optimal. The following question makes this precise, positing that the behavior of $I(AC_4, \beta)$ when β is not of the form $1/k$ or $1 - 1/k$, is similar to that of the triangle density problem solved by Razborov [10].

Question 1. For all $\beta < 1/2$, and AC_4 -good sequences (G_n) with density β , is it true that the supremum of $\rho(AC_4, (G_n))$ is achieved if G_n comprises $k = \lceil 1/\beta \rceil$ pairwise disjoint cliques, where $k - 1$ of these cliques have (asymptotically) the same size, and the remaining clique has size at most n/k ?

Liu, Mubayi, and Reiher [7, Problem 7.1.] asked if there is a red/blue colored *complete* graph H for which $I(H, \beta)$ has two global maxima (i.e., in the inducibility setting). We show the answer to their problem is yes if we do not require H to be complete (i.e., in the semi-inducibility setting).

Denote by P_{EENN} a path on five vertices with pattern edge-edge-nonedge-nonedge (see Figure 3). Define $K(a, n)$ to be the n -vertex graph consisting of a clique on a vertices and $n - a$ isolated vertices. Given graphs G and H on the same vertex set, recall that the *edit distance* between G and H is the minimum number of edges that we need to delete or add to transform G into H . The

following theorem implies that $I(P_{EENN}, \beta)$ has maximum $27/252$ iff $\beta \in \{9/16, 7/16\}$ and the asymptotic maximizer is $K(a, n)$ or its complement for $a = 3/4$.



Figure 3: Path P_{EENN} .

Theorem 1.4. *For all $\beta \in [0, 1]$, $I(P_{EENN}, \beta) = \max\{\beta^{3/2} - \beta^2, (1 - \beta)^{3/2} - (1 - \beta)^2\}$. Moreover, for a fixed β and $\varepsilon > 0$ there exist $\delta > 0$ and n_0 such for $n > n_0$, every n -vertex red/blue clique G with red density in $[\beta - \delta, \beta + \delta]$ maximizing $(\#P_{EENN}, G)$ has edit distance at most εn^2 from $K(a, n)$ or its complement for some a .*

Denote by $P_{E^\ell N^k}$ a path of length $\ell + k$ with first ℓ edges red and the remaining k edges blue. Notice P_{EENN} is $P_{E^2 N^2}$. Flag algebras experiments suggest that $K(a, n)$ and its complement are also extremal constructions for $P_{E^3 N^3}$. We expect it is the case for longer symmetric paths as well.

Conjecture 1. *For $\ell \geq 2$, $I(P_{E^\ell N^\ell}, \beta)$ is asymptotically achieved by $K(a, n)$ or its complement.*

We next consider the case of monochromatic trees. Here we can dispense with mentioning colors and our problem is to simply maximize the number of copies of a given tree T in a graph with density β . Even for the simple case when T is a star, determining $I(T, \beta)$ is quite difficult. Indeed, depending on the value of β , the maximizers are $K(a, n)$ or the complement of $K(a, n)$ for the appropriate choice of a that results in a graph of density β . This was proved by Reiher and Wagner [11]. We generalize part of their result for arbitrary trees as long as β is close to 1.

Theorem 1.5. *For every tree T there exists $\beta_T \in (0, 1)$ such that for $\beta > \beta_T$, the graph sequence (G_n) where $G_n = K((1 + o(1))\sqrt{\beta}n, n)$ achieves $I(T, \beta)$.*

Note that we need $\beta > \beta_T$ for some β_T in Theorem 1.5, since even for T consisting of two edges and $\beta < 1/2$, the complement of $K((1 + o(1))\sqrt{1 - \beta}n, n)$ has edge density β and more copies of T than $K((1 + o(1))\sqrt{\beta}n, n)$.

If we consider 2-edge-colored trees T , the problem appears to be much more complicated, perhaps even hopeless in general. We pose the following question about $I(T, \beta)$ for certain values of β .

Question 2. *Let T be a red/blue colored tree. Do there exist β_T, β'_T with $0 \leq \beta_T < \beta'_T \leq 1$ such that the following holds? Let $\beta \in [\beta_T, \beta'_T]$. Then $I(T, \beta)$ is asymptotically achieved by the following sequence of graphs (G_n) . The red subgraph R of G_n has the following form: there is a vertex partition $A \cup B = V(G_n)$, all vertices in A have asymptotically the same degree, all edges with one endpoint in A and the other endpoint in B are present, and B is an independent set.*

It is possible that the minimum number of copies of T in graphs with given density β is also asymptotically achieved by the graphs in Question 2. It was proved in [6] that when T is monochromatic, the minimum is achieved for regular graphs.

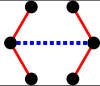
Graph H	$I(H, \beta)$	Construction	Reference
AP_4 	$\beta^2(1 - \beta)$ if $\beta \in [0, 1]$	regular	Theorem 1.1
$D(s)$ 	$\beta^{2s}(1 - \beta)$ if $\beta \in [1 - 1/2s, 1]$	regular	Theorem 1.2
AC_4 	$\beta^2(1 - \beta)$ if $\beta \in \{1/k, 1 - 1/k\}$	regular	Theorem 1.3
P_{EENN} 	$(1 - \beta)^{3/2} - (1 - \beta)^2$ if $\beta \in [0, 1/2]$ $\beta^{3/2} - \beta^2$ if $\beta \in [1/2, 1]$	clique + isolated clique $\cup$ isolated	Theorem 1.4
Tree T	$\beta \in [\beta_T, 1]$	clique $\cup$ isolated	Theorem 1.5
$S_{2,1}$ 	$\beta/4$ if $\beta \in [1/4, 1/2]$ $\beta^2(1 - \beta)$ if $\beta \in [1/2, 1]$	regular $\cup$ isolated regular	Section 7

Table 1: Summary of results from this paper. The depicted instance of $D(s)$ is $D(2)$.

We conclude the introduction by stating a conjecture related to Question 2 for the seemingly simple case of stars. Let $\mathcal{S}$ be the class of graphs obtained from a regular graph by first adding universal vertices and then adding isolated vertices or by first adding isolated vertices and then adding universal vertices. Notice that the initial regular graph may be a complete subgraph or a collection of isolated vertices. Let $S_{a,b}$ be a star with a edges (red) and b non-edges (blue).

Question 3. *Is it true that for each $a, b \geq 1$ and $\beta \in [0, 1]$ there exists an $S_{a,b}$ -good sequence (G_n) with $G_n \in \mathcal{S}$ such that*

$$I(S_{a,b}, \beta) = \rho(S_{a,b}, (G_n))?$$

In Section 7 we answer Question 3 positively for $a = 2, b = 1$ and $\beta \geq 1/4$. More precisely, we prove that

$$I(S_{2,1}, \beta) = \begin{cases} \beta/4 & \text{if } 1/4 \leq \beta \leq 1/2, \\ \beta^2(1 - \beta) & \text{if } 1/2 \leq \beta \leq 1. \end{cases} \quad (1)$$

As we have already mentioned, determining $I(H, \beta)$ includes determining $\max(H)$. While determining $I(H, \beta)$ is wide open, $\max(H)$ has been determined for all 4-vertex graphs [2] with the notable exception of a path with the pattern Edge-Edge-Nonedge. This seems to be a challenging open problem; see Problem 9.2 in [1] and Conjecture 1.1 in [2].

Our results are summarized in Table 1. In Section 2 we prove Theorem 1.1 describing $I(AP_4, \beta)$. Section 3 contains the proof of Theorem 1.2. Theorem 1.3 and Question 1 are discussed in Section 4. Theorem 1.4 is addressed in Section 5 and Section 6 contains Theorem 1.5. Finally, Section 7 discusses possible extremal constructions $\mathcal{S}$ for stars.

2 Alternating 3-edge paths

We begin with a result of Nikiforov that is a very useful tool for counting 3-edge paths in graphs. For completeness, we provide a proof.

Lemma 2.1 (Nikiforov [8]). *Let $G = (V, E)$ be a graph on n vertices with m edges. Then*

$$\sum_{uv \in E} d_u d_v \geq \frac{4m^3}{n^2},$$

where d_w denotes the degree of vertex w . Equality holds iff G is regular.

Proof. We may assume G has no isolated vertices. Set

$$S := \sum_{uv \in E} d_u d_v, \quad U := \sum_{uv \in E} \frac{1}{\sqrt{d_u d_v}}.$$

First note that for every edge uv ,

$$\frac{1}{\sqrt{d_u d_v}} \leq \frac{1}{2} \left(\frac{1}{d_u} + \frac{1}{d_v} \right).$$

Summing over all edges gives

$$U \leq \frac{1}{2} \sum_{uv \in E} \left(\frac{1}{d_u} + \frac{1}{d_v} \right) = \frac{1}{2} \sum_u d_u \cdot \frac{1}{d_u} = \frac{n}{2}.$$

For each of the m edges uv , let $x_{uv} = (d_u d_v)^{1/4}$ and $y_{uv} = (d_u d_v)^{-1/4}$. Let $x = (x_{uv})$ and $y = (y_{uv})$ be the corresponding vectors. Applying the Cauchy-Schwarz Inequality $|x \cdot y| \leq \|x\| \cdot \|y\|$ to the vectors x and y , we obtain

$$\begin{aligned} m^2 &= \left(\sum_{uv \in E} x_{uv} \cdot y_{uv} \right)^2 = |x \cdot y|^2 \leq \|x\|^2 \cdot \|y\|^2 = \left(\sum_{uv \in E} x_{uv}^2 \right) \cdot \left(\sum_{uv \in E} y_{uv}^2 \right) \\ &= \left(\sum_{uv \in E} \sqrt{d_u d_v} \right) \cdot \left(\sum_{uv \in E} \frac{1}{\sqrt{d_u d_v}} \right). \end{aligned}$$

Using $U \leq n/2$, we obtain

$$\sum_{uv \in E} \sqrt{d_u d_v} \geq \frac{m^2}{U} \geq \frac{2m^2}{n}.$$

Combining this with an application of the inequality between quadratic and arithmetic means gives

$$S = \sum_{uv \in E} d_u d_v \geq \frac{1}{m} \left(\sum_{uv \in E} \sqrt{d_u d_v} \right)^2 \geq \frac{1}{m} \left(\frac{2m^2}{n} \right)^2 = \frac{4m^3}{n^2},$$

which completes the proof. Equality holds only if all the d_u are the same due to our applications

of the arithmetic/quadratic mean inequality. $\square$

Recall that AP_4 is the alternating path of length three. A copy of AP_4 in G is a (colored) subgraph of G isomorphic to AP_4 , and hence $(\#AP_4, G)$ is twice the number of copies of AP_4 in G . Using Lemma 2.1, we obtain an upper bound on the number of copies of AP_4 in a graph with given edge density that is tight for regular graphs. This completely determines the profile for the number of AP_4 in a graph. Consequently, Theorem 1.1, which states $I(AP_4, \beta) = \beta^2(1 - \beta)$, is an immediate consequence of Theorem 2.2 below.

Theorem 2.2. *Let $G = (V, E)$ be a graph with m edges. Then*

$$\sum_{uv \notin E} d_u d_v \leq 2m^2 - \frac{2m^2}{n} - \frac{4m^3}{n^2}.$$

Equality holds only if G is regular. Moreover, the number of copies of AP_4 in G is at most

$$2m^2 - \frac{2m^2}{n} - \frac{4m^3}{n^2} - t,$$

where t is the number of 3-element vertex sets in G spanning exactly two edges.

Proof.

$$2m^2 - \frac{1}{2} \sum_v d_v^2 = \frac{1}{2} \left(\sum_v d_v \right)^2 - \frac{1}{2} \sum_v d_v^2 = \sum_{uv \in \binom{V}{2}} d_u d_v = \sum_{uv \notin E} d_u d_v + \sum_{uv \in E} d_u d_v.$$

Rearranging gives

$$\sum_{uv \notin E} d_u d_v = 2m^2 - \left(\frac{1}{2} \sum_v d_v^2 + \sum_{uv \in E} d_u d_v \right).$$

Now, if n and m are fixed, then $\sum_{uv \notin E} d_u d_v$ is maximized when $(\sum_v d_v^2)/2 + \sum_{uv \in E} d_u d_v$ is minimized. The first quantity is minimized only for regular graphs by convexity, and the second quantity is also minimized only for regular graphs due to Lemma 2.1. Therefore $\sum_{uv \notin E} d_u d_v$ is maximized only for regular graphs. The statement about AP_4 follows by observing that $\sum_{uv \notin E} d_u d_v$ is the number of copies of AP_4 together with the number of 3-element sets with exactly two edges (if the neighborhoods of u and v intersect). $\square$

2.1 Density of alternating paths of length 3 via flag algebras

Semi-inducibility problems are approachable by flag algebras as demonstrated by Chen and Noel [5] and Bodnár and Pikhurko [2]. Before finding the proof of Theorem 1.1 presented above, we had determined $I(H, \beta)$ for $1/2 \leq \beta \leq 1$ using flags on at most 4 vertices. Numerical experiments with several fixed values of $\beta < 1/2$ indicate that flag algebras may also be used for $\beta < 1/2$. However, the proof would need flags on at least 6 vertices instead of 4. We expect such a potential flag algebra

proof parametrized by β on 6 vertices to be significantly more complicated. After discovering the proof for the entire range without using flag algebras, we did not further attempt to find such a flag algebra proof. We still include the easy flag algebra proof for $\beta \geq 1/2$ here for readers familiar with the method, but we skip the introduction to the standard notation to save space.

Theorem 2.3. $I(AP_4, \beta) = \beta^2(1 - \beta)$ for $\beta \in [1/2, 1]$.

Proof. Recall that in the definition of $I(AP_4, \beta)$ the subgraph count is scaled by n^4 . However, the flag algebra densities of 4-vertex subgraphs are scaled by $\binom{n}{4}$. Hence our aim is to prove an upper bound $24\beta^2(1 - \beta)$. In flag algebras, the number of AP_4 can be written as

$$O := 4 \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} + 4 \begin{array}{c} \bullet \\ \diagup \\ \bullet \quad \bullet \\ \diagdown \\ \bullet \end{array} + 8 \begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \quad \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + 6 \begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \quad \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + 8 \begin{array}{c} \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \\ \bullet \quad \bullet \end{array}.$$

Our goal is to show $O \leq 24\beta^2(1 - \beta)$. Let $\alpha = 1 - \beta$ be the density of non-edges. We will combine O with a linear combination of the following four squares and an expression fixing the density of non-edges.

$$\begin{aligned} C_1 &:= \left[\left(\begin{array}{c} \bullet \\ \diagup \\ \boxed{1} \quad \boxed{2} \end{array} - \begin{array}{c} \bullet \\ \diagdown \\ \boxed{1} \quad \boxed{2} \end{array} \right)^2 \right] \\ C_2 &:= \left[\left(\begin{array}{c} \bullet \\ \diagup \\ \boxed{1} \quad \boxed{2} \end{array} - \begin{array}{c} \bullet \\ \diagdown \\ \boxed{1} \quad \boxed{2} \end{array} \right)^2 \right] \\ C_3 &:= 12 \left[\left(-\alpha \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \boxed{1} \quad \boxed{2} \end{array} + (1 - 2\alpha) \begin{array}{c} \bullet \\ \diagup \\ \boxed{1} \quad \boxed{2} \end{array} + (1 - \alpha) \begin{array}{c} \bullet \\ \bullet \\ \boxed{1} \quad \boxed{2} \end{array} \right)^2 \right] \\ C_4 &:= 12 \left[\left(-\alpha \begin{array}{c} \bullet \\ \diagdown \quad \diagup \\ \boxed{1} \quad \boxed{2} \end{array} + (1 - 2\alpha) \begin{array}{c} \bullet \\ \diagdown \\ \boxed{1} \quad \boxed{2} \end{array} + (1 - \alpha) \begin{array}{c} \bullet \\ \bullet \\ \boxed{1} \quad \boxed{2} \end{array} \right)^2 \right] \\ E &:= 6 \left(\begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} - \alpha \right) \end{aligned}$$

Notice that all five terms are non-negative. They can be expanded as linear combinations of 4-vertex graphs. The coefficients of these linear combinations when expanded on 4-vertex flags are in Table 2.

The following linear combination yields that the coefficient for each graph on 4 vertices is equal to $24\alpha(1 - \alpha)^2$.

$$\begin{aligned} O &+ (48\alpha^3 - 96\alpha^2 + 48\alpha)C_1 + (48\alpha^3 - 72\alpha^2 + 24\alpha + 12)C_2 + (-4\alpha + 4)C_3 \\ &+ (-4\alpha + 2)C_4 + (-12\alpha^2 + 16\alpha - 4)E = 24\alpha(1 - \alpha)^2. \end{aligned}$$

Notice that the proof works only for $0 \leq \alpha \leq 1/2$ because the factors for C_1, C_2, C_3 and C_4 in the linear combination need to be non-negative, however, the factor for C_4 is negative for $\alpha > 1/2$. Since $\alpha = 1 - \beta$, the result follows. $\square$

	O	C_1	C_2	C_3	C_4	E
	0	0	0	$12\alpha^2$	0	$0 - 6\alpha$
	4	0	0	$10\alpha^2 - 4\alpha$	$2\alpha^2$	$1 - 6\alpha$
	4	$1/6$	0	$10\alpha^2 - 8\alpha + 1$	$4\alpha^2 - 2\alpha$	$2 - 6\alpha$
	0	0	$1/2$	$6\alpha^2 - 6\alpha$	$12\alpha^2 - 12\alpha + 3$	$3 - 6\alpha$
	0	$1/2$	0	$12\alpha^2 - 12\alpha + 3$	$6\alpha^2 - 6\alpha$	$3 - 6\alpha$
	8	$-2/3$	0	0	$4\alpha^2$	$2 - 6\alpha$
	6	$-1/6$	$-1/6$	$4\alpha^2 - 6\alpha + 2$	$4\alpha^2 - 2\alpha$	$3 - 6\alpha$
	0	0	$1/6$	$4\alpha^2 - 6\alpha + 2$	$10\alpha^2 - 12\alpha + 3$	$4 - 6\alpha$
	8	0	$-2/3$	$4\alpha^2 - 8\alpha + 4$	0	$4 - 6\alpha$
	0	0	0	$2\alpha^2 - 4\alpha + 2$	$10\alpha^2 - 16\alpha + 6$	$5 - 6\alpha$
	0	0	0	0	$12\alpha^2 - 24\alpha + 12$	$6 - 6\alpha$

Table 2: Coefficients of expansions into unlabeled graphs on 4 vertices.

3 Density of alternating double stars

In this section we prove Theorem 1.2. This follows from the following theorem. Recall, that for an integer s , the alternating double-star $D(s)$ is the following labeled (and ordered) graph on $2s + 2$ vertices: $V(D(s)) = \{v_1, \dots, v_s, v, u, u_1, \dots, u_s\}$, and $E(D(s)) = \{v_1v, \dots, v_sv, uu_1, \dots, uu_s\}$ and $uv \notin E(D(s))$ (see Figure 2).

Theorem 3.1. *Fix $s \geq 1$. Then*

$$\max(D(s)) \leq \frac{(2s)^{2s}}{(2s+1)^{2s+1}}. \quad (2)$$

If G is an n -vertex graph with average degree $d \geq (1 - 1/(2s))n$, then $(\#D(s), G) \leq (n - d) \cdot n \cdot d^{2s}$.

Proof. Let G have average degree d . We have

$$(\#D(s), G) \leq 2 \sum_{uv \notin E} d_u^s d_v^s \leq \sum_{uv \notin E} (d_u^{2s} + d_v^{2s}) = \sum_u d_u^{2s} \cdot (n - 1 - d_u) < \sum_u d_u^{2s} \cdot (n - d_u). \quad (3)$$

It is easy to give an upper bound:

$$\sum_u d_u^{2s} \cdot (n - d_u) \leq \frac{1}{2s} \sum_u d_u^{2s} \cdot (2sn - 2sd_u) \leq \frac{n}{2s} \left(\frac{2sn}{2s+1} \right)^{2s+1} = \frac{(2s)^{2s}}{(2s+1)^{2s+1}} n^{2s+2}, \quad (4)$$

yielding $\max(D(s)) = (2s)^{2s} \cdot (2s+1)^{-2s-1}$.

To give an upper bound on the sum in (3), as a function of the density of the graph, we use a well-known general optimization trick (and provide its proof).

Lemma 3.2. *For an $\alpha > 0$ define X_α to be the set of vectors $(x_1, \dots, x_n)$ such that $x_i \in \{0, \alpha\}$ for all i except for possibly one coordinate. Suppose that $0 < \gamma < 1$ and $f : [0, 1] \rightarrow \mathbb{R}$ is convex in $[0, \gamma]$ and concave in $[\gamma, 1]$. Then with the restriction that $x_i \in [0, 1]$ and $\sum x_i = D$ for some $0 \leq D \leq n$, $\sum_{i=1}^n f(x_i)$ is maximized at some element of X_α , for some $\alpha \geq \gamma$.*

Proof. Suppose that $x = (x_1, \dots, x_n)$ maximizes the sum subject to $\sum x_i = D$. Then if there were $0 < x_i \leq x_j < \gamma$, then we can decrease x_i and increase x_j with the same amount to increase the sum $\sum_{i=1}^n f(x_i)$ (by convexity), so there are no two such i, j . Similarly, if $\gamma \leq x_i < x_j$ for some i, j , then we can increase x_i and decrease x_j to increase the sum (by concavity), hence there are no two such i, j , i.e., all $x_i \geq \gamma$ are equal to each other (and we define $\alpha \geq \gamma$ to be their value, setting $\alpha = \gamma$ if there is no $x_i \geq \gamma$), and all but at most one of the remaining variables are 0. We conclude that $x \in X_\alpha$ as claimed. $\square$

We apply Lemma 3.2 with $f(x) = (1-x)x^{2s} = x^{2s} - x^{2s+1}$. Using

$$f'(x) = 2s \cdot x^{2s-1} - (2s+1) \cdot x^{2s} \quad \text{and} \quad f''(x) = 2s(2s-1) \cdot x^{2s-2} - 2s(2s+1) \cdot x^{2s-1},$$

we have $f''(x) > 0$ iff $\frac{2s-1}{2s+1} > x$, i.e., we can choose $\gamma = \frac{2s-1}{2s+1}$ and $D = d$. Note that we define $x_u = d_u/n$. Using Lemma 3.2, there is an $\alpha > \gamma$ such that the maximum of f is achieved when m variables are α , (for some m) at most one is between 0 and γ , and the rest of the variables are 0. Observe that $\alpha \geq d/n$ also holds. Ignoring that one outlier variable, whose role is negligible for large n , we have that $\alpha m = d = D$, and we have the following upper bound from Lemma 3.2:

$$\sum_{i=1}^n f(x_i) \leq m \cdot f(\alpha) = \frac{d}{\alpha} (\alpha^{2s} - \alpha^{2s+1}).$$

This implies the following upper bound on (3), which essentially proves (2):

$$n^{2s+1} \frac{d}{\alpha} (\alpha^{2s} - \alpha^{2s+1}) = n^{2s+1} \cdot d \cdot \alpha^{2s-1} \cdot (1 - \alpha).$$

Now assume that $d \geq (1 - 1/(2s))n$. Then $\alpha \geq d/n \geq (2s-1)/(2s)$, and using that $\alpha^{2s-1} \cdot (1 - \alpha)$ is decreasing when $\alpha > (2s-1)/2s$, we have

$$n^{2s+1} \cdot d \cdot \alpha^{2s-1} \cdot (1 - \alpha) \leq n \cdot (n - d) \cdot d^{2s} \leq n^{2s+2} \cdot \frac{(2s)^{2s}}{(2s+1)^{2s+1}}. \quad (5)$$

□

When $s = 1$, Theorem 1.1 shows that the stronger bound $d^2(n - d)n$ holds for every d , while a similar statement does not hold for $s \geq 2$, as a smaller regular graph with additional isolated vertices contains more copies of $D(s)$ than a regular graph.

We shall later use the following proposition which is the $s = 1$ case of (3) and the first inequality in (5).

Proposition 3.3. *Let G be a graph with degree sequence $d_1, \dots, d_n$, and average degree $d \geq n/2$. Then*

$$\sum_i d_i^2(n - d_i) \leq nd^2(n - d).$$

4 Alternating C_4

In this section we present a relatively short proof of Theorem 1.3 and then discuss the case when the density $\beta \neq 1/k$.

Proof of Theorem 1.3. Recall that we are given an n -vertex graph G with density $\beta + o(1)$ and we plan to prove that $(\#AC_4, G) \leq (1 + o(1))\beta^2(1 - \beta)n^4$. Our main (simple) inequality is

$$(\#AC_4, G) \leq 2 \sum_{uv \notin E} d_u d_v.$$

An AC_4 can be obtained by choosing a non-edge uv and neighbors u' of u and v' of v . If $u' \neq v'$ and $u'v' \notin E$, then $uu'v'v$ yields a copy of AC_4 . Since each AC_4 has two non-edges, this counts every copy twice. Moreover, $(\#AC_4, G)$ is four times the number of copies of AC_4 , giving the inequality. More precisely, if we write t for the number of 3-element vertex sets that span exactly two edges and s for the number of 3-edge paths whose endpoints form a non-edge, then

$$\sum_{uv \notin E} d_u d_v = t + s + \frac{1}{2} \cdot (\#AC_4, G).$$

Let $m = |E(G)| = (1 + o(1))\beta \binom{n}{2}$. Applying Theorem 2.2

$$(\#AC_4, G) \leq 2 \sum_{uv \notin E} d_u d_v \leq 4m^2 - \frac{4m^2}{n} - \frac{8m^3}{n^2} \leq 4m^2 - \frac{8m^3}{n^2} = (1 + o(1))\beta^2(1 - \beta)n^4.$$

The first inequality above is asymptotically sharp if $s = o(n^4)$, and the second inequality is asymptotically sharp if G is close to regular. Hence, if G is the disjoint union of cliques of size $(\beta + o(1))n$, then $(\#AC_4, G) = (1 + o(1))\beta^2(1 - \beta)n^4$ as G is close to regular and $s = 0$. Such a G exists only if $k = 1/\beta$ is an integer, completing the proof of the theorem. □

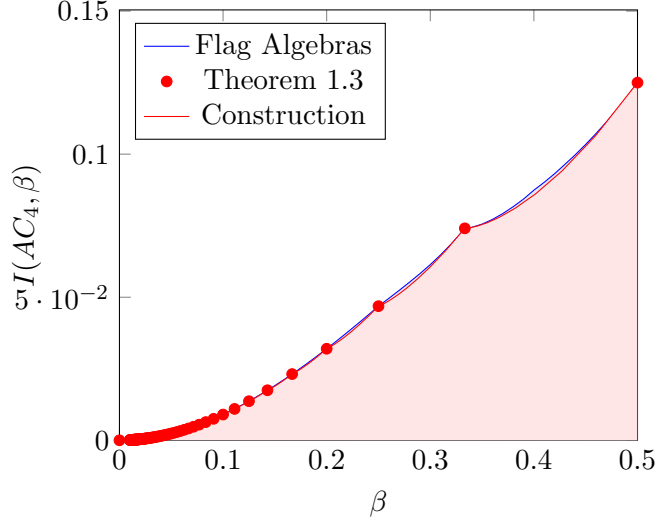


Figure 4: The upper bound obtained by flag algebras calculation on 7 vertices and the lower bound obtained by a construction.

If $1/\beta$ is not an integer, then it is impossible to construct a regular graph with the properties discussed in the proof of Theorem 1.3. Here we suggest that the extremal construction may be obtained by again taking disjoint cliques where all but one of them have the same size, and the remaining clique is smaller, see Question 1.

We tried to compare flag algebra calculations on 7 vertices and this construction for many values of β , and the numbers were close but not equal (see Figure 4). For example, even computing on 8 vertices we obtained a numerical upper bound $I(AC_4, \frac{2}{5}) \leq 0.0859499752$ (improving the bound obtained from a computation on 7 vertices) while the construction gives $I(AC_4, \frac{2}{5}) \geq 0.08566600788$.

The construction in this case (three disjoint cliques) is simple enough that one might expect that a flag algebra computation on a moderate number of vertices can “see” it if it is in fact uniquely extremal. Maybe there are other constructions we are not aware of, matching or beating our construction. On the other hand, such constructions can certainly not be regular, since imposing regularity constraints to our flag algebra calculation gives a numerical upper bound of 0.08409772.

5 Path P_{EENN} .

Theorem 1.4 is a corollary of the following two theorems that deals with $\beta \in [1/2, 1]$. See Figure 5 for an illustration of the resulting bound. The case $\beta \in [0, 1/2]$ follows from symmetry. Recall that $K(a, n)$ is the n -vertex graph comprising a clique on a vertices and $n - a$ vertices of degree 0.

Theorem 5.1. $I(P_{EENN}, \beta) = \beta^{3/2} - \beta^2$ for $\beta \in [1/2, 1]$.

Proof. To simplify the expressions in the proof we let $a = \sqrt{\beta}$. Our goal is to show $I(P_{EENN}, a^2) =$

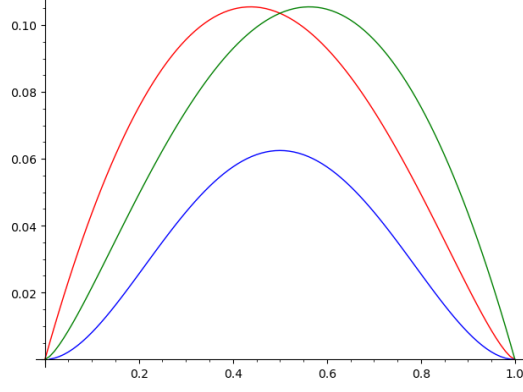


Figure 5: Red and green are bounds from clique and isolated vertices and the complement. Their maximum is $I(P_{EENN}, \beta)$. Blue is a regular graph, perhaps it is the minimum.

$a^3 - a^4$ for $a \in [1/\sqrt{2}, 1]$. The lower bound is implied by $K(a, n)$. For a convergent sequence (G_n) with the corresponding homomorphism ϕ , the scaling for flag algebras is

$$\rho(P_{EENN}, G_n) = \frac{1}{120} \phi(P_{EENN}).$$

In our flag algebra calculations, we omit writing ϕ since they are valid for all ϕ . We can rewrite P_{EENN} as a linear combination of 5-vertex subgraph densities as follows.

$$P_{EENN} = \begin{array}{c} \text{Diagram of } P_{EENN} \\ + 4 \cdot \text{Diagram 1} + 12 \cdot \text{Diagram 2} + 24 \cdot \text{Diagram 3} + 6 \cdot \text{Diagram 4} + 8 \cdot \text{Diagram 5} + 16 \cdot \text{Diagram 6} + 20 \cdot \text{Diagram 7} \\ + 12 \cdot \text{Diagram 8} + 12 \cdot \text{Diagram 9} + 20 \cdot \text{Diagram 10} + 16 \cdot \text{Diagram 11} + 2 \cdot \text{Diagram 12} + 4 \cdot \text{Diagram 13} + 8 \cdot \text{Diagram 14} + 8 \cdot \text{Diagram 15} \\ + 6 \cdot \text{Diagram 16} + 8 \cdot \text{Diagram 17} + 4 \cdot \text{Diagram 18} + 24 \cdot \text{Diagram 19} + 12 \cdot \text{Diagram 20} + 2 \cdot \text{Diagram 21} + 12 \cdot \text{Diagram 22} + 4 \cdot \text{Diagram 23} \end{array}$$

In the flag algebra language, our main goal is to prove that

$$\frac{1}{120} \text{Diagram of } P_{EENN} \leq a^3 - a^4. \quad (6)$$

Since $\beta = a^2$ is the density of edges, we can use $\begin{array}{c} \bullet \\ | \\ \bullet \end{array} - a^2 = 0$. The inequality (6) follows from

$$\begin{aligned}
0 \geq & a^2(1 - a^2) \cdot \left(\begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} - 120(a^3 - a^4) \right) + \left(120a^2(a^3 - a^4) \cdot \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} - 120(1 - a^2)(a^3 - a^4) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \right. \\
& + (-12a^6 + 120a^5 + 80a^4 - 80a^3 + 20a^2 - 20a) \cdot \begin{array}{c} \bullet \\ \diagdown \\ \bullet \end{array} + 15 \cdot B \cdot \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \left. \right) \times \left(\begin{array}{c} \bullet \\ | \\ \bullet \end{array} - a^2 \right) \\
& + 60(a - a^2) \cdot \left[\left(a \begin{array}{c} \bullet \\ | \\ \boxed{1} \end{array} \begin{array}{c} \boxed{3} \\ | \\ \boxed{2} \end{array} - (1 - a) \begin{array}{c} \bullet \\ \diagdown \\ \boxed{1} \end{array} \begin{array}{c} \boxed{3} \\ | \\ \boxed{2} \end{array} \right)^2 \right] + 30 \cdot C \cdot \left[\left(a \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \boxed{1} \quad \boxed{2} \end{array} \begin{array}{c} \boxed{3} \\ | \\ \bullet \end{array} - (1 - a) \begin{array}{c} \bullet \\ \diagdown \\ \boxed{1} \end{array} \begin{array}{c} \boxed{3} \\ | \\ \boxed{2} \end{array} \right)^2 \right], \tag{7}
\end{aligned}$$

where $B = C = \sqrt{2} - 1$ if $\beta \in [1/2, 0.8^2]$ and $B = 0.361, C = 0$ if $\beta \in (0.8^2, 1]$. We remark that there is some freedom in choosing B and C when $\beta > 1/2$, but they are fixed for $\beta = 1/2$. A crucial point for the validity of the inequality (7) is that both $(a - a^2)$ and C i.e., the multiplicative factors for sum of squares, are non-negative.

The evaluation of (7) is similar to the proof of Theorem 2.3. However, it is performed on 5-vertex graphs. Let $\mathcal{F}_5$ be the set of all 5-vertex 2-edge colored graphs up to isomorphism. A tedious yet straightforward enumeration reveals that $|\mathcal{F}_5| = 34$. Also, the corresponding coefficients are more involved. On the other hand, they need to be tight only for subgraphs of the extremal constructions. More precisely, the right hand side of (7) can be expressed as a linear combination of the elements of $\mathcal{F}_5$

$$\sum_{F \in \mathcal{F}_5} c_{a,F} \cdot F \leq \max_{F \in \mathcal{F}_5} c_{a,F} = 0. \tag{8}$$

The interested reader may find all the coefficients $c_{a,F}$ in the in the Appendix of the arXiv preprint of this paper. $\square$

The proof of Theorem 5.1 also implies the (asymptotic) stability of the extremal constructions.

Theorem 5.2. *Let $\beta \in [1/2, 1)$ be fixed. If G is a sufficiently large graph with edge density β maximizing the number of copies of P_{EENN} , then G is in edit distance at most εn^2 from a graph $K(\sqrt{\beta}n, n)$ or its complement if $\beta = 1/2$.*

Proof. If (G_n) is a convergent sequence of graphs with edge densities β achieving $I(P_{EENN}, \beta)$, all the 5-vertex subgraphs with non-zero density must have $c_{a,F} = 0$ in (8). The inspection of $c_{a,F}$ in (8) shows that $c_{a,F} = 0$ for all $a \in [\sqrt{\beta}, 1)$ only for F in

$$\mathcal{A} := \left\{ \begin{array}{c} \bullet \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \diagup \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array}, \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array}, \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \right\}.$$

Additional graphs F with at least one $a \in [\sqrt{\beta}, 1)$ such $c_{a,F} = 0$ appear only at $a = 1/\sqrt{2}$ and they

are the following four

$$\mathcal{A}_{1/2} := \left\{ \begin{array}{c} \text{pentagon with all diagonals} \\ \text{pentagon with 4 diagonals} \\ \text{pentagon with 3 diagonals} \\ \text{pentagon} \end{array} \right\}.$$

However, $c_{1/\sqrt{2}, C_5} = 0$ is an artifact of the proof. A separate flag algebra calculation tailored for $a = 1/\sqrt{2}$ has $c_{a, C_5} < 0$. We conclude that if (G_n) is a convergent sequence achieving $I(P_{EENN}, \beta)$ for $\beta \in [1/2, 1)$, then the densities of the following graphs

$$\mathcal{F} = \left\{ \begin{array}{c} \text{(a)} \\ \text{(b)} \\ \text{(c)} \\ \text{(d)} \\ \text{(e)} \end{array} \right\}$$

tend to zero since they do not appear as induced subgraphs of $\mathcal{A} \cup \mathcal{A}_{1/2} \setminus \{C_5\}$.

Let G be a sufficiently large extremal example. After applying the induced removal lemma to G , we get a graph G' in edit distance εn^2 from G , which contains none of the graphs in $\mathcal{F}$ as induced subgraphs.

Let K be the vertex set of a maximum clique in G' , let I be the set of isolated vertices in G' , and let $R = V(G') - K - I$ be the remaining vertices.

Claim 5.3. *There is at most one vertex k in K such that all the vertices r in R are adjacent to all the vertices in $K \setminus \{k\}$ and not adjacent to k .*

Proof. By the maximality of K , each vertex in R must have a non-neighbor in K . Let r be a vertex in R . Suppose for contradiction that r is not adjacent to two vertices $u, v \in K$. Since r is not isolated, it has a neighbor w . The four vertices u, v, w, r induce one of (b), (c) or (d), giving a contradiction. Hence for each vertex in r , there is at most one vertex u in K such that ur is not an edge. If $|R| = 1$, then the claim is proved.

Let r_1 and r_2 be two distinct vertices in R . Suppose for contradiction that k_1 and k_2 are two distinct vertices in K such that $r_1 k_1$ and $r_2 k_2$ are not edges. If $r_1 r_2$ is an edge, then $r_1 r_2 k_1 k_2$ induce (e) and if $r_1 r_2$ is not an edge $r_1 r_2 k_1 k_2$ induce (c). In both cases we obtain a contradiction. $\square$

Claim 5.4. *The graph induced by R is an independent set.*

Proof. If $r_1 r_2$ is an edge in $G[R]$, then $K \setminus k \cup \{r_1, r_2\}$ is a clique contradicting the maximality of K . $\square$

Claim 5.5. *Either $R = \emptyset$ or $I = \emptyset$.*

Proof. Suppose for contradiction that $r \in R$ and $i \in I$ exist. Let k be a non-neighbor of r in R and u be a neighbor of r in K . Since $\{r, i, k, u\}$ induces (a), we obtain a contradiction. $\square$

The claims imply the statement of the theorem. $\square$

6 General trees

In this section we prove Theorem 1.5. We begin by stating the following result of Reiher and Wagner [11].

Theorem 6.1 (Reiher-Wagner [11]). *Let S_k be the star with k edges, $\beta \in [0, 1]$ and $\eta = 1 - \sqrt{1 - \beta}$. Then*

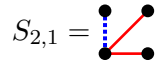
$$I(S_k, \beta) \leq \max\{\beta^{(k+1)/2}, \eta + (1 - \eta)\eta^k\}.$$

We note that the two quantities in the maximum are obtained when G_n is $K(a, n)$ or its complement, for an appropriate a ; moreover, for β sufficiently close to 1, the first term in the maximum, which corresponds to $\rho(S_k, (K(a, n)))$, is larger.

Proof of Theorem 1.5. Recall that we are given a tree T and our aim is to show that there exists $\beta_T \in (0, 1)$ such that for $\beta > \beta_T$, the graph sequence (G_n) where $G_n = K((1 + o(1))\sqrt{\beta}n, n)$ achieves $I(T, \beta)$. We proceed by induction on the number of edges of T . For the base case, we use Theorem 6.1 which applies for all stars. For the induction step, let us assume T is not a star. Then there is an edge e of T such that $T - e$ consists of two trees T_1 and T_2 , each with at least one edge. By induction there exist β_{T_1} and β_{T_2} so let $\beta_T = \max\{\beta_{T_1}, \beta_{T_2}\}$. Consider G with density $\beta > \beta_T$. Then clearly $(\#T, G) \leq (\#T_1, G) \cdot (\#T_2, G)$. Both of the latter quantities are asymptotically maximized when G is a clique and isolated vertices by induction. Moreover, for such a choice of G we have $(\#T, G) \geq (\#T_1, G) \cdot (\#T_2, G)$ up to lower order terms. Indeed, for every two disjoint copies of T_1 and T_2 in G , and for every two vertices v_1 and v_2 in T_1 and T_2 , respectively, the pair v_1v_2 is adjacent in G . An appropriate choice of v_1 and v_2 yields a copy of T . $\square$

7 Stars

In this section we investigate Question 3. Here, we have edges as red and non-edges as blue. Let $S_{a,b}$ be a star with a edges (red) and b non-edges (blue), where $a \geq b$ and $a \geq 2$. We will consider



as the main example.

Recall from the introduction that $\mathcal{S}$ is the class of graphs obtained from a regular graph by first adding universal vertices and then adding isolated vertices or by first adding isolated vertices and then adding universal vertices. The class $\mathcal{S}$ is a combination of several different constructions that seem to appear to be maximizing (or minimizing) $\#S_{a,b}$ for n -vertex graphs for some β . Here we list these special cases of $\mathcal{S}$ and their resulting densities.

(i) The disjoint union of $(1 - s)n$ isolated vertices and a regular graph R on sn vertices for some $s \in (0, 1)$. The graph R has density x , resulting in degree $s nx$ for every vertex in R . This implies

$\beta = s^2x$, solving for x yields $x = \beta/s^2$. The density of labeled $S_{a,b}$ as a function of s is then

$$f(s) = s(sx)^a(1-sx)^b = s \left(\frac{\beta}{s}\right)^a \left(1 - \frac{\beta}{s}\right)^b.$$

By solving $f'(s) = 0$ using Wolfram Alpha² we obtain

$$s = \frac{\beta(a+b-1)}{a-1}, \quad x = \frac{(a-1)^2}{\beta(a+b-1)^2}. \quad (9)$$

Since $s < 1$, (9) provides an upper bound on β . Similarly, since $x \leq 1$, (9) gives a lower bound on β . These bounds are

$$\beta \in \left[\frac{(a-1)^2}{(a+b-1)^2}, \frac{a-1}{a+b-1} \right].$$

After substitution of s from (9) to f we obtain a linear estimate for the density of labeled $S_{a,b}$ as a function of β

$$\ell(\beta) := \beta \frac{(a-1)^{a-1}b^b}{(a-1+b)^{a-1+b}}.$$

(ii) A clique of size $(1-s)n$ completely connected to a regular graph R of size sn . This is the complement of (i), so the previous formulas give us the density for $S_{b,a}$ for edge density $1-\beta$. The density of $S_{a,b}$ as a function of β in this case is

$$\ell c(\beta) := (1-\beta) \frac{a^a(b-1)^{b-1}}{(a+b-1)^{a+b-1}} \quad \text{for} \quad \beta \in \left[1 - \frac{b-1}{a+b-1}, 1 - \frac{(b-1)^2}{(a+b-1)^2} \right].$$

(iii) For a regular graph of edge density β , the density of labeled $S_{a,b}$ is

$$r(\beta) := \beta^a(1-\beta)^b.$$

(iv) A graph G of edge density β with the vertex set partitioned into parts X, Y, Z , where $G[X, Y]$ is complete bipartite, $G[Y]$ is a clique and there are no other edges. The optimal sizes of X, Y and Z and the resulting density of labeled $S_{a,b}$ come from the following polynomial program.

$$s(\beta) := \begin{cases} \text{maximum} & xy^a(1-y)^b + y(x+y)^a(1-x-y)^b \\ \text{subject to} & x+y \leq 1 \\ & 2xy + y^2 = \beta \\ & x, y \geq 0. \end{cases}$$

For fixed values of a, b and β the program has one variable and can be solved using calculus (or more accurately, a computer algebra program). Unfortunately, the solution as a function of β is a root of a polynomial expression depending on a and b . This construction has two special cases

²command: `derivative by s for s(d/s)^a(1-d/s)^b and determine its roots`

when $X = \emptyset$ or $Z = \emptyset$ that are easier to handle and seem to appear often.

(ivX) If $X = \emptyset$ then the clique Y has $\beta^{1/2}n$ vertices and the rest are isolated vertices in Z . The density of labeled $S_{a,b}$ is

$$c(\beta) := \beta^{1/2}\beta^{a/2}(1 - \beta^{1/2})^b.$$

(ivZ) If $Z = \emptyset$ then Y is a set of $(1 - (1 - \beta)^{1/2})n$ universal vertices and $G[X]$ is an independent set. Notice that this is a complement of (ivX). The density of labeled $S_{a,b}$ is

$$cc(\beta) := (1 - \beta)^{1/2}(1 - (1 - \beta)^{1/2})^a(1 - \beta)^{b/2}.$$

(v) The complement of construction (iv). That means a graph G of edge density β with the vertex set partitioned into sets X, Y, Z where Z is the set of universal vertices, $G[X]$ induces a clique and there are no other edges.

$$cs(\beta) := \begin{cases} \text{maximum} & xy^b(1 - y)^a + y(x + y)^b(1 - x - y)^a \\ \text{subject to} & x + y \leq 1 \\ & 2xy + y^2 = 1 - \beta \\ & x, y \geq 0. \end{cases}$$

With these constructions, we state a conjecture of the profile for $S_{2,1}$.

Conjecture 2. *For every $S_{2,1}$ -good (G_n)*

$$\left. \begin{array}{ll} cc(\beta) & \text{if } \beta \in [0, x] \\ c(\beta) & \text{if } \beta \in [x, 1] \end{array} \right\} \leq \rho(S_{2,1}, (G_n)) \leq \begin{cases} s(\beta) & \text{if } \beta \in [0, y] \\ c(\beta) & \text{if } \beta \in [y, 1/4] \\ \ell(\beta) & \text{if } \beta \in [1/4, 1/2] \\ r(\beta) & \text{if } \beta \in [1/2, 1], \end{cases}$$

where $x \in (0, 1)$ is a solution to $cc(x) = c(x)$, i.e. $16x^3 - 40x^2 + 41x - 16 = 0$ and $y \in (0, 1/4)$ is a root of a degree 6 polynomial coming from determining the threshold when the optimal solution to $s(\beta)$ happens at $x = 0$.

The visualization of the conjecture is in Figure 6. Notice that the upper bound in the conjecture describes $I(S_{2,1}, \beta)$. Since the number of labeled copies of $S_{2,1}$ in an n -vertex graph $G = (V, E)$ is upper bounded by $\sum_{v \in V} d_v^2 \cdot (n - d_v)$, Proposition 3.3 proves the upper bound from the conjecture for $\beta \geq 1/2$.

The method of the proof of Proposition 3.3 proves the conjecture for $\beta \geq 1/4$. Indeed, the application of Lemma 3.2 gives as an extremal construction a regular graph plus isolated vertices, which is the construction for $I(S_{2,1}, \beta)$ when $\beta \geq 1/4$. More precisely, the application of Lemma 3.2 yields that the maximum for the sum of functions $f(x) = (1 - x)x^2$ is achieved if there are m values, each

are equal to α . The sum is $m\alpha^2(1 - \alpha) = \beta n\alpha(1 - \alpha)$ since $m\alpha = \beta n$.³ Since $\alpha(1 - \alpha)$ is concave, maximized at $\alpha = 1/2$ and $\beta \leq \alpha$, the maximum of the sum is achieved at $\alpha = \max\{1/2, \beta\}$. In order for the degree sequence of m entries each of value αn to be realized, we need $m \geq \alpha n$, which is equivalent to the condition $\beta n \geq \alpha^2 n$. Since $\alpha = \max\{1/2, \beta\}$, the condition holds for $\beta \geq 1/4$.

We used flag algebras for calculating both upper and lower bounds at multiple fixed values of β and the results were numerically matching the conjecture. While $\ell c(\beta)$ and $cs(\beta)$ are not used in Conjecture 2, we included them for completeness since they are useful for describing $I(S_{a,b}, \beta)$ for other values of a and b .

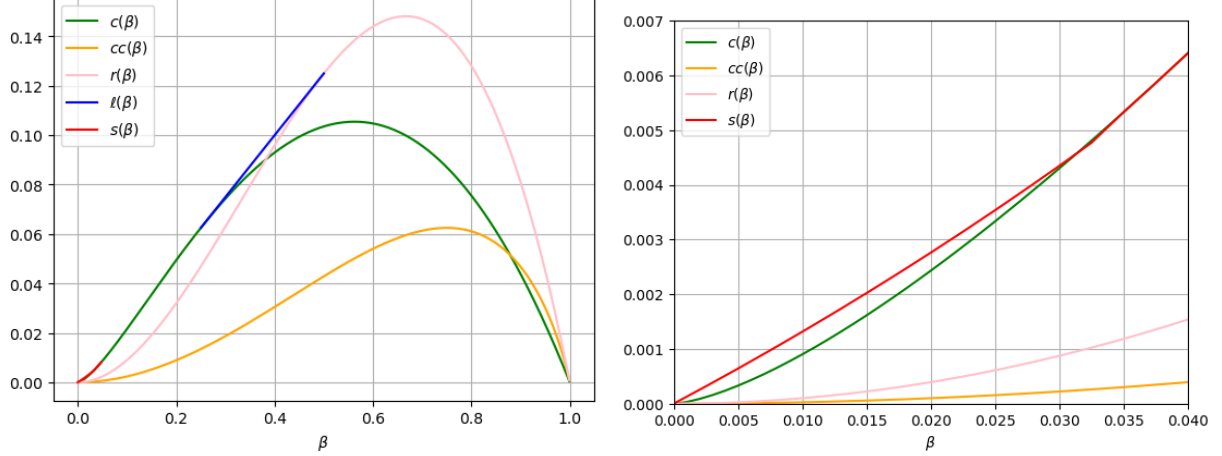


Figure 6: Functions used in Conjecture 2.

8 Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1928930, while the authors were in residence at the Simons Laufer Mathematical Sciences Institute in Berkeley, California, during the spring of 2025. The computer calculations were performed mainly on the Alderaan cluster. The Alderaan cluster was funded by the National Science Foundation Campus Computing program grant OAC-2019089 and a contribution from the Simons Foundation. It is operated by the Center for Computational Mathematics, Department of Mathematical and Statistical Sciences, College of Liberal Arts and Sciences at the University of Colorado Denver.

³Here $\beta \in [0, 1]$ while in the referenced proof $d \in [0, n]$.

References

- [1] Abdul Basit, Bertille Granet, Daniel Horsley, André Kündgen, and Katherine Staden. The semi-inducibility problem, 2025. [arXiv:2501.09842](#).
- [2] Levente Bodnár and Oleg Pikhurko. Semi-inducibility of 4-vertex graphs, 2025. [arXiv:2510.24336](#).
- [3] Levente Bodnár and Oleg Pikhurko. Some exact inducibility-type results for graphs via flag algebras, 2025. [arXiv:2507.01596](#).
- [4] Hao Chen, Felix Christian Clemen, and Jonathan A. Noel. Maximizing alternating paths via entropy, 2025. [arXiv:2505.03903](#).
- [5] Hao Chen and Jonathan A. Noel. On alternating 6-cycles in edge-coloured graphs, 2025. [arXiv:2505.09809](#).
- [6] D. Dellamonica, Jr., P. Haxell, T. Łuczak, D. Mubayi, B. Nagle, Y. Person, V. Rödl, and M. Schacht. Tree-minimal graphs are almost regular. *J. Comb.*, 3(1):49–62, 2012. doi:10.4310/JOC.2012.v3.n1.a2.
- [7] Xizhi Liu, Dhruv Mubayi, and Christian Reiher. The feasible region of induced graphs. *J. Combin. Theory Ser. B*, 158:105–135, 2023. doi:10.1016/j.jctb.2022.09.003.
- [8] Vladimir Nikiforov. The minimum number of 4-cliques in graphs with triangle-free complement, 2005. [arXiv:math/0501211](#).
- [9] Nicholas Pippenger and Martin Charles Golumbic. The inducibility of graphs. *Journal of Combinatorial Theory, Series B*, 19(3):189–203, December 1975. doi:10.1016/0095-8956(75)90084-2.
- [10] Alexander A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007. doi:10.2178/jsl/1203350785.
- [11] Christian Reiher and Stephan Wagner. Maximum star densities. *Studia Sci. Math. Hungar.*, 55(2):238–259, 2018. doi:10.1556/012.2018.55.2.1395.

Additional material

The following material is not part of the manuscript. It is not intended for publication. It includes more details about some of the flag algebras calculations.

A P_{EENN}

This part provides the expansion of (7). Let $\mathcal{F}_5$ be the set of 5-vertex graphs, up to isomorphism. We want to rewrite the right hand side as

$$\sum_{F \in \mathcal{F}_5} c_{a,F} F \leq \max_{F \in \mathcal{F}_5} c_{a,F} = 0.$$

The values of $c_{a,F}$ are the following polynomials in a , B and C .

$$\begin{aligned} \sum_{F \in \mathcal{F}_5} c_{a,F} F = & 0 \cdot \begin{array}{c} \bullet \\ \bullet \end{array} + 0 \cdot \begin{array}{c} \bullet \\ \bullet \end{array} + 0 \cdot \begin{array}{c} \bullet \\ \bullet \end{array} + 0 \cdot \begin{array}{c} \bullet \\ \bullet \end{array} + 0 \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-12a^8 + 12a^7 + 4a^6 - 4a^5 + 12a^4 - 1.5Ba^2 - 16a^3 + 4a^2) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-36a^8 + 36a^7 + 12a^6 - 12a^5 + 18a^4 - 4.5Ba^2 - 30a^3 + 12a^2) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-72a^8 + 72a^7 + 24a^6 - 24a^5 + 24a^4 - 9Ba^2 + 6Ca^2 - 48a^3 - 12Ca + 24a^2 + 6C) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (4a^2 - 4a) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-24a^8 + 24a^7 + 8a^6 - 8a^5 + 12a^4 - 3Ba^2 - 18a^3 + 10a^2 - 4a) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-24a^8 + 24a^7 + 8a^6 - 8a^5 + 12a^4 - 3Ba^2 - 26a^3 + 18a^2 - 4a) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-12a^8 + 12a^7 + 16a^6 - 16a^5 + 2a^4 - 1.5Ba^2 - 2a^3 + 4a^2 + 1.5B - 4a) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-48a^8 + 48a^7 + 28a^6 - 28a^5 + 12a^4 - 6Ba^2 - 20a^3 + 12a^2 + 1.5B - 4a) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \\ & + (-60a^8 + 60a^7 + 32a^6 - 32a^5 + 16a^4 - 7.5Ba^2 + Ca^2 - 36a^3 - 2Ca + 24a^2 + 1.5B + C - 4a) \cdot \begin{array}{c} \bullet \\ \bullet \end{array} \end{aligned}$$

$$\begin{aligned}
& + (-48a^8 + 48a^7 + 16a^6 - 16a^5 + 16a^4 - 6Ba^2 - 24a^3 + 16a^2 - 8a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-24a^8 + 24a^7 + 8a^6 - 8a^5 + 12a^4 - 3Ba^2 - 28a^3 + 22a^2 - 6a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-36a^8 + 36a^7 + 36a^6 - 36a^5 + 2a^4 - 4.5Ba^2 - 4a^3 + 6a^2 + 3B - 4a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-48a^8 + 48a^7 + 40a^6 - 40a^5 + 8a^4 - 6Ba^2 - 20a^3 + 16a^2 + 3B - 4a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-72a^8 + 72a^7 + 60a^6 - 60a^5 + 6a^4 - 9Ba^2 - 10a^3 + 8a^2 + 4.5B - 4a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-60a^8 + 60a^7 + 56a^6 - 56a^5 + 8a^4 - 7.5Ba^2 + Ca^2 - 24a^3 - Ca + 20a^2 + 4.5B - 4a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-108a^8 + 108a^7 + 108a^6 - 108a^5 - 13.5Ba^2 + 9B) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-72a^8 + 72a^7 + 96a^6 - 96a^5 - 9Ba^2 + 6Ca^2 - 12a^3 - 6Ca + 12a^2 + 9B) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (6a^4 - 6a^3 + 6a^2 - 6a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-36a^8 + 36a^7 + 24a^6 - 24a^5 + 8a^4 - 4.5Ba^2 - 12a^3 + 10a^2 + 1.5B - 6a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-48a^8 + 48a^7 + 40a^6 - 40a^5 + 8a^4 - 6Ba^2 - 16a^3 + 14a^2 + 3B - 6a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-36a^8 + 36a^7 + 48a^6 - 48a^5 + 6a^4 - 4.5Ba^2 - 18a^3 + 18a^2 + 4.5B - 6a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-48a^8 + 48a^7 + 16a^6 - 16a^5 + 16a^4 - 6Ba^2 - 24a^3 + 16a^2 - 8a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-60a^8 + 60a^7 + 80a^6 - 80a^5 - 10a^4 - 7.5Ba^2 + 10a^3 + 7.5B) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-72a^8 + 72a^7 + 72a^6 - 72a^5 - 9Ba^2 - 2a^3 + 4a^2 + 6B - 2a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-60a^8 + 60a^7 + 56a^6 - 56a^5 + 6a^4 - 7.5Ba^2 - 12a^3 + 10a^2 + 4.5B - 4a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \\
& + (-84a^8 + 84a^7 + 88a^6 - 88a^5 - 2a^4 - 10.5Ba^2 + 2a^3 + 7.5B) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array}
\end{aligned}$$

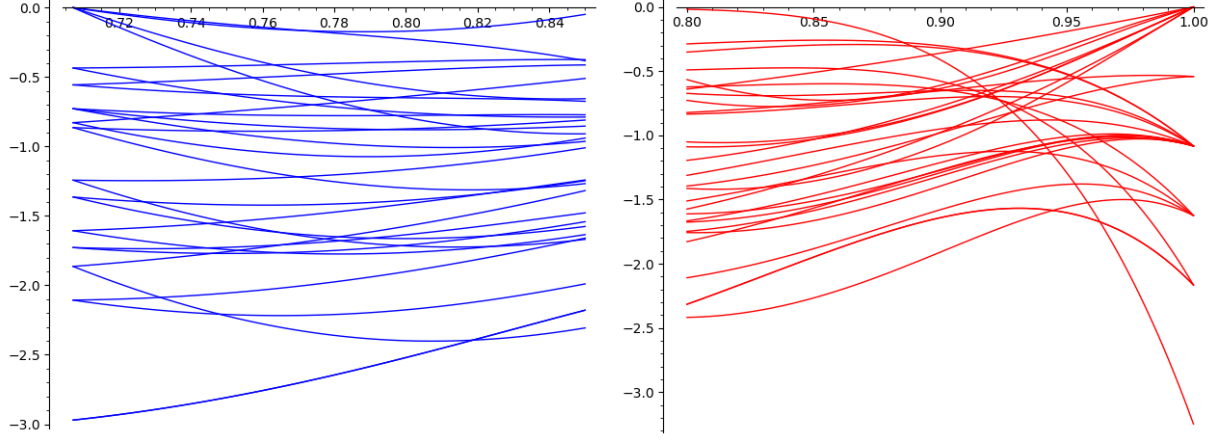


Figure 7: Case $B = C = \sqrt{2} - 1$ on the left and on the right $B = 0.361$ and $C = 0$.

$$\begin{aligned}
& + (-48a^8 + 48a^7 + 64a^6 - 64a^5 + 2a^4 - 6Ba^2 - 6a^3 + 6a^2 + 6B - 2a) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \\
& + (-72a^8 + 72a^7 + 72a^6 - 72a^5 - 9Ba^2 + 2Ca^2 + 6B) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \\
& + (-36a^8 + 36a^7 + 48a^6 - 48a^5 - 4.5Ba^2 + 3Ca^2 + 4.5B) \text{ } \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array}
\end{aligned}$$

Plots of non-zero polynomials $c_{a,F}$ are depicted in Figure 7 for both options of values of B and C .

B Flag Algebras Code

Theorems 2.3 and 5.1 use flag algebras calculations. The code below can be used on a temporary website <https://lidicky.name/review-semi> to check the calculations in this paper. The website is only to assist the reviewers to check the calculation if they wish. The website is work in progress and it is not for general public use.

Code for Theorem 2.3. Direct link.

```
# The following code sums to 0 for any pa
d pa = 0.3 # change this number if you wish
1 4 0 2 0 0 1 0 2 EC - 24*pa*(1-pa)^2 0 0
+{48*pa^3 - 96*pa^2 + 48*pa } [ ( 1 3 2 2 2 1 - 1 3 2 2 1 2 )^2 ] \\
+{ 48*pa^3 - 72*pa^2 + 24*pa + 12 } [ ( 1 3 2 1 2 1 - 1 3 2 1 1 2 )^2 ] \\
+{-4*pa + 4}*12 [(-1*pa 3 2 2 2 2 + {1-2*pa} 3 2 2 2 1 + {1-pa} 3 2 2 1 1 ) ^2 ] \\
+{-4*pa + 2}*12 [(-1*pa 3 2 1 2 2 + {1-2*pa} 3 2 1 2 1 + {1-pa} 3 2 1 1 1 )^2] \\
+{-12*pa^2 + 16*pa - 4}*6(1 2 0 1 - pa 0 0 )
@4
```

Code for Theorem 5.1. Direct link.

```
# pe goes from 1/2 to 0.5625 to 1
# hence pa goes from sqrt(1/2)=0.707 to 0.75 to 1
d pa = 0.85 # change here as you like
d pe = pa^2
# Regime from edge density 1/2 to 0.8^2
#d pC = (2^(1/2)-1)
#d pB = (2^(1/2)-1)

# Regime from edge density 0.8^2 to 1
d pB = 0.361
d pC = 0

1*(1-pe)*pe( (1 5 0 2 0 0 0 2 0 0 1 0 1 EC ) - 120*{pa^3-pe^2} 0 0 ) \\
+120*pe*(-1*pe^2 + pa^3) 3 0 2 2 2 * ( 1 2 0 2 - pe 0 0 ) \\
-120*(1-pe)*(pa^3 - pe^2) 3 0 1 1 1 * ( 1 2 0 2 - pe 0 0 ) \\
+60*(pa-pa^2) [ ( pa 4 3 1 1 1 2 1 1 -{1-pa} 4 3 1 1 1 2 2 2 )^2 ] \\
+ 20*(-6*pa^6 + 6*pa^5 + 4*pa^4 - 4*pa^3 + pa^2 - pa) 3 0 1 1 2 * ( 1 2 0 2 - pe 0 0 ) \\
+ 15*pB 3 0 1 2 2 * ( 1 2 0 2 - pe 0 0 ) \\
+ 30*pC [ ( pa 4 3 1 2 2 2 2 2 -{1-pa} 4 3 1 2 1 2 1 2 )^2 ]
@5
```