On a Conjecture of Nagy on Extremal Densities

A. Nicholas Day* Amites Sarkar†

October 12, 2020

Abstract

We disprove a conjecture of Nagy on the maximum number of copies $N(G, H)$ of a fixed graph G in a large graph H with prescribed edge density. Nagy conjectured that for all G, the quantity $N(G, H)$ is asymptotically maximised by either a quasi-star or a quasi-clique. We show this is false for infinitely many graphs, the smallest of which has 6 vertices and 6 edges. We also propose some new conjectures for the behaviour of $N(G, H)$, and present some evidence for them.

1 Introduction

Let G be a fixed small graph, let $\beta \in (0, 1)$, and let n be a large positive integer. The problem of asymptotically minimizing the number of copies of G in a large graph H on n vertices, with edge density β, is a very well-studied problem in extremal graph theory. It generalises the forbidden subgraph problem, and has received much attention in recent years. For instance, a famous conjecture of Sidorenko [15] states that when G is bipartite, the minimiser H is quasirandom, and a celebrated theorem of Reiher [13] solves the problem for complete graphs (the minimiser is close to a Turán graph).

In this paper we will study the opposite problem: given G, how do we *maximise* the number of copies of G in H? As before, H will have order n and edge density β, and, as before, we are mainly interested in asymptotics: we will write "maximiser" for "asymptotic maximiser" throughout. This problem also has a long history, going back at least to Ahlswede and Katona [1], who studied the case when $G = P_2$, the path with two edges. (Throughout this paper, P_l denotes a path with l *edges*.) Roughly speaking, Ahlswede and Katona proved that, for $\beta > \frac{1}{2}$, the maximiser is a *quasi-clique*, i.e., a clique K, with another vertex joined to a subset of $V(K)$, together with some isolated vertices; we note that the size of the clique is uniquely determined

*Institutionen för Matematik och Matematisk Statistik, Umeå Universitet, 901 87 Umeå, Sweden. Email: nicholas.day@umu.se. Research supported by Swedish Research Council grant 2016-03488.

†Department of Mathematics, Western Washington University, Bellingham, Washington 98225, USA.

by n and β. When $\beta < \frac{1}{2}$, they proved further that the maximiser is instead a *quasi-star*: the complement of a quasi-clique; the parameters of the quasi-star are again uniquely determined by n and β. In short, the maximiser is first a quasi-star and then a quasi-clique, with the "flip" occurring at $\beta = \frac{1}{2}$. The paper [1] in fact contains an exact result for $G = P_2$, which is surprisingly complicated.

On the other hand, for some graphs G, the maximiser is always a quasi-clique, regardless of β. Alon showed that this is the case for any graph with with $\alpha^*(G) = v/2$, where $\alpha^*(G)$ is the fractional independence number of G, and v is the number of vertices of G; see Section 2 for a definition of the fractional independence number of a graph. A result of Janson, Oleszkiewicz and Ruciński [8] can be used to show that this condition is in fact an "if and only if" condition, that is, if the maximiser of G is a quasi-clique for all β, then $\alpha^*(G) = v/2$. Alon's result (which was originally formulated in a different way) was generalised to hypergraphs by Friedgut and Kahn [6].

These results leave many questions open. To restate the basic one: given a small graph G on v vertices, we would like to know which large graphs H asymptotically maximise $N(G, H)$, the number of unlabelled copies of G in H, where H runs over all graphs on n vertices and edge density β. Interest in this problem was revitalised by its connection with graphons, and subsequently by the work of Nagy [11], who solved it for $G = P_4$. Nagy's result is that for P_4, as for P_2, the maximiser is first a quasi-star and then a quasi-clique, with the flip occurring this time at $\beta = 0.0865...$, instead of $\frac{1}{2}$. By contrast, odd-length paths such as P_3 are covered by Alon's theorem: for them, the maximiser is always a quasi-clique.

Note that, up until now, we have been tacitly assuming that the maximiser is in some sense unique; however, this might conceivably not be the case, so we should strictly speaking write "an (asymptotic) maximiser", rather than "the maximiser".

At the end of his paper [11], Nagy posed three questions. The first asked for an exact, not just asymptotic, result for P_4. The second asked whether, for every graph G, and every edge density β, the quantity $N(G, H)$ is always (asymptotically) maximised when H is either a quasi-star or a quasi-clique. The third question was more cautious: given G, is there always a nontrivial threshold $\beta_G < 1$, such that the quasi-clique (asymptotically) maximises $N(G, H)$ for $\beta > \beta_G$?

While the first question remains out of reach, the third was recently answered in the affirmative by Gerbner, Nagy, Patkós and Vizer [7], using a recent result of Reiher and Wagner [14] (who had in turn extended an earlier result of Kenyon, Radin, Ren and Sadun [10]). Reiher and Wagner answered Nagy's second question affirmatively for stars, i.e., they proved that the maximiser is first a quasi-star and then a quasi-clique when G is a star, with the flip always occurring at $\beta < 1$. Both [10] and [14] make extensive use of graphons.

In this paper, we give a negative answer to the second question. Specifically, we exhibit a 6-vertex graph G_6, such that neither the quasi-star nor the quasi-clique asymptotically maximise $N(G_6, H)$ at any edge density $\beta \in (0, 0.016)$. More generally, we show that any graph G on v

vertices satisfying $\alpha^*(G) > \max(\alpha(G), v/2)$ is also a counterexample.

The plan of this paper is as follows. In Section 2 we introduce some basic definitions and notation, as well as the graphs $T_n^e(q)$, which lie at the heart of the paper. Section 3 contains our main result, which provides a family of counterexamples to Nagy's conjecture. In Section 4 we describe the prior work of Janson, Oleszkiewicz and Ruciński mentioned above, and discuss how it relates to these counterexamples. Finally, in Section 5 we propose some new conjectures, which we hope will inspire further progress in this area.

2 Definitions and notation

For a fixed small graph G on v vertices, we write

$$\mathrm{ex}(n, e, G) = \max\{N(G, H) : |H| = n, e(H) \leqslant e\},$$

where $N(G, H)$ is the number of unlabelled copies of G in H. We will also wish to consider labelled copies of G; in this case we will fix a labelling G_l of G, and work with $N_l(G_l, H) = N(G, H)|\mathrm{AutG}|$ instead. For example, $N(P_4, C_7) = 7$ and $N(P_4, K_5) = 60$. With G and $\beta \in [0, 1]$ fixed, a family of graphs $(H_n)_{n\geq 1}$, such that each H_n has n vertices and edge density $\beta+O(1/n)$, is an *asymptotic maximiser* for G at edge density β if

$$N(G, H_n) = (1 + o(1))\mathrm{ex}(n, \lfloor \beta n^2/2 \rfloor, G).$$

A graph homomorphism from G_l to H is a map $f : V(G_l) \to V(H)$ such that $\{f(u), f(w)\} \in E(H)$ for all $\{u, w\} \in E(G_l)$. We write $\hom(G_l, H)$ for the number of homomorphisms from G_l to H. Given a family of graphs $(H_n)_{n\geqslant 1}$ such that $|V(H_n)| = n$, we have that $N_l(G, H_n) \leqslant \hom(G_l, H_n)$ and also $N_l(G, H_n) = (1 + o(1)) \hom(G_l, H_n)$ as $n \to \infty$. As such, we define

$$t(G_l, H_n) = \lim_{n\to\infty} \frac{N_l(G_l, H_n)}{n^v} = \lim_{n\to\infty} \frac{\hom(G_l, H_n)}{n^v};$$

the limit will exist for all the families H_n we consider. We will switch between working with $\hom(G_l, H_n)$, $N_l(G, H_n)$ and $t(G_l, H_n)$, depending on which is convenient at the relevant time.

Given a graph G, a function $\phi : V(G) \to [0, 1]$ such that $\phi(u)+\phi(w) \leqslant 1$ for all $\{u, w\} \in E(G)$ is known as a *fractional independence weighting* of G. The *fractional independence number* of G, written $\alpha^*(G)$, is defined as the maximum of $\sum_{u\in V(G)} \phi(u)$ over all fractional independence weightings of G. We will make crucial use of the following result of Nemhauser and Trotter [12]: any graph G has a maximal weighting (one that realises $\alpha^*(G)$) in which all the weights are either $0, \frac{1}{2}$ or 1. As is customary, we write $\alpha(G)$ for the usual independence number of G, that is, $\alpha(G)$ is the size of the largest independent set in G. Given an independent set $X \subseteq V(G)$, the function $\phi : V(G) \to \{0, 1\}$, with $\phi(u) = 1$ if $u \in X$ and $\phi(u) = 0$ otherwise, is a fractional independence weighting of G. Thus we have that $\alpha^*(G) \geqslant \alpha(G)$ for all graphs G.

We now turn to some specific families of graphs. Given n and $e \leq \binom{n}{2}$, there is a unique quasi-clique K_n^e with n vertices and e edges. To define it, we first write $e = \binom{a}{2} + b$, where $0 \leq b < a$. The graph K_n^e is a complete graph K_a with a vertices, with an additional vertex joined to b vertices of K_a, and $n - a - 1$ isolated vertices. Likewise, there is a unique quasi-star S_n^e with n vertices and e edges; this is just the complement of $K_n^{e'}$, where $e' = \binom{n}{2} - e$.

Here we will be interested in asymptotics only. Thus we will replace the number of edges e by the (asymptotic) *edge density* $\beta = 2e/n^2$. For asymptotic purposes, K_n^e is a clique of size $\sqrt{\beta}n$, and $V(S_n^e)$ can be partitioned into two sets R_S (red vertices) and B_S (blue vertices), where the red vertices span a clique of size $(1 - \sqrt{1-\beta})n$, the blue vertices form an independent set of size $\sqrt{1-\beta}n$, and every blue vertex is joined to every red vertex. Here, and in what follows, we omit floor functions for ease of notation; our "approximate versions" of K_n^e, S_n^e and (in the next paragraph) $T_n^e(q)$ will not have exactly n vertices and e edges. However, we will have, for instance, $e(T_n^e(q)) = (1 + O(1/n))\, e$, and this is more than enough for our asymptotic estimates.

Let $q \in [0, 1]$. The following graph $T = T_n^e(q)$, with (asymptotically) n vertices and e edges, will prove useful. We partition the vertices of T into three sets Y_T (yellow), R_T (red) and B_T (blue), with the following sizes:

$$\begin{aligned} |Y_T| &= \sqrt{\beta}qn, \\ |R_T| &= \left(1 - \sqrt{1 - \beta\left(1 - q^2\right)}\right) n, \\ |B_T| &= \left(\sqrt{1 - \beta\left(1 - q^2\right)} - \sqrt{\beta}q\right) n. \end{aligned}$$

The sets Y_T and R_T both span cliques, while B_T is an independent set. Also, every vertex in R_T is connected to every vertex in Y_T and B_T. It is easy to check that $T_n^e(q)$ has the required number of vertices and edges. Moreover, we have that $T_n^e(0) = S_n^e$ and $T_n^e(1) = K_n^e$, so that $T_n^e(q)$ interpolates between S_n^e and K_n^e. See Figure 1 for a picture of S_n^e and $T_n^e(q)$.

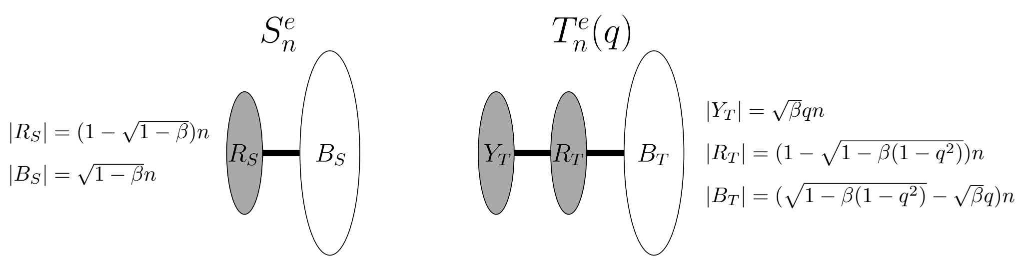


Figure 1: The graphs S_n^e and $T_n^e(q)$ where $e = \frac{\beta n^2}{2}$. The shaded sets are cliques, while the white sets are independent. A black line between two sets indicates they are fully connected.

3 Main result

In this section we prove the following theorem, which disproves the conjecture of Nagy.

Theorem 1. *Let G_l be a graph on v vertices such that $\alpha^*(G_l) > \max\left(\alpha\left(G_l\right), \frac{v}{2}\right)$. Fix $q \in (0,1)$. Then there exists $\epsilon = \epsilon(G_l, q) > 0$ such that, for all $\beta \in (0, \epsilon)$, we have*

$$t\left(G_l, T_n^e\left(q\right)\right) > \max\left(t\left(G_l, K_n^e\right), t\left(G_l, S_n^e\right)\right).$$

We remark that there are infinitely many graphs G that have $\alpha^*(G) > \max\left(\alpha\left(G\right), \frac{v}{2}\right)$. For example, for $a \geqslant 3$, $b \geqslant 2$, consider the following graph that has $a+b+1$ vertices. We start by taking a clique K_a, we then add a single vertex u to our graph and connect it to one vertex of K_a. We finally add b more vertices to our graph, and connect them all to u. It is easy to see that $\alpha^*(G) = b + \frac{a}{2}$, $\alpha(G) = b+1$ and $\frac{v}{2} = \frac{a+b+1}{2}$. The smallest such graph satisfying $\alpha^*(G) > \max\left(\alpha\left(G\right), \frac{v}{2}\right)$ occurs when $a = 3$ and $b = 2$. We call this graph G_6, and we study it more carefully in Section 3.1.

The proof of Theorem 1 will follow from the two homomorphism counting lemmas below; however we first need to introduce some more notation. Given a labelled graph G_l, let $\Phi(G_l)$ be the set of fractional independence weightings of G_l in which every vertex receives a weight in $\{0, \frac{1}{2}, 1\}$. Given $\phi \in \Phi(G_l)$, let

$$R_\phi = \Big\{w \in V(G) : \phi(w) = 0\Big\},$$
$$Y_\phi = \Big\{w \in V(G) : \phi(w) = \frac{1}{2}\Big\},$$
$$B_\phi = \Big\{w \in V(G) : \phi(w) = 1\Big\},$$

and let $r_\phi = |R_\phi|$, $y_\phi = |Y_\phi|$, and $b_\phi = |B_\phi|$. For all $q \in [0,1]$, we define

$$y(q) = \sqrt{\beta} q,$$
$$r(q) = 1 - \sqrt{1 - \beta(1-q^2)},$$
$$b(q) = \sqrt{1 - \beta(1-q^2)} - \sqrt{\beta} q.$$

Recall that for any $q \in [0,1]$, the graph $T_n^e(q)$ has three vertex classes Y_T, R_T and B_T, with $|Y_T| = y(q)n$, $|R_T| = r(q)n$, and $|B_T| = b(q)n$, where Y_T and R_T both span cliques, B_T is an independent set, and every vertex in R_T is connected to every vertex in Y_T and B_T.

Lemma 1. *Let G_l be a labelled graph, and fix $\beta, q \in [0,1]$. Then*

$$t\left(G_l, T_n^e\left(q\right)\right) = \sum_{\phi \in \Phi(G_l)} y(q)^{y_\phi} r(q)^{r_\phi} b(q)^{b_\phi}.$$

Lemma 2. *Let G_l be a labelled graph on v vertices with no isolated vertices. Fix $q \in (0,1)$, and let $\beta \to 0$. Then there exist constants $C_1 = C_1(G_l, q) > 0$ and $C_2 = C_2(G_l) > 0$ such that the following all hold:*

1. $t\left(G_l, K_n^e\right) = \beta^{\frac{v}{2}}$,
2. $t\left(G_l, T_n^e(q)\right) = C_1\left(\beta^{v-\alpha^*(G)} + O\left(\beta^{v-\alpha^*(G)+\frac{1}{2}}\right)\right)$,
3. $t\left(G_l, S_n^e\right) = C_2\left(\beta^{v-\alpha(G)} + O\left(\beta^{v-\alpha(G)+1}\right)\right)$.

In Corollary 3 below we give explicit values of the constants C_1 and C_2. We remark that, just as with the constants C_1 and C_2, the constants hidden in the big O notation in this lemma may depend on G_l and q, but no other variables.

Proof of Lemma 1. Given a homomorphism f from G_l to $T_n^e(q)$, we give a weighting ϕ_f to the vertices of G_l in the following way:

$$\phi_f(u) = \begin{cases} 0 & \text{if } f(u) \in R_T, \\ \frac{1}{2} & \text{if } f(u) \in Y_T, \\ 1 & \text{if } f(u) \in B_T. \end{cases}$$

It is easy to see that ϕ_f is a fractional independence weighting of G_l. Given $\phi \in \Phi(G_l)$, let $\hom_\phi(G_l, T_n^e(q))$ be the number of homomorphisms f from G_l to $T_n^e(q)$ such that $\phi_f = \phi$, and let

$$t_\phi(G_l, T_n^e(q)) = \lim_{n\to\infty} \frac{\hom_\phi(G_l, T_n^e(q))}{n^v}.$$

A homomorphism f from G_l to $T_n^e(q)$ has the property that $\phi_f = \phi$ if and only if $f(R_\phi) \subseteq R_T$, $f(Y_\phi) \subseteq Y_T$, and $f(B_\phi) \subseteq B_T$. Therefore

$$t_\phi\left(G_l, T_n^e(q)\right) = y(q)^{y_\phi} r(q)^{r_\phi} b(q)^{b_\phi}. \tag{1}$$

As each homomorphism from G_l to $T_n^e(q)$ gives rise to a fractional independence weighting of G_l as described above, we have that

$$t(G_l, T_n^e(q)) = \sum_{\phi\in\Phi(G_l)} t_\phi(G_l, T_n^e(q)). \tag{2}$$

Combining (1) with (2) gives the result. □

Proof of Lemma 2. We start by proving the first part of the lemma. The quasi-clique K_n^e consists of a clique X on $\sqrt{\beta}n$ vertices, and $(1-\sqrt{\beta})n$ isolated vertices. As G_l has no isolated vertices, a map $f : V(G_l) \to V(K_n^e)$ is a homomorphism from G to K_n^e if and only if $f\left(V\left(G_l\right)\right) \subseteq V(X)$, and so $t\left(G_l, K_n^e\right) = \beta^{\frac{v}{2}}$, as required.

We now proceed by proving the remaining two parts of the lemma. Recall that the Taylor series for $1-\sqrt{1-x}$ about 0 is $\frac{x}{2}+O(x^2)$. Thus, as $\beta \to 0$, we have that

$$\begin{aligned} r(q) &= \frac{\beta\left(1-q^2\right)}{2} + O\left(\beta^2\right), \\ b(q) &= 1 + O\left(\sqrt{\beta}\right). \end{aligned}$$

Therefore, combining this with (1) from the proof of Lemma 1, we have that

$$t_\phi\left(G_l, T_n^e(q)\right) = \left(\frac{1-q^2}{2}\right)^{r_\phi} q^{y_\phi}\left(\beta^{\left(r_\phi+\frac{y_\phi}{2}\right)} + O\left(\beta^{\left(r_\phi+\frac{y_\phi}{2}+\frac{1}{2}\right)}\right)\right), \tag{3}$$

for all $\phi \in \Phi(G_l)$. Suppose first that $q = 0$, so that we are counting homomorphisms into S_n^e. In order for (3) to not equal zero, we must have that $y_\phi = 0$, which corresponds precisely to there being no vertex $u \in V(G_l)$ such that $\phi(u) = \frac{1}{2}$. In this case, we can rewrite (3) as

$$t_\phi(G_l, S_n^e) = 2^{-r_\phi}\left(\beta^{r_\phi} + O\left(\beta^{(r_\phi+1)}\right)\right). \tag{4}$$

We remark that since $q = 0$, we have $b(q) = 1 + O(\beta)$ rather than $b(q) = 1 + O(\sqrt{\beta})$, and so our correction term in (4) is an improvement over that in (3). Among all $\phi \in \Phi(G_l)$ such that $y_\phi = 0$, we have that $r_\phi \geqslant v - \alpha(G_l)$, and equality occurs if and only if $\sum_{u\in V(G_l)} \phi(u) = \alpha(G_l)$. Let C_2' be the number of $\phi \in \Phi(G_l)$ such that $y_\phi = 0$ and $\sum_{u\in V(G_l)} \phi(u) = \alpha(G_l)$. Then, combining (4) with (2) from the proof of Lemma 1, we have that

$$t(G_l, S_n^e) = 2^{\alpha(G_l)-v} C_2' \left(\beta^{v-\alpha(G_l)} + O\left(\beta^{v-\alpha(G_l)+1}\right)\right),$$

completing the proof of the third part of the lemma.

To prove the second part of the lemma, we proceed in a similar fashion. Fix $q \in (0, 1)$. For all q in this range, we have that $y(q), r(q), b(q) > 0$. For all $\phi \in \Phi(G_l)$ we have that $r_\phi + \frac{y_\phi}{2} \geqslant v - \alpha^*(G_l)$, and equality occurs if and only if $\sum_{u\in V(G_l)} \phi(u) = \alpha^*(G_l)$. Thus for a suitable constant C_1, as in the previous case, we have that

$$t\left(G_l, T_n^e(q)\right) = C_1\left(\beta^{v-\alpha^*(G_l)} + O\left(\beta^{v-\alpha^*(G_l)+\frac{1}{2}}\right)\right),$$

as required. □

Corollary 3. *Let $\alpha = \alpha(G_l)$ and $\alpha^* = \alpha^*(G_l)$. For each $0 \leqslant i \leqslant v - \alpha_*$, let $\tilde{C}_i$ be the number of $\phi \in \Phi(G_l)$ such that $\sum_{u \in V(G_l)} \phi(u) = \alpha^*$ and $r_\phi = i$. Let $A(G_l)$ be the number of independent sets X in G_l such that $|X| = \alpha$. Then*

$$C_1 = \sum_{i=0}^{v-\alpha_*} \tilde{C}_i \left(\frac{1-q^2}{2}\right)^i q^{2(v-\alpha^*-i)},$$

and

$$C_2 = 2^{\alpha - v} A(G_l).$$

Proof. The calculation for C_1 follows directly from the proof of Lemma 2, by taking care to calculate the "suitable constant" mentioned at the end of the proof.

To calculate C_2, we first note that if $\phi \in \Phi(G_l)$ is such that $y_\phi = 0$ and $\sum_{u \in V(G_l)} \phi(u) = \alpha$, then B_ϕ is an independent set in G_l, and $b_\phi = \alpha(G)$. On the other hand, given an independent set $X \in V(G_l)$, the weighting $\phi : V(G_l) \to \{0, 1\}$, given by $\phi(u) = 1$ if and only if $u \in X$, is a fractional independence weighting of G_l with $y_\phi = 0$ and $\sum_{u \in V(G_l)} \phi(u) = |X|$. Thus C_2', as defined in the proof of the theorem, is equal to $A(G_l)$, and so the second part of the corollary follows. □

Theorem 1 follows immediately from Lemma 2. Indeed, if $\alpha^*(G) > \max\left(\alpha(G), \frac{v}{2}\right)$, and $q \in (0, 1)$, then both $t(G_l, K_n^e) = o(t(G_l, T_n^e(q)))$ and $t(G_l, S_n^e) = o(t(G_l, T_n^e(q)))$ as $\beta \to 0$.

Another consequence of Lemma 2 is that, as $\beta \to 0$, we have that $t(G_l, K_n^e) = o(t(G_l, S_n^e))$ if $\alpha(G) > v/2$ and $t(G_l, S_n^e) = o(t(G_l, K_n^e))$ if $\alpha(G) < v/2$. When $\alpha(G) = v/2$, we may apply the following result, which was proved independently by many people (see Cutler and Radcliffe [4] for references and a short proof).

Theorem 2. *[4] If G is a graph with n vertices, $\alpha(G) \leq l$ and $0 \leq k \leq n$, then*

$$i_k(G) \leq i_k(K_{n_1} \cup K_{n_2} \cup \cdots \cup K_{n_l}),$$

where $\sum n_i = n, n_1 \leq n_2 \leq \cdots \leq n_l \leq n_1 + 1$, and $i_k(G)$ denotes the number of independent sets of size k in G.

Taking $l = k = v/2$, we see that $i_k(G) \leq i_k(K_2 \cup K_2 \cup \cdots \cup K_2) = 2^{v/2}$, so that, in our notation, $A(G_l) \leq 2^{v/2}$ when $\alpha(G) = v/2$. Together with Lemma 2 and Corollary 3, this implies the following.

Theorem 3. *With asymptotic notation as $\beta \to 0$,*

$$\max\{t(G_l, S_n^e), t(G_l, K_n^e)\} \sim t(G_l, K_n^e) \text{ if and only if } \alpha(G) \leq v/2.$$

3.1 An explicit counterexample

Throughout this subsection, G_l will be the (labelled) graph with $V(G_l) = [6]$ and

$$E(G_l) = \{\{1,2\},\{1,3\},\{2,3\},\{3,4\},\{4,5\},\{4,6\}\},$$

as in Figure 2, and T_n^e will be the graph $T_n^e(1/\sqrt{2})$. Let G_6 be an unlabelled copy of G_l. We will show that, for $\beta \in (0, 0.016)$, the graph T_n^e has many more copies of G_6 than either K_n^e or S_n^e.

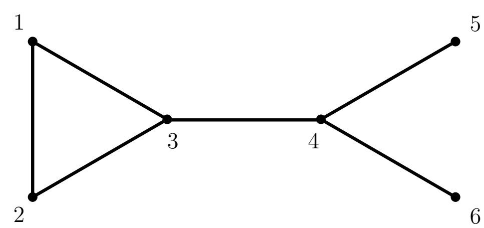


Figure 2: The labelled graph G_l; the unlabelled version is G_6.

Theorem 4. *For $\beta \in (0, 0.016)$, we have $t(G_l, T_n^e) > t(G_l, K_n^e) > t(G_l, S_n^e)$.*

Proof. First, as in the proof of Lemma 2, it is easy to see that $t(G_l, K_n^e) = \beta^3$. We next turn to calculating $t(G_l, S_n^e)$. Recall that $S_n^e = T_n^e(0)$, and that $y(0) = 0, r(0) = 1 - \sqrt{1-\beta}$ and $b(0) = \sqrt{1-\beta}$. By Lemma 1, we have that

$$t(G_l, S_n^e) = \sum_{\phi \in \Phi'(G_l)} r(0)^{r_\phi} b(0)^{b_\phi}, \tag{5}$$

where $\Phi'(G_l)$ is the set of fractional weightings of G in which every vertex receives weight 0 or 1. Given such a fractional weighting ϕ, let c_ϕ be a colouring of the vertices of G where vertices such that $\phi(u) = 0$ are coloured red, and vertices with $\phi(u) = 1$ are coloured blue. In Figure 3 below, we classify the elements of $\Phi'(G_l)$ by the number of blue vertices in their corresponding colourings. Note that no such colouring can have more than $\alpha(G_l) = 3$ blue vertices.
From Figure 3 and (5) we see that

$$t(G_l, S_n^e) = r^6 + 6r^5 b + 9r^4 b^2 + 3r^3 b^3 = r^3\left(6r^2 b + 9rb^2 + 3b^3\right), \tag{6}$$

where $r = r(0)$ and $b = b(0)$.

No. of blue vertices in c_ϕ	No. of $\phi \in \Phi'(G_l)$
0	1
1	6
2	9
3	3

Figure 3: The number of $\phi \in \Phi'(G_l)$ whose colouring has a given number of blue vertices.

To calculate $t(G_l, T_n^e)$, we now let

$$y = y\left(\frac{1}{\sqrt{2}}\right) = \sqrt{\frac{\beta}{2}},$$
$$r = r\left(\frac{1}{\sqrt{2}}\right) = 1 - \sqrt{1 - \frac{\beta}{2}},$$
$$b = b\left(\frac{1}{\sqrt{2}}\right) = \sqrt{1 - \frac{\beta}{2}} - \sqrt{\frac{\beta}{2}}.$$

In a similar fashion to the above, we need to classify all $\phi \in \Phi(G_l)$. Given $\phi \in \Phi(G_l)$, let c_ϕ be a colouring of the vertices of G where vertices such that $\phi(u) = \frac{1}{2}$ are coloured yellow, vertices such that $\phi(u) = 0$ are coloured red, and vertices with $\phi(u) = 1$ are coloured blue. Again, as above, one can list all $\phi \in \Phi(G_l)$ by keeping track of how many red and blue vertices the colouring c_ϕ has. We omit the table listing these colourings, but as before we see that

$$\begin{aligned} t(G_l, T_n^e) &= (y+r)^6 + 2(y+r)^3 r^2 b + 2(y+r)^2 r^3 b + 2(y+r)^4 rb \\ &+ 2r^4 b^2 + 6(y+r) r^3 b^2 + (y+r)^3 rb^2 + 3r^3 b^3. \end{aligned} \tag{7}$$

Numerically, we find that the theorem holds for $\beta \in [0, 0.016]$; see Figure 4 for a graph of the the functions $t(G_l, T_n^e), t(G_l, K_n^e)$ and $t(G_l, S_n^e)$ over this interval. We remind the reader that $e = \frac{\beta n^2}{2}$, and also that the quantities r and b are different in equations (6) and (7). Namely they are $r(0)$ and $b(0)$, and $r(1/\sqrt{2})$ and $b(1/\sqrt{2})$ respectively.

□

Note that we do not claim that T_n^e is the maximiser for the graph G_6. We have only shown that there exists β such that the maximiser for G_6 at edge density β is neither K_n^e nor S_n^e. Nonetheless, we do believe that, for all graphs G, and for all edge densities $\beta \in [0, 1]$, some graph family $H_n = T_n^e(q)$ is the maximiser. We refer the reader to Section 5 for further details on this.

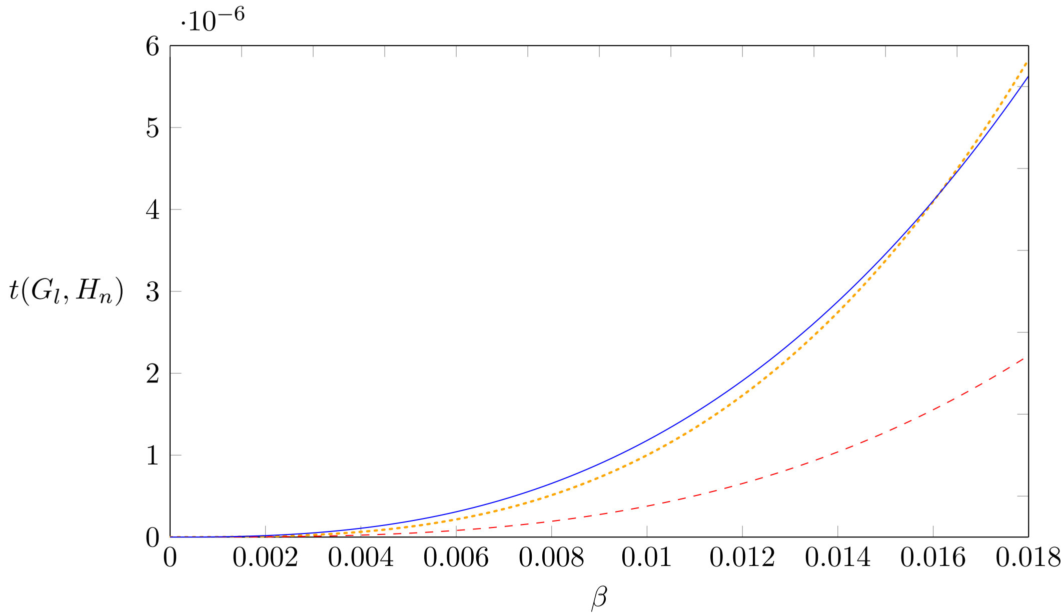


Figure 4: A graph comparing the functions $t(G_l, K_n^e)$, $t(G_l, S_n^e)$ and $t(G_l, T_n^e)$ on the interval $\beta \in [0, 0.018]$. The blue solid line is the function $t(G_l, T_n^e)$, the yellow dotted line is the function $t(G_l, K_n^e)$, and the red dashed line is the function $t(G_l, S_n^e)$. We have that $t(G_l, K_n^e) = t(G_l, T_n^e)$ when $\beta \approx 0.01613474$.

4 The random connection

With some effort, a counterexample to Nagy's conjecture can also be read out of some previous results of Janson, Oleszkiewicz and Ruciński [8]. (We discovered this paper only after we had proved Theorem 1.) As part of their celebrated study on the upper tail for subgraph counts in random graphs, Janson, Oleszkiewicz and Ruciński proved the following (in our notation).

Theorem 5. *Let G be a graph on v vertices with fractional independence number $\alpha^*(G)$. Then, with $\beta = 2e/n^2$,*

$$N(n, e, G) = \Theta(n^v \beta^{v-\alpha^*(G)}).$$

Since it is easy to see that $N(G, K_n^e) = \Theta(n^v \beta^{v/2})$ and $N(G, S_n^e) = \Theta(n^v \beta^{v-\alpha(G)})$, Theorem 5 by itself shows that neither the quasi-clique nor the quasi-star asymptotically maximises $N(G, H)$, at sufficiently small edge density β, if $\alpha^*(G) > \max(\alpha(G), v/2)$. To disprove Nagy's conjecture, one only has to exhibit a single graph satisfying the last condition (for instance, G_6 - there is no such graph on five or fewer vertices).

It is worth describing the lower bound construction in [8], and its relationship to the graphs $T_n^e(q)$. (The construction in [8] is expressed in terms of the solution to a linear program; we rephrase it in our notation.) Given a graph G on v vertices with fractional independence number $\alpha^*(G)$, let ϕ be a weighting of $V(G)$ realizing $\alpha^*(G)$. As noted in Section 2, we may assume that ϕ takes values in $\{0, \frac{1}{2}, 1\}$. Let c be a sufficiently small constant. For each vertex $u \in V(G)$, we "blow up" u to an independent set B_u of size $cn\beta^{1-\phi(u)}$. For each edge $\{u, w\} \in E(G)$, we put a complete bipartite graph (containing $c^2n^2\beta^{2-\phi(u)-\phi(v)} \leq c^2n^2\beta$ edges) between B_u and B_v. Call the resulting graph H. For sufficiently small c, the graph H has at most n vertices and at most $|E(G)|c^2n^2\beta \leq \beta n^2/2$ edges. Moreover, H contains $\prod_u cn\beta^{1-\phi(u)} = c^v n^v \beta^{v-\alpha^*(G)}$ copies of G.

Janson, Oleszkiewicz and Ruciński made no attempt to optimise the constant, and indeed it is not hard to see that one can improve on their construction by making $H[B_u]$ a clique whenever $\phi(u) \neq 1$, and adjusting the sizes of the B_u. In other words, we amalgamate those B_u for which $\phi(u)$ is constant, to get just three sets B_0, B_1 and $B_{1/2}$, "fill in" B_0 and $B_{1/2}$ with cliques, and impose the conditions $|H| = n$ and $|E(H)| = e$ while keeping $|B_i| = \Theta(n\beta^{1-i})$. The result of doing this is just the graph $T_n^e(q)$; the sets $B_0, B_{1/2}$ and B_1 are just R_T, Y_T and B_T respectively. Thus, for the lower bound, one only needs to consider a one-parameter family $T_n^e(q)$, instead of a separate construction for each G, and, moreover, this family $T_n^e(q)$ simply consists of graphons with at most three "steps". We conjecture in the next section that some $T_n^e(q)$ is always asymptotically optimal.

For completeness, we sketch the proof of the upper bound from [8]. To do this, we first re-examine the lower bound construction, where each vertex u in the small graph G is "blown up" to an independent set B_u of size cn^{x_u}, for some $0 \leq x_u \leq 1$, and in which the sought-after copies of G are *compatible* with the partition $(B_u)_{u \in V(G)}$, i.e., we only look for copies of G where each $u \in V(G)$ is located in B_u. Now, a simple random argument [6, 8] shows that, with G fixed and $|G| = v$, any large graph H has a vertex partition $V(H) = \bigcup_{u \in V(G)} B_u$ in which at least $v^{-v}N(G, H)$ of the $N(G, H)$ copies of G in H are compatible (in the above sense) with the partition. So it is enough to show, given a graph H, together with a partition of $V(H)$ into $v = |V(G)|$ parts, labelled with the vertices of G, that H contains at most $\Theta(n^v\beta^{v-\alpha^*(G)})$ *compatible* copies of G.

Fixing G, and given a partition of $V(H)$, we now aim to choose the edges of H so as to maximise the number $N^c(G, H)$ of compatible copies of G in H. Clearly, if there is no edge from u to w in G, we should not put any edges between B_u and B_w in H. For edges $\{u, w\}$ of G, if we make $E(B_u, B_w)$ a complete bipartite graph, we have exactly the lower bound construction. The question remains: can we increase $N^c(G, H)$ by increasing the sizes of the parts B_u, while thinning out the edge sets $E(B_u, B_w)$ for $\{u, w\} \in E(G)$? It turns out that the answer is no.

To see this, we again revisit the lower bound construction, in which the parts B_u have sizes

cn^{x_u}, where the vertex weights x_u comprise a solution to the following linear program.

$$\textbf{Maximise} \quad \sum_u x_u \quad \textbf{subject to} \quad 0 \le x_u \le 1 \text{ and } \; uw \in E(G) \Rightarrow x_u + x_v \le 2 - \epsilon. \tag{8}$$

Here, $\epsilon = -\log(\beta/2)/\log n$, so that $\beta/2 = n^{-\epsilon}$. Given a weighting ϕ of $V(G)$ realizing $\alpha^*(G)$, a solution to (8) can be obtained by setting

$$x_u = 1 - \epsilon(1 - \phi(u)). \tag{9}$$

The dual program is to find nonnegative edge weights y_{uw} and vertex weights z_u of G as below.

$$\textbf{Minimise} \quad \sum_u z_u + (2-\epsilon) \sum_{uw \in E(G)} y_{uw} \quad \textbf{subject to} \quad u \in V(G) \Rightarrow z_u + \sum_{uw \in E(G)} y_{uw} \ge 1. \tag{10}$$

By linear programming duality, the minimum in (10) is exactly the maximum in (8), and, by (9), this maximum is just $v - \epsilon(v - \alpha^*(G))$ (yielding the lower bound $N^c(G, H) \ge Cn^v \beta^{v - \alpha^*(G)}$).

Now each compatible copy of G in H may be considered as a v-vertex hyperedge on $V(H)$; together these form the hypergraph $\mathcal{H}$, whose edges correspond to compatible copies of G in H. Rationalizing a solution to (10) by $a_{uw} = \lceil M y_{uw} \rceil$ and $b_u = \lceil M z_u \rceil$, where M is a large positive integer, we form a sequence of subsets of $V(H)$ by taking each B_u b_u times and each $B_u \cup B_w$ a_{uw} times. By construction, and (10), each vertex in each B_u is covered by at least M of these subsets. For a subset $V' \subset V(H)$, write $\mathrm{Tr}(\mathcal{H}, V') = \{h \cap V' : h \in \mathcal{H}\}$ for the trace of $\mathcal{H}$ on V'. Then, by Shearer's lemma [3],

$$\begin{aligned} N^c(G, H) = |\mathcal{H}| &\le \left(\prod_{u \in V(G)} |\mathrm{Tr}(\mathcal{H}, B_u)|^{b_u} \prod_{uw \in E(G)} |\mathrm{Tr}(\mathcal{H}, B_u \cup B_w)|^{a_{uw}} \right)^{1/M} \\ &\le \left(\prod_{u \in V(G)} n^{b_u} \prod_{uw \in E(G)} n^{(2-\epsilon)a_{uw}} \right)^{1/M} \to \prod_{u \in V(G)} n^{z_u} \prod_{uw \in E(G)} n^{(2-\epsilon) y_{uw}} \\ &= n^v (\beta/2)^{v - \alpha^*(G)}, \end{aligned}$$

as $M \to \infty$, where, in the last line, we have used the duality theorem of linear programming. This is the sought-after upper bound.

We mention for completeness that Janson, Oleszkiewicz and Ruciński's results were generalised to hypergraphs by Dudek, Polcyn and Ruciński [5].

5 Conjectures

In this section, we make some new conjectures about the asymptotic value of $\mathrm{ex}(n, e, G)$. These are essentially the simplest modifications of Nagy's conjecture which fit the known data.

First we define the concept of an *upper profile boundary*. For a fixed labelled graph G_l on v vertices, we look at the number of homomorphisms $\hom(G_l, H)$ from G_l to H, where H ranges over the set of all unlabelled graphs. To each graph H with n vertices and e edges, we associate the point

$$p(G_l, H) = (2e/n^2, \hom(G_l, H)/n^v) \in [0,1]^2,$$

whose x-coordinate is the edge density of H, and whose y-coordinate is the *homomorphism density* of G_l in H. In this way, each labelled graph G_l gives rise to a *profile* $P(G_l) \in [0,1]^2$, defined as the closure of the set of all the points $p(G_l, H)$. The *upper profile boundary* of G_l is the upper boundary of $P(G_l)$; it is not hard to see that this boundary is the graph of a function $f(G_l, \beta)$ of the edge density β. See [9] (page 28) for a picture of the profile of K_3.

We return to the graphs $T_n^e(q)$. For a given graph G_l, and a given edge density β (of H), define $f_T(G_l, \beta)$ by the formula

$$f_T(G_l, \beta) = \sup_{q \in [0,1]} t(G_l, T_n^e(q)),$$

where $e = \beta n^2/2$, and let $q(G_l, \beta) \in [0,1]$ be the value of q at which the supremum is attained. In other words, the function $f_T(G_l, \beta)$ is the normalised asymptotic number of copies of G_l in the optimised T-graph. Now we are ready to state our conjectures.

Conjecture 1. For all graphs G_l and all $\beta \in [0,1]$, we have that $f(G_l, \beta) = f_T(G_l, \beta)$. In other words, for all graphs G_l and all edge densities β, some graph family $H_n = T_n^e(q)$ asymptotically maximises $\hom(G_l, H_n)$ and $N_l(G_l, H_n)$.

Conjecture 2. For each graph G_l, we have that $q(G_l, \beta)$ is an increasing function of β.

A slightly stronger version of Conjecture 1 can be most clearly stated in terms of the "STK notation". Indeed, it appears that, for each graph G_l, there is a partition of the set $[0,1]$ of edge densities into three sets S, T and K (we suppress the dependence on G_l) such that for $\beta \in S$, the quasi-star (asymptotically) maximises $N(G_l, H)$, for $\beta \in T$ some graph $T_n^e(t)$ (with $t \in (0,1)$) maximises $N(G_l, H)$, and for $\beta \in K$, the quasi-clique K_n^e maximises $N(G_l, H)$. If in addition Conjecture 2 holds, these partitions have a particularly simple form. Indeed, in keeping with the theorem of Reiher and Wagner [14], only four possibilities can arise:

Type **K**: $K = [0,1]$,
Type **SK**: $S = [0, \gamma]$ and $K = [\gamma, 1]$, for some $\gamma \in (0,1)$,
Type **TK**: $T = [0, \gamma]$ and $K = [\gamma, 1]$, for some $\gamma \in (0,1)$,
Type **STK**: $S = [0, \gamma], T = [\gamma, \delta]$ and $K = [\delta, 1]$, for some $0 < \gamma < \delta < 1$.

With this notation, Alon [2] characterised graphs of type **K**, Ahlswede and Katona [1] proved that P_2 is type **SK**, Nagy [11] proved that P_4 is type **SK** and conjectured that all graphs are either type **K** or **SK**, and Reiher and Wagner [14] proved that stars are type **SK**, enabling Gerbner, Nagy, Patkós and Vizer [7] to prove that the type always ends in **–K**. In contrast, the results of Janson, Oleszkiewicz and Ruciński only have a bearing on the *start* of the type; for instance, the type of G_6 cannot begin with either **S–** or **K–**, and we conjecture that it is in fact **TK**. We remark that we are unaware of any graphs of type **STK**, and would be very interested to know whether or not such graphs exist. See Figure 5 in the Appendix for a summary of the various types of all connected[1] graphs on at most 5 vertices.

We conclude our paper with a weaker version of Conjecture 1, concerning the behaviour of f and f_T as $\beta \to 0$:

Conjecture 3. As $\beta \to 0$, we have that $f(G_l, \beta) \sim f_T(G_l, \beta)$ for all graphs G_l.

6 Acknowledgements

The work of both authors was supported by STINT foundation grant IB2017-7360. Research of the first author was funded by a grant from the Swedish Research Council, whose support is also gratefully acknowledged.

References

[1] R. Ahlswede and G. Katona, Graphs with maximal number of adjacent pairs of edges, *Acta Math. Acad. Sci. Hungar.* **32** (1978), 97–120.

[2] N. Alon, On the number of subgraphs of prescribed type of graphs with a given number of edges, *Israel Journal of Mathematics* **38** (1981), 116–130.

[3] F. Chung, R. Graham, P. Frankl and J. Shearer, Some intersection theorems for ordered sets and graphs, *J. Combin. Theory Ser. A* **43** (1986), 23–37.

[4] J. Cutler and J. Radcliffe, Extremal problems for independent set enumeration, *Electronic J. of Combin.* **18** (2011), Paper 169.

[5] A. Dudek, J. Polcyn and A. Ruciński, Subhypergraph counts in extremal and random hypergraphs and the fractional q-independence, *Journal of Combinatorial Optimization* **19** (2010), 184–199.

[6] E. Friedgut and J. Kahn, On the number of copies of one hypergraph in another, *Israel Journal of Mathematics* **105** (1998), 251–256.

[1] Any graph on 5 or fewer vertices that is not connected must contain either an isolated vertex or an isolated edge, and so can be reduced to a smaller graph.

[7] D. Gerbner, D. Nagy, B. Patkós and M. Vizer, On the maximum number of copies of H in graphs with given size and order, arXiv:1810.00817v1.

[8] S. Janson, K. Oleszkiewicz and A. Ruciński, Upper tails for subgraph counts in random graphs, *Israel Journal of Mathematics* **142** (2004), 61–92.

[9] L. Lovász, Large Networks and Graph Limits, AMS Colloquium Publications, 2012.

[10] R. Kenyon, C. Radin, K. Ren and L. Sadun, Multipodal structure and phase transitions in large constrained graphs, *J. Stat. Phys.* **168** (2017), 233–258.

[11] D. Nagy, On the number of 4-edge paths in graphs with given edge density, *Combin. Probab. Comput.* **26** (2017), 431–447.

[12] G. Nemhauser and L. Trotter, Properties of vertex packing and independence system polyhedra, *Mathematical Programming* **6** (1974), 48–61.

[13] C. Reiher, The clique density theorem, *Annals of Mathematics* **184** (2016), 683–707.

[14] C. Reiher and S. Wagner, Maximum star densities, *Studia Sci. Math. Hungar.* **55** (2018), 238–259.

[15] A. Sidorenko, A correlation inequality for bipartite graphs, *Graphs and Combinatorics* **9** (1993), 201–204.

Appendix

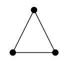
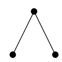
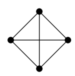
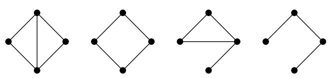
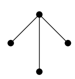
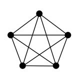
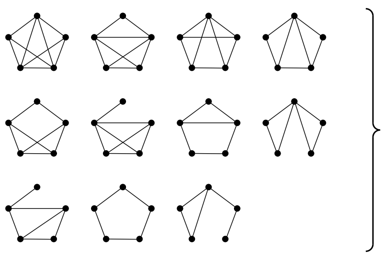
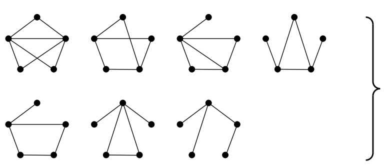
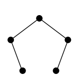
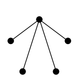
Connected graphs on 3 vertices	$\alpha(G)$	$\alpha^*(G)$	Type	Proof
	1	$\frac{3}{2}$	**K**	• Alon, (Kruskal, Katona)
	2	2	**SK**	• Ahlswede, Katona
Connected graphs on 4 vertices				
	1	2	**K**	• Alon, (Kruskal, Katona)
	2	2	**K**	• Alon
	3	3	**SK**	• Kenyon, Radin, Ren, Sadun • Reiher, Wagner
Connected graphs on 5 vertices				
	1	$\frac{5}{2}$	**K**	• Alon, (Kruskal, Katona)
	2	$\frac{5}{2}$	**K**	• Alon
	3	3	**SK or STK**	• Lemma 2 + Corollary 3 (assuming Conjectures 1 and 2)
	3	3	**SK**	• Nagy
	4	4	**SK**	• Kenyon, Radin, Ren, Sadun • Reiher, Wagner

Figure 5: All the known types of connected graphs on 5 or fewer vertices. The types of all but 7 of these graphs are known based on the results of the listed authors. If one assumes that Conjectures 1 and 2 are true, then the remaining 7 graphs must be of type **SK** or type **STK**.