Maximising the Number of Cycles in Graphs with Forbidden Subgraphs

Natasha Morrison* Alexander Roberts† Alex Scott†‡

March 20, 2020

Abstract

Fix $k \geq 2$ and let H be a graph with $\chi(H) = k+1$ containing a critical edge. We show that for sufficiently large n, the unique n-vertex H-free graph containing the maximum number of cycles is $T_k(n)$. This resolves both a question and a conjecture of Arman, Gunderson and Tsaturian [4].

1 Introduction

For a graph G, let $c(G)$ be the number of cycles in G. The problem of bounding $c(G)$ for various classes of graph has a long history: for example, an upper bound on $c(G)$ in terms of the cyclomatic number of G was given by Ahrens [1] in 1897; while a lower bound is implicit in work of Kirchhoff [19] from fifty years earlier.

For graphs on n vertices, the number of cycles is clearly maximized by the complete graph, which has $\sum_{i=3}^{n}(i!/2i)\binom{n}{i}$ cycles. But what happens if we constrain the structure of G by forbidding some subgraph? In other words, what is the maximal number of cycles in an H-free graph on n vertices (here a graph is *H-free* if it does not contain a subgraph isomorphic to H)? For graphs G and H, let $c(G)$ be the number of cycles in G and let

$$m(n;H) := \max\{c(G) : |V(G)| = n, H \not\subseteq G\}.$$

The problem of determining $m(n,H)$ was introduced by Durocher, Gunderson, Li and Skala [9] (who studied $m(n,K_3)$) and will be the focus of this paper.

*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, United Kingdom and Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brasil. Research partially supported by CNPq and Sidney Sussex College, Cambridge.
E-mail: morrison@dpmms.cam.ac.uk.
†Mathematical Institute, University of Oxford, Woodstock Road, Oxford, United Kingdom.
E-mail: {robertsa, scott}@maths.ox.ac.uk.
‡Supported by a Leverhulme Trust Research Fellowship.

The problem of maximizing the number of *edges* in an H-free graph has been extensively studied. Indeed, Turán [23] proved that the unique n-vertex K_{k+1}-free graph with the maximum number of edges is the complete k-partite graph with all classes of size $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$, which is known as the *Turán graph* $T_k(n)$. More generally, the classical Turán problem asks for the maximum number of edges in an H-free graph: this is the *extremal number* $\mathrm{ex}(n; H)$ and the *extremal graphs* are $\mathrm{EX}(n; H) = \{G : |V(G)| = n, H \not\subseteq G\}$, that is the H-free graphs on n vertices with $\mathrm{ex}(n; H)$ edges. For further detail, we refer to [7].

Much less is known about maximizing the number of *cycles* in H-free graphs. Durocher, Gunderson, Li and Skala [9] investigated $m(n, K_3)$, and conjectured that the maximum is attained by the Turán graph $T_2(n)$. This conjecture was proved for large n by Arman, Gunderson and Tsaturian [4], who showed that, for $n \geq 141$, $T_2(n)$ is the unique triangle-free graph containing $m(n; K_3)$ cycles. They made the following natural further conjecture.

Conjecture 1.1 (Arman, Gunderson and Tsaturian [4])**.** *For any $k > 1$, for sufficiently large n, $T_2(n)$ is the unique n-vertex C_{2k+1}-free graph containing $m(n; C_{2k+1})$ cycles.*

A partial result towards this conjecture is given in [4], where it is shown that $m(n; C_{2k+1}) = O(c(T_2(n)))$. They also ask about a different generalisation.

Question 1.2 (Arman, Gunderson and Tsaturian [4])**.** For $k \geq 4$, what is $m(n; K_k)$? Is $T_{k-1}(n)$ the K_k-free graph containing $m(n; K_k)$ cycles?

In this paper we prove Conjecture 1.1 for any fixed k and sufficiently large n and answer Question 1.2 affirmatively for sufficiently large n. In fact we prove a much more general result. In what follows we say that an edge e of a graph H is *critical* if $\chi(H \backslash \{e\}) = \chi(H) - 1$. Our main result is the following.

Theorem 1.3. *Let $k \geq 2$ and let H be a graph with $\chi(H) = k+1$ containing a critical edge. Then for sufficiently large n, the unique n-vertex H-free graph containing the maximum number of cycles is the Turán graph $T_k(n)$.*

The condition that H has a critical edge is necessary, since if H does not have a critical edge we can add an edge to the relevant Turán graph without creating a copy of H (and the addition of this edge will increase the number of cycles). Conjecture 1.1 follows from Theorem 1.3 as an odd cycle contains a critical edge.

By using the same techniques as in the proof of Theorem 1.3, we are able to obtain a bound on the number of cycles in an H-free graph for any fixed graph H (not just critical ones).

Theorem 1.4. *Let $k \geq 2$ and let H be a fixed graph with $\chi(H) = k + 1$. Then*

$$m(n; H) \leq \left(\frac{k-1}{k}\right)^n n^n e^{-(1-o(1))n}.$$

The Turán graph gives a lower bound showing that this bound is tight up to the $o(1)$ term in the exponent.

In this paper we concern ourselves with maximising cycles of any length in a graph with a forbidden subgraph. The related problem of maximising copies of a single graph in a graph with a collection of forbidden subgraphs has received a great deal of attention. For a graph G and family of graphs $\mathcal{F}$, define $\mathrm{ex}(n, G, \mathcal{F})$ to be the maximum possible number of copies of G in a graph containing no member of $\mathcal{F}$. The value of $\mathrm{ex}(n, G, \mathcal{F})$ is of particular interest when the graphs being studied are cycles (see [2, 6, 10] for results concerning other graphs). Improving on earlier work of Bollobás and Győri [8] and Győri and Li [16], Alon and Shikhelman [2] gave bounds for $\mathrm{ex}(n, K_3, C_{2k+1})$, when $k \geq 2$. Using flag algebras, Hatami, Hladký, Král', Norine, and Razborov [17] showed that the unique triangle-free graph with maximum number of copies of C_5 is the balanced blow up of C_5. Also using flag algebras, Grzesik [14] determined $\mathrm{ex}(n, C_5, K_3)$. More recently, Grzesik and Kielak [15] determined $\mathrm{ex}(n, C_{2k+1}, \mathcal{F})$, where $k \geq 3$ and $\mathcal{F}$ is the family of odd cycles of length at most $2k - 1$. They also asymptotically determine $\mathrm{ex}(n, C_{2k+1}, C_{2k-1})$.

The rest of paper is organised as follows. Section 2 contains a number of lemmas about counting cycles in complete k-partite graphs (Lemmas 2.1-2.6). These will be used in Section 4 for the proof of Theorem 1.3. The statements are very natural but our proofs are unfortunately technical, so we defer these to Section 5. In Section 3 we prove Lemma 3.2 and use similar techniques to prove Theorem 1.4. The proof of Theorem 1.3 is completed in Section 4. We conclude the paper in Section 6 with some related problems and open questions. We conclude the current section with a sketch of the proof of Theorem 1.3.

1.1 Outline of Proof

In what follows we fix H to be a graph with $\chi(H) = k + 1$ that contains a critical edge and assume that n is sufficiently large. As usual, for a graph F we will write $e(F) := |E(F)|$ and in the particular case of the Turán graph, we will write $t_k(n) := |E(T_k(n))|$. Let G be an n-vertex H-free graph with $c(G) = m(n; H)$. As $T_k(n)$ is H-free, we have that $m(n; H) \geq c(T_k(n))$. We will suppose that G is not $T_k(n)$ and obtain a contradiction by showing that $c(G) < c(T_k(n))$.

The first step in the proof (Lemma 4.1) is to show that G with $c(G) \geq c(T_k(n))$ contains at least $e(T_k(n)) - O(n \log^2 n)$ edges. In order to prove this, we will need a bound on the number of cycles an n-vertex H-free graph with $m \geq \beta(H) \cdot n$ edges can contain, where β is some constant depending on H. Such a bound is provided by Lemma 3.2.

Given Lemma 4.1, we are able to apply the following stability result from [21].

Theorem 1.5 (Theorem 1.4 [21]). *Let H be a graph with a critical edge and $\chi(H) = k + 1 \geq 3$, and let $f(n) = o(n^2)$ be a function. If G is an H-free graph with n vertices and $e(G) \geq t_k(n) - f(n)$ then G can be made k-partite by deleting $O(n^{-1} f(n)^{3/2})$ edges.*

Since we have $f(n) = O(n \log^2 n)$, this will imply that G is a sublinear number of edges away from being k-partite. We then take a k-partition of G which minimises the number of edges within classes and carefully bound (given that G is not $T_k(n)$) the number of cycles G can contain that do not use edges within classes (Lemma 4.2). We conclude the proof by separately counting the cycles in G that use edges within classes and observing that the total number of cycles in G is not large enough, a contradiction.

2 Counting Cycles in Complete k-partite Graphs

In this section we state some results about the number of cycles in complete k-partite graphs. These are needed in Section 4 for the proof of Theorem 1.3, but may be of independent interest. Despite the simplicity of the statements, the proofs are annoyingly technical, and so we will give them later in Section 5.

The first gives a bound on the number of cycles in $T_k(n)$. In what follows we write $h(G)$ for the number of Hamiltonian cycles in G (a Hamiltonian cycle of a graph is a cycle covering all of the vertices). We also define $c_r(G)$ to be the number of cycles of length r in G.

Lemma 2.1.

$$c_{2\lfloor n/2\rfloor}(T_2(n)) \sim \pi 2^{-n} n^n e^{-n},$$

and for fixed $k \geq 3$,

$$h(T_k(n)) = \Omega\left(\left(\frac{k-1}{k}\right)^n n^{n-\frac{1}{2}} e^{-n}\right).$$

Since $c(G) \geq h(G)$ for all G, if follows that $c(T_k(n)) = \Omega\left(\left(\frac{k-1}{k}\right)^n n^{n-\frac{1}{2}} e^{-n}\right)$. Arman [3, Theorems 5.22 and 5.26] proves similar results here and also provides an upper bound for $c(T_k(n))$.

Lemma 2.2. *Let $k \geq 2$ and G be an n-vertex k-partite graph. Then for any r, $c_r(T_k(n)) \geq c_r(G)$. Furthermore, when $n \geq 5$, $c(T_k(n)) > c(G)$ for any n-vertex k-partite graph G not isomorphic to $T_k(n)$.*

In particular, Lemma 2.2 implies that the Turán graph $T_k(n)$ has the most Hamilton cycles amongst all k-partite graphs on n vertices.

In order to state the next few lemmas we require some more technical definitions. For $\underline{a} = (a_1, \ldots, a_k) \in \mathbb{N}^k$, we define $K_{\underline{a}}$ to be the complete k-partite graph with vertex classes $V_1, \ldots, V_k$, where $|V_i| = a_i$. Let v be some vertex in $V(K_{\underline{a}})$. We define $h_v(j, K_{\underline{a}})$ to be the number of permutations $v_1 \cdots v_n$ of the vertices of $K_{\underline{a}}$, such that $v_1 = v$, $v_2 \in V_j$ and $v_1 \cdots v_n$ is a Hamilton cycle (we count permutations rather than cycles, so that we count a cycle $v_1 \cdots v_n$ with v_2 and v_n from the same vertex class twice). Note that if we count the Hamilton cycles by considering $v_1 \cdots v_n$ with v_1 fixed, by counting the number of cycles visiting each other vertex class first, then each cycle will be counted twice due to the choice of orientation. So for $v \in V_i$, we have

$$h(K_{\underline{a}}) = \frac{1}{2} \sum_{j \neq i} h_v(j, K_{\underline{a}}). \tag{2.1}$$

The next lemma will allow us to count cycles more accurately in complete k-partite graphs that are not balanced.

Lemma 2.3. *Let $k \geq 3$. Let $\underline{b} = (b_1, \ldots, b_k)$, $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ be such that $b_i \geq b_j$ if and only if $c_i \geq c_j$, and that $K_{\underline{b}} \cong T_k(n)$. Denote the vertex classes of $K_{\underline{c}}$ by $V_1, \ldots, V_k$, and vertex classes of $K_{\underline{b}}$ by $V'_1, \ldots, V'_k$. Then if $v \in V_1, w \in V'_1$, then*

$$h_v(2, K_{\underline{c}}) \leq h_w(2, T_k(n)) \prod_{i=1}^{k} e^{\left|\log(\frac{b_i}{c_i})\right|}.$$

We now bound the proportion of Hamilton cycles starting from a fixed vertex that immediately pass through a fixed vertex class. This will be important when we bound the cycles in a non-complete k-partite graph.

Lemma 2.4. *Let $k \geq 3$, and suppose $T_k(n)$ has vertex classes $V_1, \ldots, V_k$ (arbitrarily ordered independently of class size). Then for n sufficiently large, if $v \in V_1$,*

$$h_v(2, T_k(n)) \geq \frac{2}{3k} h(T_k(n)).$$

The next two lemmas give a recursive bound on the number of Hamilton cycles in $T_k(n)$. This will allow us to bound the number of cycles in the Turán graph in terms of the number of Hamilton cycles it contains. Throughout the chapter we will make use of the notation $(n)_i := n \cdot (n-1) \cdots (n-(i-1))$.

Lemma 2.5. *For $k, n \in \mathbb{N}, k \geq 3$ and $i \in [n]$,*

$$h(T_k(n)) \geq (n-1)_i \left(\frac{k-2}{k}\right)^i h(T_k(n-i)).$$

Lemma 2.6. *For $k, n \in \mathbb{N}, k \geq 3$:*

$$c(T_k(n)) \leq e^{\frac{2k}{k-2}} h(T_k(n)).$$

Finally, we have similar results when $k = 2$. This case is slightly different to when $k \geq 3$ as $T_2(n)$ only contains even cycles.

Lemma 2.7. *For $n \in \mathbb{N}$ and $i = o(n)$, we have*

$$c(T_2(n-i)) \leq 2e \left(\frac{4}{n}\right)^i c_{2\lfloor \frac{n}{2} \rfloor}(T_2(n)).$$

3 Counting Cycles in H-free Graphs

Fix H to be a graph with $\chi(H) = k+1$ containing a critical edge. The first aim of this section is to prove a lemma bounding the number of cycles in an n-vertex H-free graph containing a fixed number of edges. We will need the following theorem of Simonovits [22].

Theorem 3.1 (Simonovits [22, Theorem 2.3]). *Let H be a graph with $\chi(H) = k + 1 \geq 3$ that contains a critical edge. Then there exists some n_0 such that, for all $n \geq n_0$, we have* $\mathrm{EX}(n; H) = \{T_k(n)\}$.

Given H, define $n_0'(H)$ to be the smallest value of n_0 such that Theorem 3.1 holds and choose $n_0(H) \geq n_0'(H)$ such that $\mathrm{ex}(n; H) \geq 10n$ for each $n \geq n_0$. We define $\beta(H) := 10n_0$.

In a recent paper, Arman and Tsaturian [5] consider the maximum number of cycles in a graph with a fixed number of edges: They show that if G is an n-vertex graph with m edges, then

$$c(G) \leq \begin{cases} \frac{3}{4}\Delta(G)\left(\frac{m}{n-1}\right)^{n-1} & \text{for } \frac{m}{n-1} \geq 3, \\ \frac{3}{4}\Delta(G) \cdot \left(\sqrt[3]{3}\right)^m, & \text{otherwise.} \end{cases}$$

This general bound is not strong enough for us: comparing this bound with the bounds given in Lemma 2.1, we see that a graph with at least as many cycles as $T_k(n)$ has at least $(1 + o(1))\, e^{-1} t_k(n)$ edges. However under the additional assumption that our graph does not contain a forbidden subgraph H, we are able to prove the following lemma which we will later use to show that an H-free graph with at least as many cycles as $T_k(n)$ has at least $(1 + o(1))\, t_k(n)$ edges. We remark that when m is close to $t_k(n)$, the bound we gives beats the general bound of Arman and Tsaturian by an exponential factor.

Lemma 3.2. *Let H be a fixed graph with $\chi(H) = k + 1 \geq 3$ containing a critical edge. For n sufficiently large, let G be an H-free graph with n vertices and m edges where $t_k(n) - 10n \geq m \geq \beta(H) \cdot n$ (recall the definition of $\beta(H)$ from just after Theorem 3.1). Then $c(G) = O\left(\lambda^n n^{n+2} \left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\lambda} - \lambda n}\right)$, where*

$$\lambda := 1 - \left(1 - \frac{2k}{k-1}\frac{m}{(n-3)^2}\right)^{\frac{1}{2}}. \tag{3.1}$$

The next lemma bounds the maximum number of paths that an H-free graph G can contain between two fixed vertices. For $x, y \in V(G)$, define $p_{x,y}$ to be the number of paths between x and y in G.

Lemma 3.3. *Let H be a graph with $\chi(H) = k + 1 \geq 3$ that contains a critical edge. For n sufficiently large, let G be an H-free graph with n vertices and m edges where $t_k(n) - 10n \geq m \geq \beta(H) \cdot n$ (recall the definition of $\beta(H)$ from just after Theorem 3.1). Then for any $x, y \in V(G)$,*

$$p_{x,y}(G) = O\left(\lambda^n n^n \left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\lambda} - \lambda n}\right),$$

where λ is as defined in (3.1).

Lemma 3.2 follows easily from Lemma 3.3.

Proof of Lemma 3.2. Observe that for each edge $e = xy$ in G, the number of cycles containing e is at most $p_{x,y}$. Thus, by Lemma 3.3

$$\begin{aligned} c(G) &\leq \sum_{xy \in E(G)} p_{x,y}(G) \\ &= O\left(m\lambda^n n^n \left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\lambda} - \lambda n} \right) \\ &= O\left(\lambda^n n^{n+2} \left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\lambda} - \lambda n} \right), \end{aligned}$$

as required. □

Before proving Lemma 3.3, we prove the following Lemma which allows us to consider an integer valued linear optimisation problem to find upper bounds for the number of paths between vertices in graphs with a forbidden subgraph.

Lemma 3.4. *Let H be a graph with $\chi(H) \geq 3$. Let G be an H-free graph with n vertices and m edges, and let x, y be vertices of G. Then $p_{x,y}(G)$ is bounded by the maximum value of the product*

$$\prod_{i=2}^{n} \max\{r_i, 1\} \tag{3.2}$$

under the following set of constraints:

(i) $r_i \in \mathbb{Z}_{\geq 0}$*, for* $2 \leq i \leq n$*,*

(ii) $\sum_{i=2}^{n} r_i \leq m$*, and*

(iii) $\sum_{i=2}^{t} r_i \leq \mathrm{ex}(t; H)$*, for* $2 \leq t \leq n$*.*

Proof of Lemma 3.4. Fix $x, y \in V(G)$. We define a sequence of vertices $(x_i)_{i \in [n]}$ and a sequence of graphs $(G_i)_{i \in [n]}$ as follows. Let $x_1 = x$ and $G_1 = G$. For $i \geq 2$, given x_{i-1} and G_{i-1}, let $G_i = G_{i-1} \setminus x_{i-1}$ and choose x_i with $p_{x_i,y}(G_i)$ as large as possible.

We count the number of paths between x and y by summing over possibilities for the second vertex in a path. We get the following inequality

$$\begin{aligned} p_{x,y}(G) &= \sum_{z \in N(x)} p_{z,y}(G \setminus \{x\}) \\ &\leq d_G(x_1) \cdot \max\{p_{z,y}(G_2) : z \in N(x_1)\} \\ &= d_G(x_1) p_{x_2,y}(G_2). \end{aligned}$$

Repeating this process gives

$$p_{x_1,y}(G) \leq \prod_{i=1}^{\ell} d_{G_i}(x_i),$$

where ℓ is minimal such that $\max\{p_{x_{\ell+1},y}(G_{\ell+1}) : x_{\ell+1} \in N_{G_\ell}(x_\ell)\} = 1$.

For $1 \le i \le \ell$, let $d_i := d_{G_i}(x_i)$. Note that the d_i are positive integers and that $\sum_{i=1}^{\ell} d_i \le m$. Also note that for any $t \in \{1, \dots, \ell\}$, we have

$$\sum_{i=t}^{\ell} d_i \le e(G_t).$$

Therefore, as G_t is an $(n-t+1)$-vertex H-free graph, $\sum_{i=t}^{\ell} d_i \le \mathrm{ex}(n-t+1; H)$. The result follows by letting $r_i = 0$ for $i = 2, \dots, n-\ell$ and $r_i = d_{n+1-i}$ for $i = n+1-\ell, \dots, n$. $\square$

We now prove Lemma 3.3.

Proof of Lemma 3.3. Following on from the proof of Lemma 3.4, we consider a relaxation of the constraints given in the statement of Lemma 3.4. Recall that $n_0 := n_0(H)$ is such that $\mathrm{ex}(s; H) = t_k(s)$ and $\mathrm{ex}(s; H) \ge 10s$ for all $s \ge n_0$. We look to maximise

$$\prod_{i=2}^{n} \max\{r_i, 1\}, \tag{3.3}$$

under the following relaxed constraints:

(a) $r_i \in \mathbb{Z}_{\ge 0}$, for $i > n_0$,

(b) $r_i \in \mathbb{R}_{\ge 0}$, for $i \le n_0$,

(c) $\sum_{i=2}^{n} r_i \le m$, and

(d) $\sum_{i=2}^{t} r_i \le \mathrm{ex}(t; H)$, for each $n_0 \le t \le n$.

Since $m \ge \beta(H)n$, we have $\frac{m}{n} \ge \frac{10t_k(n_0)}{n_0-1}$. Now let $(r_i)_{i=2}^{n}$ be a sequence maximising (3.3) subject to (a)-(d). We may assume that $r_2, \dots, r_{n_0}$ and $r_{n_0+1}, \dots, r_n$ are in increasing order as this will not violate (a)-(d).

Claim 3.5. *There is some $I \in [n_0+1, n-2]$ such that:*

(i) $r_i = \frac{t_k(n_0)}{n_0-1}$, *for* $i \le n_0$,

(ii) $r_i = t_k(i) - t_k(i-1)$, *for* $n_0 + 1 \le i \le I$, *and*

(iii) $r_i \in \{r_I, r_I + 1\}$, *for* $i > I$.

Proof of Claim. Let $T = \sum_{i=2}^{n_0} r_i$. Then $(r_2, \dots, r_{n_0}) = (0, \dots, 0, \frac{T}{S}, \dots, \frac{T}{S})$ for some $S \in [n_0 - 1]$ (or else we can increase $\prod_{i=2}^{n_0} r_i$). We may assume that T is an integer as we can replace T by $\lceil T \rceil$ and still satisfy (a)-(d). Differentiation of the function $j(x) = \left(\frac{T}{x}\right)^x$ shows that if $T \ge en_0$, then $S = n_0 - 1$ and so $r_i = \frac{T}{n_0-1}$ for each $i \in [n_0]$.

Suppose that $T < e \cdot n_0$. Then since $\frac{m}{n} \ge \beta(H) \ge \frac{10t_k(n_0)}{n_0-1}$, there must be a $j > n_0$ such that $r_j \ge \frac{t_k(n_0)}{n_0-1} \ge 10$. Choose j to be minimal with this property. It can easily be verified that increasing r_2 by 2 and decreasing r_j by 2 gives a sequence which satisfies (a)-(d) but gives a larger product. Therefore it must be the case that $T \geq e \cdot n_0$ and so $S = n_0 - 1$.

Now suppose that (i) doesn't hold and so $e \cdot n_0 \leq T < t_k(n_0)$. Since $\frac{m}{n} \geq \frac{10t_k(n_0)}{n_0-1}$, there exists some $j > n_0$ such that $r_j > \frac{5t_k(n_0)}{n_0-1}$. Choose j to be minimal with this property and define $(s_i)_{i=2}^n$ by $s_i = \frac{T+1}{n_0-1}$ for $i \leq n_0$, $s_j = r_j - 1$ and $s_i = r_i$ otherwise. Then $(s_i)_{i=2}^n$ is a sequence satisfying (a)-(d) which gives a larger product, a contradiction. Therefore $T = t_k(n_0)$ and (i) holds.

Now suppose that (ii) does not hold and so $r_{n_0+1} < t_k(n_0+1) - t_k(n_0)$. Since $\frac{m}{n} \geq 2(t_k(n_0+1) - t_k(n_0))$, there must be a $j > n_0$ such that $r_j > t_k(n_0+1) - t_k(n_0)$. Choose j to be minimal with this property and define $(s_i)_{i=2}^n$ by $s_{n_0+1} = r_{n_0+1} + 1$, $s_j = r_j - 1$ and $s_i = r_i$ otherwise. Then $(s_i)_{i=2}^n$ is a sequence satisfying (a)-(d) which gives a larger product, a contradiction. Therefore $r_{n_0+1} = t_k(n_0+1) - t_k(n_0)$ and (ii) holds.

Let $j > n_0 + 1$ be minimal such that $\sum_{i=1}^j r_i \leq t_k(j) - 1$ (such a j must exist since $m < t_k(n)$) and set $I = j - 1$. If (iii) does not hold then there exists some $t \geq j$ such that $r_j + 1 < r_t$. Let t be minimal with this property, and define $s_j := r_j + 1$, $s_t := r_t - 1$, and $s_i := r_i$ for all $i \notin \{j, t\}$. The sequence $(s_i)_{i\in[n]}$ satisfies (a)-(d) but

$$\prod_{i=2}^{n} \max\{r_i, 1\} < \prod_{i=2}^{n} \max\{s_i, 1\},$$

a contradiction. Therefore $(r_i)_{i=1}^n$ satisfies properties (i)-(iii), completing the proof of the claim.

Finally note that $I \leq n-2$ follows from $m \leq t_k(n) - 10n$. $\square$

Putting the values for r_i from the claim into (3.3), we see that

$$\begin{aligned} p_{x,y} &\leq \left(\frac{t_k(n_0)}{n_0-1}\right)^{n_0-1} \prod_{i=n_0+1}^{I} [t_k(i) - t_k(i-1)] \prod_{i=I+1}^{n} r_i \\ &= O\left(\prod_{i=2}^{n} s_i\right), \end{aligned} \tag{3.4}$$

where (s_i) is some sequence such that $s_i = t_k(i) - t_k(i-1)$ for $i \in \{2, \ldots, I\}$, $s_i \in \{s_I, s_I + 1\}$ for $i > I$, and $m = \sum_{i=2}^n s_i$.

Note that $s_i = t_k(i) - t_k(i-1) = (i-1) - \left\lfloor \frac{i-1}{k} \right\rfloor$ for $i \leq I$. Then the sequence $(s_i)_{i=2}^I$ is just the natural numbers up to $I - 1 - \left\lfloor \frac{I-1}{k} \right\rfloor$ with a repetition at each multiple of $k-1$. In other words,

$$\left\{s_i : i \in \{2, \ldots, I\} \setminus \left\{\ell k + 1 : \ell \leq \frac{I-1}{k}\right\}\right\} = \left[I - 1 - \left\lfloor \frac{I-1}{k} \right\rfloor\right]$$

and $s_{\ell k+1} = \ell(k-1)$ for each $\ell \leq \frac{I-1}{k}$. Letting $b = \left\lfloor \frac{I-1}{k} \right\rfloor$ we have

$$\prod_{i=2}^{I} s_i = (s_I)! \prod_{j=1}^{b} j(k-1) = s_I! b! (k-1)^b. \tag{3.5}$$

The remaining $n-I$ elements of the product $\prod_{i=2}^n s_i$ are all at most s_I+1. Therefore, by (3.4) and (3.5) we have

$$\begin{aligned}
p_{x,y} &= O\left(\prod_{i=2}^n s_i\right) \\
&= O\left(s_I!b!(k-1)^b(s_I+1)^{n-I}\right) \\
&= O\left(s_I!b!(k-1)^b s_I^{n-I} e^{\frac{n}{s_I}}\right). \qquad (3.6)
\end{aligned}$$

Applying Stirling's approximation and simplifying, (3.6) yields

$$p_{x,y} = O\left(s_I^{n+s_I+1/2-I} b^{b+1/2}(k-1)^b \exp\left\{\frac{n}{s_I} - I\right\}\right).$$

Since $s_I = I-1-\left\lfloor \frac{I-1}{k}\right\rfloor \geq (k-1)\frac{I-1}{k}$ and $b = \left\lfloor \frac{I-1}{k}\right\rfloor \leq \frac{I-1}{k}$, we have $b \leq \frac{s_I}{k-1}$. Therefore

$$\begin{aligned}
p_{x,y} &= O\left(s_I^{n-b-1/2}\left(\frac{s_I}{k-1}\right)^{b+1/2}(k-1)^b \exp\left\{\frac{n}{s_I} - I\right\}\right) \\
&= O\left(s_I^n \exp\left\{\frac{n}{s_I} - I\right\}\right).
\end{aligned}$$

Note that $s_I \in \left[\frac{k-1}{k}(I-1), \frac{k-1}{k}I\right]$ and so

$$\begin{aligned}
p_{x,y} &= O\left((I-1)^n\left(\frac{k-1}{k}\right)^n\left(1+\frac{1}{I-1}\right)^n \exp\left\{\frac{kn}{(k-1)(I-1)} - (I-1)\right\}\right) \\
&= O\left((I-1)^n\left(\frac{k-1}{k}\right)^n \exp\left\{\frac{(2k-1)n}{(k-1)(I-1)} - (I-1)\right\}\right).
\end{aligned}$$

Substituting $I-1=\alpha n$ gives

$$p_{x,y} = O\left(\alpha^n n^n\left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\alpha}-\alpha n}\right). \qquad (3.7)$$

It remains to determine the value of α. We do this by counting edges. Since $m=\sum_i s_i$, we see that

$$m \geq t_k(I) + s_I\,(n-I)\,. \qquad (3.8)$$

Arguing as for (3.5), we see that

$$\begin{aligned}
t_k(I) &= \sum_{i=1}^{s_I} i + (k-1)\sum_{j=1}^{b} j \\
&= \frac{1}{2}(s_I^2 + s_I + (k-1)(b^2+b)).
\end{aligned}$$

If we put this value for $t_k(I)$ into (3.8) we see that

$$\begin{aligned} m &\geq \frac{1}{2}(s_I^2 + s_I + (k-1)(b^2+b)) + s_I\,(n-(I-1)) - s_I \\ &= \frac{1}{2}\left(s_I^2 - s_I + (k-1)(b^2+b)\right) + s_I\,(n-(I-1))\,. \end{aligned}$$

Now consider that $b = \left\lfloor \frac{I-1}{k} \right\rfloor \geq \frac{I-1}{k} - 1$, so that $b^2 + b \geq \left(\frac{I-1}{k}\right)^2 - \frac{I-1}{k}$. Recall also that $s_I \geq \frac{k-1}{k}(I-1)$ and so

$$\begin{aligned} m \geq{}& \frac{1}{2}\left(\left(\frac{k-1}{k}\right)^2 (I-1)^2 - \frac{k-1}{k}(I-1) + \frac{k-1}{k^2}(I-1)^2 - \frac{k-1}{k}(I-1)\right) \\ &+ \frac{k-1}{k}(I-1)n - \frac{k-1}{k}(I-1)^2 \\ \geq{}& \frac{k-1}{k}n(I-1) - \frac{k-1}{2k}(I-1)^2 - 3\frac{k-1}{k}(I-1). \end{aligned}$$

Substituting $(I-1) = \alpha n$ and rearranging gives

$$\left(\left(1-\frac{3}{n}\right) - \alpha\right)^2 \geq \left(1-\frac{3}{n}\right)^2 - \frac{2k}{k-1}\frac{m}{n^2}.$$

Recall that $I \leq n-2$ and so $\alpha \leq \left(1-\frac{3}{n}\right)$. On the other side of the inequality, $\left(1-\frac{3}{n}\right)^2 - \frac{2k}{k-1}\frac{m}{n^2}$ is positive since $m \leq t_k(n) - 10n$. Therefore we can take square roots and rearrange to get

$$\alpha \leq \left(1-\frac{3}{n}\right) - \left(\left(1-\frac{3}{n}\right)^2 - \frac{2k}{k-1}\frac{m}{n^2}\right)^{\frac{1}{2}} = \left(1-\frac{3}{n}\right)\lambda.$$

Since the expression $\alpha^n n^n \left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\alpha} - \alpha n}$ is increasing in α when $\alpha \leq 1 - \frac{2}{n}$, (3.7) is maximised by setting $\alpha = \left(1-\frac{3}{n}\right)\lambda$. We are then done since

$$\left(1-\frac{3}{n}\right)^n \lambda^n n^n \left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\left(1-\frac{3}{n}\right)\lambda} - \left(1-\frac{3}{n}\right)\lambda n} = O\left(\lambda^n n^n \left(\frac{k-1}{k}\right)^n e^{\frac{2k-1}{(k-1)\lambda} - \lambda n}\right).$$

□

Theorem 1.4 follows easily from the idea of this proof by applying the following theorem of Erdős and Simonovits.

Theorem 3.6 (Erdős and Simonovits [11, Theorem 1])**.** *Let H be a graph with $\chi(H) = k+1$. Then,*

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n;H)}{\binom{n}{2}} = 1 - \frac{1}{k}.$$

Proof of Theorem 1.4. Let $\varepsilon > 0$. By Theorem 3.6 and the fact that $t_k(n) \sim \left(1 - \frac{1}{k}\right)\binom{n}{2}$, we know that for n sufficiently large, $\mathrm{ex}(n;H) \le (1+\varepsilon)t_k(n)$. Thus, for n sufficiently large, $\mathrm{ex}(s;H) \le (1+\varepsilon)t_k(s)$ for all $n^{\frac{1}{2}} \le s \le n$. For ease of notation, let $n_1 := n^{\frac{1}{2}}$.

To bound the number of cycles in the graph, we wish to bound $p_{x,y}(G)$ for $x, y \in V(G)$. From Lemma 3.4, we see that it is enough to bound the product

$$\prod_{i=2}^{n} \max\{r_i, 1\},$$

where (r_i) satisfies the relaxed conditions:

(i) $r_i \in \mathbb{R}^+$, for all i, and

(ii) $\sum_{i=2}^{t} r_i \le (1+\varepsilon)t_k(t)$, for each $n_1 \le t \le n$.

It is easily seen that this expression is maximised when $r_i := \frac{(1+\varepsilon)t_k(n_1)}{n_1 - 1}$ for $i = 2, \ldots, n_1$ and $r_i = (1+\varepsilon)(t_k(i) - t_k(i-1))$ otherwise. Therefore, we arrive at the following bound:

$$\begin{aligned}
\prod_{i=2}^{n} r_i &\le \left(\frac{(1+\varepsilon)t_k(n_1)}{n_1 - 1}\right)^{n_1 - 1} \prod_{i=n_1+1}^{n} (1+\varepsilon)(t_k(i) - t_k(i-1)) \\
&= O\left(e^{n_1} \prod_{i=2}^{n} (1+\varepsilon)(t_k(i) - t_k(i-1))\right) \\
&= O\left(e^{\varepsilon n + n_1} \prod_{i=2}^{n} (t_k(i) - t_k(i-1))\right). \qquad (3.9)
\end{aligned}$$

Recall from (3.5) that, defining $b = \left\lfloor \frac{n-1}{k} \right\rfloor$, we have

$$\prod_{i=2}^{n} (t_k(i) - t_k(i-1)) = (n-1-b)!b!(k-1)^b.$$

Applying Stirling's approximation and simplifying gives

$$\begin{aligned}
\prod_{i=2}^{n} (t_k(i) - t_k(i-1)) &= O\left((n-1-b)^{n-1-b+1/2} b^{b+1/2} e^{-n} (k-1)^b\right) \\
&= O\left(\left(\frac{k-1}{k}\right)^n n^{n+1} e^{-n}\right).
\end{aligned}$$

Putting this into (3.9) gives

$$p_{x,y} = O\left(\left(\frac{k-1}{k}\right)^n n^{n+1} e^{\varepsilon n + n_1 - n}\right). \tag{3.10}$$

Now, as in the proof of Lemma 3.2, we see that by (3.10) and the fact that $n_1 = o(n)$,

$$\begin{aligned} c(G) &\leq \sum_{xy \in E(G)} p_{x,y} \\ &= O\left(n^2 \left(\frac{k-1}{k}\right)^n n^{n+1} e^{\varepsilon n + n_1 - n}\right) \\ &= O\left(\left(\frac{k-1}{k}\right)^n n^n e^{-(1-2\varepsilon)n}\right). \end{aligned}$$

Since ε is arbitrary, we have our result. □

4 Proof of Theorem 1.3

Here we complete the proof of Theorem 1.3. This will follow from the next two lemmas.

The first gives a lower bound on the number of edges in an extremal graph. (See also [3, Theorem 5.3.2] for a K_{k+1} version.)

Lemma 4.1. *Let H be a graph $\chi(H) = k+1 \geq 3$ containing a critical edge. For sufficiently large n, let G be an n-vertex H-free graph with m edges and $c(G) \geq c(T_k(n))$. Then $m \geq \frac{n^2(k-1)}{2k} - O\left(n \log^2(n)\right)$.*

Given this lemma, we can apply Theorem 1.5 to show that any extremal graph G is close to being k-partite. We then carefully count the number of cycles in such a graph. In what follows, for a graph G and a k-partition of its vertices, we call edges within a vertex class *irregular* and those between vertex classes *regular*. Define a *best* k-partition of a graph G to be one which minimises the number of irregular edges contained within G. The next lemma counts the cycles using only regular edges if G is not $T_k(n)$. Recall that $c_r(G)$ is the number of cycles of length r in G.

Lemma 4.2. *Let H be a graph with $\chi(H) = k+1 \geq 3$ containing a critical edge. Suppose $G \ncong T_k(n)$ is an n-vertex H-free graph with $c(G) \geq c(T_k(n))$. Then for sufficiently large n, the number of cycles using only regular edges in the best k-partition of G is at most:*

$$\begin{cases} c(T_k(n)) - \frac{1}{16k} h(T_k(n)) & \text{for } k \geq 3, \\ c(T_2(n)) - \frac{1}{8} c_{2\lfloor \frac{n}{2} \rfloor}(T_2(n)) & \text{for } k = 2. \end{cases}$$

Given Lemmas 4.1 and 4.2, we now complete the proof of Theorem 1.3. We will then prove the lemmas themselves. The main work remaining for Theorem 1.3 is to count the number of cycles using irregular edges.

Proof of Theorem 1.3. Let H be a graph with a critical edge with chromatic number $\chi(H) = k+1 \geq 3$, and suppose G is an n-vertex H-free graph with $c(G) = m(n; H)$. Then, in particular, $c(G) \geq c(T_k(n))$. Suppose for a contradiction that G is not isomorphic to $T_k(n)$. Fix a best k-partition of G: by Lemma 4.1 and Theorem 1.5, we know that for sufficiently large n, the graph G has at most $n^{0.55}$ irregular edges in its best k-partition.

Let $c^I(G)$ be the number of cycles in G containing at least one irregular edge and let $c^R(G)$ be the number of cycles in G using only regular edges. If $c^I(G) = o(h(T_k(n)))$, then by applying Lemma 4.2 and taking n sufficiently large, we have $c(G) = c^R(G) + c^I(G) < c(T_k(n))$. Thus $c^I(G) = \Omega(h(T_k(n)))$.

Let E_I be the set of irregular edges in G. For each non-empty $A \subsetneq E_I$, let C_A be the set of cycles C in G such that $E(C) \cap E_I = A$ and such that C contains at least one regular edge. Fix A such that C_A is non-empty and fix an edge $a_1a_2 \in A$. (Note that A must be a vertex-disjoint union of paths or else it would not be possible to have a cycle using all edges in A.) For any cycle $C = x_1x_2 \cdots x_j$ in C_A, with $x_1 = a_1$ and $x_2 = a_2$, define $S(C)$ to be the directed cycle $x_1x_2 \cdots x_j$ (so for all i, the edge x_ix_{i+1} is directed towards x_{i+1}, where indices are taken modulo j).

For each $C \in C_A$, the orientation of $S(C)$ induces an orientation f_C on the edges of A. Given a fixed orientation f of A, we write

$$C_A(f) := \{C \in C_A : f_C = f\}.$$

We will bound the size of each $C_A(f)$. A bound on $c^I(G)$ will then follow by summing over all possible A and f.

Let G/A be the graph obtained by contracting every edge in A. Then remove the remaining irregular edges to form J (so J is an H-free k-partite graph with $n - |A|$ vertices, as A is a vertex-disjoint union of paths, and each edge of A lies inside some vertex class of our k-partition). For each cycle C in $C_A(f)$, we obtain an oriented cycle $g(C)$ in J by replacing each maximal path $u_1 \cdots u_j$ in $E(C) \cap A$ oriented from u_1 to u_j by u_1. As C contains at least one regular edge, $g(C)$ is either an edge or cycle in J.

We claim that g is injective on $C_A(f)$. Indeed suppose that there exists a cycle $C \in C_A(f)$. Recall that A is a vertex-disjoint union of paths and furthermore that f orients the paths of A. Denote these oriented paths $(u_i^1)_{i\in[\ell_1]}, \ldots, (u_i^t)_{i\in[\ell_t]}$. Each cycle $C \in C_A(f)$ must contain these oriented paths as segments (each edge of A must be contained in C and it is not possible to break up a path or else a vertex must be adjacent to more than two edges in the cycle). Therefore we have an inverse of g which takes a cycle from $g(C_A(f))$ and replaces each instance of u_1^j with the path $u_1^j \cdots u_{\ell_j}^j$.

As J is a k-partite graph on $n - |A|$ vertices, by Lemma 2.2 we have

$$c(J) \leq c(T_k(n - |A|)).$$

Recall that for each $C \in C_A(f)$, $g(C)$ is either an edge or a cycle in J. We therefore have

$$|C_A(f)| \leq 2 \cdot c(T_k(n - |A|)) + 2|E(T_k(n))| \leq 4 \cdot c(T_k(n - |A|)),$$

for sufficiently large n by applying Lemma 2.1 and recalling that $|A| \leq n^{0.55}$. Let F_A be the set of all possible orientations f of A. We have

$$c^I(G) \leq |E_I|^{|E_I|} + \sum_{A \subseteq E_I} \sum_{f \in F_A} |C_A(f)|, \tag{4.1}$$

where the first term counts cycles that contain only irregular edges and the second term counts cycles in $c^I(G)$ that contain both a regular and irregular edge.

We will bound the second term of this expression. Recalling that there are at most $n^{0.55}$ irregular edges, we get that

$$\sum_{A \subseteq E_I} \sum_{f \in F_A} |C_A(f)| \leq \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} 2^i \cdot 4 \cdot c(T_k(n-i)).$$

For $k \geq 3$, we now apply Lemma 2.6 and Lemma 2.5 for each i in the sum,

$$\begin{aligned}
\sum_{A \subseteq E_I} \sum_{f \in F_A} |C_A(f)| &\leq \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} e^{\frac{2k}{k-2}} 2^{i+2} h(T_k(n-i)) \\
&\leq 4e^{\frac{2k}{k-2}} \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} \left(\frac{2k}{k-2}\right)^i \frac{h(T_k(n))}{(n-1)_i} \\
&\leq 4e^6 h(T_k(n)) \sum_{i \geq 1} n^{0.55i} \left(\frac{6}{n-n^{0.55}}\right)^i \\
&= o\left(h(T_k(n))\right).
\end{aligned}$$

We have $|E_I|^{|E_I|} \leq (n^{0.55})^{n^{0.55}}$ which is $o(h(T_k(n)))$ by Lemma 2.1. Therefore, using (4.1) we see that $c^I(G) = o(h(T_k(n))$, a contradiction. Therefore G is isomorphic to $T_k(n)$.

Similarly for $k = 2$, we apply Lemma 2.7 to get

$$\begin{aligned}
\sum_{A \subseteq E_I} \sum_{f \in F_A} |C_A(f)| &\leq \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} 2^i \cdot 8e \cdot \left(\frac{4}{n}\right)^i c_{2\lfloor n/2 \rfloor}(T_2(n)) \\
&\leq 8e \cdot c_{2\lfloor n/2 \rfloor}(T_2(n)) \sum_{i=1}^{n^{0.55}} n^{0.55i} \left(\frac{8}{n}\right)^i \\
&= o\left(c_{2\lfloor n/2 \rfloor}\left(T_2(n)\right)\right),
\end{aligned}$$

and we conclude as before. □

We now present the proofs of Lemmas 4.1 and 4.2.

Proof of Lemma 4.1. First suppose that $m = O(n)$. We can then crudely bound $p_{x,y}(G)$ by Lemma 3.4. By (3.2) and constraints (i) and (ii) above we have

$$p_{x_1,y}(G) \leq \max_{\ell} \prod_{i=1}^{\ell} r_i \leq \max_{\ell} \left(\frac{m}{\ell}\right)^{\ell}.$$

The function $f(x) = \left(\frac{m}{x}\right)^x$ is maximised at $x = \frac{m}{e}$ and so $p_{x_1,y}(G) \leq e^{\frac{m}{e}} = e^{O(n)}$. This is asymptotically smaller than $c(T_k(n))$ by Lemma 2.1.

So $m \neq O(n)$. Suppose that $m \leq t_k(n) - 10n$ (otherwise we are done so assume) so that we obtain a bound for $c(G)$ from Corollary 3.2. Dividing this bound by $c(T_k(n)) = \Omega((\frac{k-1}{k})^n n^{n-\frac{1}{2}} e^{-n})$ gives

$$\frac{c(G)}{c(T_k(n))} = O\left(\lambda^n n^{2.5} e^{\frac{2k-1}{(k-1)\lambda}+(1-\lambda)n}\right), \tag{4.2}$$

where λ is defined in (3.1).

If we take the logarithm of the right hand side and call it R for ease of notation, we get

$$\begin{aligned} R &\leq 2.5\log(n) + n(\log(\lambda) + (1-\lambda)) + \frac{2k-1}{(k-1)\lambda} + O(1) \\ &\leq 2.5\log(n) + n(\log(\lambda) + (1-\lambda)) + 3\lambda^{-1} + O(1). \end{aligned}$$

First assume that $\lambda \leq 1 - n^{-\frac{1}{2}}\log(n)$: we will show that then $R \to -\infty$ and so (4.2) is $o(1)$.

If $\lambda \leq e^{-2}$, then $\log(\lambda) + (1-\lambda) \leq \frac{\log(\lambda)}{2}$. Furthermore we see from (3.1) that $\lambda = \Omega\left(\frac{m}{n^2}\right)$ and so $\lambda^{-1} = o(n)$. Therefore

$$\begin{aligned} R &\leq 2.5\log(n) + \frac{n}{2}\log(\lambda) + o(n) \\ &\leq 2.5\log(n) - n + o(n) \to -\infty, \end{aligned}$$

as n tends to infinity.

Otherwise, $\lambda^{-1} \leq e^2$ and since (by assumption) $\lambda \leq 1 - n^{-\frac{1}{2}}\log(n)$, we may apply Taylor's theorem to see

$$\begin{aligned} R &\leq 2.5\log(n) - n(1-\lambda)^2 + 3e^2 \\ &\leq 2.5\log(n) - \log^2(n) + 3e^2 \to -\infty, \end{aligned}$$

as n tends to infinity.

In either case R tends to $-\infty$ for sufficiently large n, and we must have that $c(G) < c(T_k(n))$, a contradiction.

Therefore $\lambda > 1 - \log(n)n^{-\frac{1}{2}}$. Equation (3.1) now allows us to conclude that $m \geq t_k(n) - O\left(n\log^2(n)\right)$, as required. $\square$

For the proof of Lemma 4.2 we require the Erdős-Stone Theorem [12].

Theorem 4.3 (Erdős-Stone [12])**.** *Let $k \geq 2$, $t \geq 1$, and $\varepsilon > 0$. Then for n sufficiently large, if G is a graph on n vertices with*

$$e(G) \geq \left(1 - \frac{1}{k-1} + \varepsilon\right)\binom{n}{2},$$

then G must contain a copy of $T_k(kt)$.

We now apply this theorem to complete the proof of Lemma 4.2.

Proof of Lemma 4.2. Let the best k-partition of G, be $V_1, \dots, V_k$. By Lemma 4.1, $e(G) > t_k(n) - O\left(n \log^2 n\right)$, and so Theorem 1.5 tells us that G contains $t_k(n)(1-o(1))$ edges between its vertex classes $V_1, \dots, V_k$. We therefore have $|V_i| = \frac{n}{k}(1+o(1))$ for each i. Also note that G cannot be k-partite (else $c(G) < c(T_k(n))$ by Lemma 2.2). Therefore G must contain an irregular edge. Now we count the cycles in G which contain only regular edges. Note that if we define G^R to be $G \backslash E_I$, where E_I is the set of irregular edges, then G^R is k-partite; $G^R \subseteq K_{\underline{a}}$ for some $\underline{a} = (a_1, \dots, a_k) \in \mathbb{N}^k$.

Let t be such that $H \subseteq T_k(tk) + e$, where e is any edge inside a vertex class of $T_k(tk)$. Pick an irregular edge uv: without loss of generality we may assume $uv \in V_1$. We first show that u and v cannot have $\frac{n}{10k}$ common neighbours in every other vertex class. Suppose otherwise and form a set Q by picking $\frac{n}{10k}$ vertices in $N(u) \cap N(v) \cap V_i$ for $i = 2, \dots, k$ and picking $\frac{n}{10k}$ vertices in V_1 to be in Q.

The graph $G^R[Q]$ does not contain a copy of $T_k(tk)$: if it did, it would contain a copy of $T_k(tk) + e$ and hence a copy of H. So then applying Theorem 4.3, there are $\Omega(n^2)$ regular edges that are not present in G, a contradiction. Thus, without loss of generality, $|N(u) \cap N(v) \cap V_2| < \frac{n}{10k}$ and, again without loss of generality, $|N(v) \cap V_2| \leq \frac{5n}{8k}$ (since $|V_2| = \frac{n}{k}(1+o(1))$ and we may assume that n is large).

When $k \geq 3$, this means that G cannot contain at least $\frac{3}{8}$ of the Hamilton cycles contained in $K_{\underline{a}}$ which start from v and then go to vertex class V_2. Recall that $h_v(i, K_{\underline{a}})$ is the number of permutations of $V(K_{\underline{a}}) = \{v_1, \dots, v_n\}$ such that $v_1 = v$, $v_2 \in V_i$ and $v_1 \cdots v_n$ is a Hamilton cycle. Since cycles may be counted at most twice due to orientation when considering permutations, the number of Hamilton cycles in $K_{\underline{a}}$ which start from v and then go to vertex class V_2 is at least $\frac{1}{2} h_v(2, K_{\underline{a}})$. By applying (2.1), we get

$$\begin{aligned} c(G^R) &\leq c(K_{\underline{a}}) - \frac{3}{8} \cdot \frac{1}{2} h_v(2, K_{\underline{a}}) \\ &= \sum_{r=3}^{n-1} c_r(K_{\underline{a}}) + \frac{1}{2} \sum_{i=3}^{k} h_v(i, K_{\underline{a}}) + \left(\frac{1}{2} - \frac{3}{16}\right) h_v(2, K_{\underline{a}}). \end{aligned}$$

Let $\underline{b} = (b_1, \dots, b_n)$, be such that $b_i \geq b_j$ if and only if $a_i \geq a_j$, and that $K_{\underline{b}} \cong T_k(n)$. Recall that $a_i = \frac{n}{k}(1+o(1))$ and so $\prod_{i=1}^{k} e^{\left|\log\left(\frac{b_i}{a_i}\right)\right|} = (1+o(1))$. Therefore by applying Lemmas 2.3 and 2.4 we get

$$\begin{aligned} c(G^R) &\leq \sum_{r=3}^{n-1} c_r(K_{\underline{a}}) + \prod_{i=1}^{k} e^{\left|\log\left(\frac{b_i}{a_i}\right)\right|} \left[\frac{1}{2} \sum_{i=3}^{k} h_v(i, T_k(n)) + \left(\frac{1}{2} - \frac{3}{16}\right) h_v(2, T_k(n))\right] \\ &= \sum_{r=3}^{n-1} c_r(K_{\underline{a}}) + (1+o(1)) \left(c_n(T_k(n)) - \frac{3}{16} h_v(2, T_k(n))\right) \\ &\leq (1+o(1)) \left(c(T_k(n)) - \frac{1}{8k} h(T_k(n))\right). \end{aligned}$$

Finally, we can apply Lemma 2.6 to get

$$\begin{aligned}
c(G^R) &\leq (1+o(1))\left(c(T_k(n)) - \frac{1}{24k}h(T_k(n)) - \frac{1}{12k}h(T_k(n))\right) \\
&\leq (1+o(1))\left(c(T_k(n))\left(1 - \frac{e^{-\frac{2k}{k-2}}}{24k}\right) - \frac{1}{12k}h(T_k(n))\right),
\end{aligned}$$

and so for n sufficiently large, $c(G^R) \leq c(T_k(n)) - \frac{1}{16k}h(T_k(n))$.

For $k = 2$, first consider that if $|V_1|$ and $|V_2|$ differ in size by more than 1, then G^R contains no cycle of length $2\lfloor n/2 \rfloor$. Counting cycles by length and applying Lemma 2.2 gives

$$\begin{aligned}
c(G^R) &= \sum_{r=2}^{\lfloor n/2 \rfloor - 1} c_{2r}(G^R) \\
&\leq \sum_{r=2}^{\lfloor n/2 \rfloor - 1} c_{2r}(T_2(n)) \\
&= c(T_2(n)) - c_{2\lfloor n/2 \rfloor}(T_2(n)).
\end{aligned}$$

Therefore assume that $|V_1|$ and $|V_2|$ differ in size by at most 1 (so G^R is a subgraph of $T_2(n)$). Recall (from the third paragraph of this proof) that G^R contains a vertex v with degree at most $5n/16$. Therefore, when applying the argument for $k \geq 3$, we lose at least a quarter of the cycles of length $2\lfloor n/2 \rfloor$ which contain v from $T_2(n)$. Note that v is present in at least half of the cycles of length $2\lfloor n/2 \rfloor$ in $T_2(n)$ and so $c(G^R) \leq c(T_2(n)) - \frac{1}{8}c_{2\lfloor \frac{n}{2} \rfloor}(T_k(n))$. $\square$

5 Counting Cycles in Complete multi-partite Graphs

In this section we present the proofs for the lemmas concerning counting cycles in complete multi-partite graphs that we stated in Section 2. We start with some preliminary lemmas. In order to state these we require some technical definitions.

Define a *code* on an alphabet $\mathcal{A}$ to be a string of letters $a_1 \cdots a_n$ where each a_i is in $\mathcal{A}$. For $k \geq 3$, we now discuss a way to count the number of Hamilton cycles in a k-partite graph G. Suppose each vertex class V_i of G is ordered. Consider a code $a_1 \cdots a_n$, where each $a_i \in [k]$. From such a code, we attempt to construct a Hamilton cycle $v_1 \cdots v_n$ in G as follows: for $j = 1, \ldots, n$ let $p(j) := |\{\ell \leq j : a_\ell = a_j\}|$. Define v_j to be the $p(j)$-th vertex in V_{a_j}. For $v_1 \cdots v_n$ to be a Hamilton cycle, each letter must appear in the code $a_1 \cdots a_n$ the correct number of times ($|\{j : a_j = i\}| = |V_i|$, for each $i \in [k]$) and any two consecutive letters of the code must be distinct ($a_j \neq a_{j+1}$ for each $j \in [n-1]$, and $a_1 \neq a_n$).

For a code $a_1 \cdots a_n$, with each $a_i \in [k]$, we say that the code is in Q if $a_i \neq a_{i+1}$ for each i, where indices are taken modulo n (so each pair of consecutive letters are distinct). For $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$, we say that the code is in $P_{\underline{c}}$ if there are c_i copies of i, for each $i \in [k]$. Finally we say that a code is in $P_{n,k}$ if it is in $P_{\underline{d}}$, where $\underline{d} = (d_1, \ldots, d_k) \in \mathbb{N}^k$ is such that $d_1 \leq d_2 \leq \ldots \leq d_k \leq d_1 + 1$ and $\sum_i d_i = n$.

In what follows it will be useful to consider a random code, so let $C_{n,k}$ denote the random code $C_{n,k} = a_1 \cdots a_n$, where each a_i is independently and uniformly distributed on $[k]$.

Enumerate the vertex set $V(K_{\underline{c}}) = \{v_1, \ldots, v_n\}$. We can count the number of Hamilton cycles in $K_{\underline{c}}$ by considering the probability that a permutation σ of $[n]$ picked uniformly gives a Hamilton cycle $v_{\pi(1)} \ldots v_{\pi(n)}$. Since we have a choice of orientation and starting vertex, each Hamilton cycle will be counted $2n$ times, and so

$$h(K_{\underline{c}}) = \frac{n!}{2n} \mathbb{P}\left[v_{\pi(1)} \ldots v_{\pi(n)} \text{ is a Hamilton cycle}\right]. \tag{5.1}$$

For $i \in [n]$, define $b_i \in [k]$ such that $v_{\pi(i)} \in V_{b_i}$. Then $b_1 \cdots b_n$ has the same distribution as $C_{n,k}$ conditioned on the event $\{C_{n,k} \in P_{\underline{c}}\}$. Note further that $v_{\pi(1)} \ldots v_{\pi(n)}$ is a Hamilton cycle if and only if $b_1 \cdots b_n \in Q$. Putting these into (5.1) gives

$$\begin{aligned} h(K_{\underline{c}}) &= \frac{n!}{2n} \mathbb{P}[b_1 \cdots b_n \in Q] \\ &= \frac{n!}{2n} \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}]. \end{aligned} \tag{5.2}$$

Obtaining good bounds on the probability that a random code is in Q (and similarly in $P_{\underline{c}}$) is relatively easy but approximating the probability of the intersection of the events proves more tricky. The following lemma will help us bound (5.2) from below, in order to prove Lemma 2.1.

Lemma 5.1. *Let $k \geq 2$ and suppose $C_{n,k} = a_1 \cdots a_n$ where the a_i are independent and identically uniformly distributed on $[k]$. If $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$, then*

$$\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}] \geq \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}],$$

and in particular,

$$\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}] \geq \mathbb{P}[C_{n,k} \in Q].$$

Proof. Let $k \geq 2$ and suppose $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$. Suppose that there exist some i and j such that $c_i \leq c_j - 2$, and let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_t = c_t$ for $t \neq i, j$. It is sufficient to show that $\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}'}] \geq \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}]$ – we may inductively find an i and j until the c_a differ by at most one and $\underline{c}$ corresponds to the vertex class sizes of a Turán graph.

Fix a subset A of $[n]$ with $|A| = n - (c_i + c_j)$ and let $R_{A,\underline{c}}$ be the event that $C_{n,k}$ is in $P_{\underline{c}}$, that $A = \{\ell : a_\ell \neq i, j\}$, and that $a_\ell \neq a_{\ell+1}$ for all ℓ in A and $a_n \neq a_1$ if both n and 1 are in A. $R_{A,\underline{c}}$ can be thought of as the event that everything in the code except the letters with values i and j behave well. Now note that we can partition over all the sets of size $n - (c_i + c_j)$ in $[n]$, and get the expression

$$\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}] = \sum_{A \in \binom{[n]}{n-(c_i+c_j)}} \mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}}] \cdot \mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}].$$

Note that given $P_{\underline{c}}$ holds, we may as well identify i and j when considering whether $R_{A,\underline{c}}$ holds. As such, $\mathbb{P}[R_{A,\underline{c}}|C_{n,k} \in P_{\underline{c}}]$ is constant with respect to c_i and c_j with fixed $c_i + c_j$. This in turn, means that $\mathbb{P}[R_{A,\underline{c}}|C_{n,k} \in P_{\underline{c}}] = \mathbb{P}[R_{A,\underline{c'}}|C_{n,k} \in P_{\underline{c'}}]$ and so to prove the first statement of the lemma, it is sufficient to show that

$$\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}] \leq \mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c'}}], \tag{5.3}$$

for each $A \subseteq [n]$, with $|A| = n - (c_i + c_j)$.

Let $A \subseteq [n]$, with $|A| = n - (c_i + c_j)$ and condition on the event $R_{A,\underline{c}}$ (note that we may assume that this event is not null else we have nothing to prove). If we consider $C_{n,k}$ as a code that is a cycle (imagine joining a_1 to a_n), then the occurrences of i, j form a collection of segments of total length $c_i + c_j$ with c_i copies of i and c_j copies of j. Conditioning just on $R_{A,\underline{c}}$, we have choice over where we place the i and j letters in the segments. Since we must have c_i total copies of i in the segments, there are $\binom{c_i+c_j}{c_i}$ such choices of placement of the i and j letters. Conditional on $R_{A,\underline{c}}$, the i and j placements are uniformly distributed on these $\binom{c_i+c_j}{c_i}$ choices. Conditional on $R_{A,\underline{c}}$, for the code $C_{n,k}$ to be in Q, the segments all have to be a string of letters alternating between i and j. As such the first letter of a segment dictates the remainder of that segment.

Let the lengths of the $\{i, j\}$-segments of $C_{n,k}$ be $r_1, \ldots, r_m$ and let s_{odd} and s_{even} be the number of odd length $\{i, j\}$-segments and even length $\{i, j\}$-segments respectively. We are then able to compute $\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}]$ by considering the starting letter of each $\{i, j\}$-segment. Suppose that t of the s_{odd} $\{i, j\}$-segments with odd length start with i. Then in the code, there will be $s_{\text{odd}} - 2t$ more appearances of j, than of i. Therefore, since $C_{n,k} \in P_{\underline{c}}$, we must have $2t - s_{\text{odd}} = c_i - c_j$ and so $t = \frac{s_{\text{odd}}+c_i-c_j}{2}$. Note that if $s_{\text{odd}} + c_i - c_j$ is odd, then $\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}] = 0$ since t must be an integer (and so we have nothing to prove). Therefore we assume that $s_{\text{odd}} + c_i - c_j$ is even in what follows.

We can specify such a code by choosing the starting letter of each even interval arbitrarily and choosing exactly t odd intervals to start with i. Comparing this with all possible choices of placements of i and j letters, we obtain

$$\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}] = \frac{2^{s_{\text{even}}}\binom{s_{\text{odd}}}{t}}{\binom{c_i+c_j}{c_i}}, \tag{5.4}$$

$$\begin{aligned} \mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c'}}] &= \frac{2^{s_{\text{even}}}\binom{s_{\text{odd}}}{t+1}}{\binom{c'_i+c'_j}{c'_i}} \\ &= \frac{2^{s_{\text{even}}}\binom{s_{\text{odd}}}{t+1}}{\binom{c_i+c_j}{c_i+1}}. \end{aligned} \tag{5.5}$$

Writing $b = c_j - c_i$ and dividing (5.4) by (5.5), we get

$$\begin{aligned}
\frac{\mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}}]}{\mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}'}]} &= \frac{c_j(s_{\text{odd}} + c_i - c_j + 2)}{(c_i + 1)(s_{\text{odd}} + c_j - c_i)} \\
&= \frac{(c_i + b)(s_{\text{odd}} - b + 2)}{(c_i + 1)(s_{\text{odd}} + b)} \\
&= \frac{c_i s_{\text{odd}} + 2c_i - bc_i + bs_{\text{odd}} + 2b - b^2}{c_i s_{\text{odd}} + bc_i + b + s_{\text{odd}}} \\
&= 1 - (b-1)\frac{2c_i + b - s_{\text{odd}}}{c_i s_{\text{odd}} + bc_i + b + s_{\text{odd}}}. \qquad (5.6)
\end{aligned}$$

Since there can be at most $c_i + c_j = 2c_i + b$ odd length $\{i, j\}$-segments, we have $2c_i + b \geq s_{\text{odd}}$, and $b \geq 2$. The right hand side of (5.6) must be less than or equal to 1 and so

$$\mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}}] \leq \mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}'}],$$

as required for (5.3). This completes the proof of the first statement of the lemma. For the second statement we partition $\mathbb{P}[C_{n,k} \in Q]$ over the $P_{\underline{c}}$ to give

$$\begin{aligned}
\mathbb{P}[C_{n,k} \in Q] &= \sum_{\underline{c}} \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{c}}] \\
&= \sum_{\underline{c}} \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}] \mathbb{P}[C_{n,k} \in P_{\underline{c}}] \\
&\leq \sum_{\underline{c}} \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}] \mathbb{P}[C_{n,k} \in P_{\underline{c}}] \\
&= \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}],
\end{aligned}$$

as required. □

We now use Lemma 5.1 to bound from below the number of Hamilton cycles in $T_k(n)$ and in turn prove Lemma 2.1.

Proof of Lemma 2.1. Let $k \geq 3$ and suppose $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$. Recalling (5.2), we note that if $c_i \leq c_j - 2$ and we let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_t = c_t$ otherwise, then applying Lemma 5.1 gives

$$h(K_{\underline{c}'}) \geq h(K_{\underline{c}}). \qquad (5.7)$$

Furthermore, we have

$$\begin{aligned} h(T_k(n)) &= \frac{n!}{2n}\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}] \\ &\geq \frac{n!}{2n}\mathbb{P}[C_{n,k} \in Q] \\ &= \frac{n!}{2n}\mathbb{P}[a_n \neq a_1, a_{n-1} | a_{n-1} \neq \cdots \neq a_1] \prod_{i=2}^{n-1} \mathbb{P}[a_i \neq a_{i-1} | a_{i-1} \neq \cdots \neq a_1] \\ &\geq \frac{n!}{2n}\left(\frac{k-2}{k}\right)\left(\frac{k-1}{k}\right)^{n-2} \\ &= \Omega\left(n^{n-\frac{1}{2}}e^{-n}\left(\frac{k-1}{k}\right)^n\right), \end{aligned}$$

as required.

For $k = 2$ we apply a simple counting argument. The number of cycles of length $t = 2\left\lfloor \frac{n}{2} \right\rfloor$ in $T_2(n)$ this is easily counted by ordering both colour classes and accounting for starting vertex and orientation. Therefore we get

$$c_t(T_2(n)) = \frac{\left(\left\lfloor \frac{n}{2} \right\rfloor\right)_{\frac{t}{2}} \left(\left\lceil \frac{n}{2} \right\rceil\right)_{\frac{t}{2}}}{2t} = \frac{\left\lfloor \frac{n}{2} \right\rfloor! \left\lceil \frac{n}{2} \right\rceil!}{4\left\lfloor \frac{n}{2} \right\rfloor},$$

and the result follows by applying Stirling's approximation. □

We now use a counting argument to prove Lemma 2.2.

Proof of Lemma 2.2. As before, let $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ be such that $\sum_i c_i = n$. If there exists i and j such that $c_i \leq c_j - 2$, and we let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_\ell = c_\ell$ otherwise. We are going to show that $c_r(K_{\underline{c}'}) \geq c_r(K_{\underline{c}})$, for all r.

Without loss of generality, we may assume that $i = 2$ and $j = 1$. We can count the number of cycles of a given length, r, by choosing r vertices and then counting the number of Hamilton cycles in graph induced by this cycle and then summing over all choices of r vertices:

$$c_r(K_{\underline{c}}) = \sum_{\substack{\underline{a} \in \prod_{i=1}^k \{0,\ldots,c_i\}: \\ \sum_{i=1}^k a_i = r}} \left[\left(\prod_{i=1}^k \binom{c_i}{a_i}\right) \cdot h\big(K_{\underline{a}}\big)\right].$$

Fix a copy K of $K_{\underline{c}}$ with vertex classes $V_1, \ldots, V_k$ and choose $v \in V_1$; then define K' to be $K \setminus v$ with a vertex v' added to V_2 which is a neighbour of all vertices not in V_2. We can see that K' is a copy of $K_{\underline{c}'}$. Using this coupling to compare $c_r(K_{\underline{c}})$ and $c_r(K_{\underline{c}'})$, we only need to consider cycles in K containing v and the cycles in K' containing v'. We write $c_{r,v}(G)$ to be the number of cycles of length r in G containing vertex v. In what follows we denote the unit vector in direction m by $\underline{e}_m = (y_1, \ldots, y_k)$, where $y_m = 1$ and $y_\ell = 0$ otherwise. Since we already assume that v is in our cycle, we then choose $r-1$ other vertices and count the number of Hamilton cycles on the induced subgraph to express $c_{r,v}(K)$ as

$$\begin{aligned}
&\sum_{\substack{\underline{a}\in\{0,\dots,c_1-1\}\times\prod_{i=2}^{k}\{0,\dots,c_i\}:\\ \sum_{i=1}^{k}a_i=r-1}} \left[\binom{c_1-1}{a_1}\cdot\left(\prod_{i=2}^{k}\binom{c_i}{a_i}\right)\cdot h\big(K_{\underline{a}+\underline{e}_1}\big)\right]\\
&=\sum_{\substack{a_1\in\{0,\dots,c_1-1\}\\ a_2\in\{0,\dots,c_2\}}}\left[\binom{c_1-1}{a_1}\binom{c_2}{a_2}\sum_{\substack{(a_3,\dots,a_k)\in\prod_{i=3}^{k}\{0,\dots,c_i\}:\\ \sum_{i=1}^{k}a_i=r-1}}\left[\left(\prod_{i=3}^{k}\binom{c_i}{a_i}\right)\cdot h\big(K_{\underline{a}+\underline{e}_1}\big)\right]\right]
\end{aligned}$$

and similarly we may express $c_{r,v'}(K')$ as

$$\begin{aligned}
&\sum_{\substack{a_1\in\{0,\dots,c_1-1\}\\ a_2\in\{0,\dots,c_2\}}}\left[\binom{c_1-1}{a_1}\binom{c_2}{a_2}\sum_{\substack{(a_3,\dots,a_k)\in\prod_{i=3}^{k}\{0,\dots,c_i\}:\\ \sum_{i=1}^{k}a_i=r-1}}\left[\left(\prod_{i=3}^{k}\binom{c_i}{a_i}\right)\cdot h\big(K_{\underline{a}+\underline{e}_2}\big)\right]\right]\\
&=\sum_{\substack{a_1\in\{0,\dots,c_1-1\}\\ a_2\in\{0,\dots,c_2\}}}\left[\binom{c_1-1}{a_1}\binom{c_2}{a_2}\sum_{\substack{(a_3,\dots,a_k)\in\prod_{i=3}^{k}\{0,\dots,c_i\}:\\ \sum_{i=1}^{k}a_i=r-1}}\left[\left(\prod_{i=3}^{k}\binom{c_i}{a_i}\right)\cdot h\big(K_{\underline{a}'+\underline{e}_1}\big)\right]\right],
\end{aligned}$$

where $\underline{a}'=(a_2,a_1,a_3,a_4,\dots,a_k)$ is the vector $\underline{a}$ with the first two values switched.

Define:

$$\eta(a_1,a_2,\underline{c},r):=\sum_{\substack{(a_3,\dots,a_n)\in\prod_{i=3}^{k}\{0,\dots,c_i\}:\\ \sum_{i=1}^{k}a_i=r-1}}\left[\left(\prod_{i=3}^{k}\binom{c_i}{a_i}\right)h\big(K_{\underline{a}+\underline{e}_1}\big)\right].$$

Then

$$c_{r,v}(K)=\sum_{\substack{a_1\in\{0,\dots,c_1-1\}\\ a_2\in\{0,\dots,c_2\}}}\binom{c_1-1}{a_1}\binom{c_2}{a_2}\eta(a_1,a_2,\underline{c},r) \tag{5.8}$$

and

$$c_{r,v'}(K')=\sum_{\substack{a_1\in\{0,\dots,c_1-1\}\\ a_2\in\{0,\dots,c_2\}}}\binom{c_1-1}{a_1}\binom{c_2}{a_2}\eta(a_2,a_1,\underline{c},r). \tag{5.9}$$

If we subtract (5.9) from (5.8) and split the sums depending on the values of a_1 and a_2, we get

$$
\begin{aligned}
c_{r,v'}(K') - c_{r,v}(K) = & \sum_{0 \le a_2 < a_1 \le c_2} \binom{c_1 - 1}{a_1}\binom{c_2}{a_2} \left(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r)\right) \\
& + \sum_{0 \le a_1 < a_2 \le c_2} \binom{c_1 - 1}{a_1}\binom{c_2}{a_2} \left(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r)\right) \\
& + \sum_{0 \le a_2 \le c_2 < a_1 \le c_1 - 1} \binom{c_1 - 1}{a_1}\binom{c_2}{a_2} \left(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r)\right).
\end{aligned}
$$

If we swap around the values of a_1 and a_2 in the second line of this expression, we get

$$
\begin{aligned}
& c_{r,v'}(K') - c_{r,v}(K) \\
& = \sum_{0 \le a_2 < a_1 \le c_2} \left(\binom{c_1 - 1}{a_1}\binom{c_2}{a_2} - \binom{c_1 - 1}{a_2}\binom{c_2}{a_1} \right) \Big(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r) \Big) \\
& + \sum_{\substack{a_1 \in \{c_2 + 1, \dots, c_1 - 1\} \\ a_2 \in \{0, \dots, c_2\}}} \binom{c_1 - 1}{a_1}\binom{c_2}{a_2} (\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r)).
\end{aligned}
\tag{5.10}
$$

From (5.7), we obtain that if $x > y$, then we have $\eta(x, y, \underline{c}, r) \le \eta(y, x, \underline{c}, r)$. Thus in the first sum of (5.10), when $a_1 > a_2$, we have $\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r) \ge 0$. At the same time, note that since $c_1 - 1 > c_2$,

$$
\binom{c_1 - 1}{x}\binom{c_2}{y} - \binom{c_1 - 1}{y}\binom{c_2}{x} > 0
$$

if and only if $x > y$. Combining these, we must have that for all $0 \le a_2 < a_1 \le c_2$

$$
\left(\binom{c_1 - 1}{a_1}\binom{c_2}{a_2} - \binom{c_1 - 1}{a_2}\binom{c_2}{a_1} \right) \Big(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r) \Big) \ge 0
$$

and so the first sum is positive.

In the second sum of (5.10), $a_1 > a_2$ and (5.7) tells us $\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r) \ge 0$. Thus the second sum is positive as well. We are then able to conclude that $c_{r,v'}(K') \ge c_{r,v}(K)$ as required.

All that remains is to prove that $c(T_k(n)) > c(G)$ for any k-partite graph G. Suppose that $G = K_{\underline{c}^0}$ where $\underline{c}^0 = (c_1^0, \dots, c_k^0) \in \mathbb{N}^k$ is such that $\sum_{i=1}^k c_i^0 = n$. While there exist some i and j such that $c_i^\ell \le c_j^\ell - 2$, define $\underline{c}^{\ell+1} = (c_1^{\ell+1}, \dots, c_k^{\ell+1})$ by $c_i^{\ell+1} = c_i^\ell + 1$, $c_j^{\ell+1} = c_j^{\ell+1} - 1$ and $c_r^{\ell+1} = c_r^\ell$ otherwise. Suppose that this process terminates with $\underline{c}^I$, so $T_k(n) \simeq K_{\underline{c}^I}$. Note that by successive applications of (5.7), we have $h(G) \le h(K_{\underline{c}^{I-1}})$.

We will now show that $h(G) < h(K_{\underline{c}^{I-1}})$. In order to do this, we have to consider (5.6) a bit more closely. If $h(G) = h(K_{\underline{c}^{I-1}})$, then at each application of (5.7), we have equality. So let us suppose, in order to obtain a contradiction, that $h(K_{\underline{c}^{I-1}}) = h(K_{\underline{c}^I})$, for some I. In this case, we must have that $s_{\mathrm{odd}} = c_i^{I-1} + c_j^{I-1}$ for all $A \in \binom{[n]}{n - (c_i^I + c_j^I)}$.

Say that a code $a_1, \ldots, a_n$ has an *ij transition* if there exists some s such that $\{a_s, a_{s+1}\} = \{i, j\}$ where indices are taken modulo n. For a fixed A, if $s_{\text{odd}} = c_i^{I-1} + c_j^{I-1}$ then there can be no ij in any code in Q conditional on $R_{A,\underline{c}^I}$. Therefore if $h(K_{\underline{c}^{I-1}}) = h(K_{\underline{c}^I})$ then there are no codes in $Q \cap P_{\underline{c}^I}$ with an ij transition. However we will show that we can construct such a code with an ij transition, and hence obtain our contradiction. We now split into two cases dependent on whether $c_i^I = c_j^I - 1$ or $c_i^I = c_j^I$.

First suppose that the $c_i^I = c_j^I - 1$. Since $K_{\underline{c}^I}$ is balanced, all vertex classes are of size c_i^I or c_j^I. In any Hamilton cycle of $K_{\underline{c}^I}$, there must be a transition from a vertex class of size c_i^I to a vertex class of size c_j^I and so by symmetry there must be a Hamilton cycle with a ij transition.

Now suppose that $c_i^I = c_j^I$. If all the vertex class sizes of $K_{\underline{c}^I}$ are the same, then we are done by symmetry. Similarly if the vertex class sizes of $K_{\underline{c}^I}$ are $c_i^I - 1$ and c_i^I, then there must be a transition between two classes of size c_i^I and so we are done by symmetry. Finally it remains to consider when $c_i^I = c_j^I$ and the vertex class sizes of $K_{\underline{c}^I}$ are c_i^I and $c^I + 1$. Consider a permutation $\pi = \pi_1 \cdots \pi_k$ such that $\pi_{k-1} = i$, $\pi_k = j$ and $\{\pi_1, \ldots, \pi_r\} = \{l : c_l^I = c_i^I + 1\}$. If $r = 1$ and $k = 3$, then $c_i^I \geq 2$ (else there are only four vertices) and so the code $\pi_1\pi_2\pi_1\pi_3(\pi_1\pi_2\pi_3) \cdots (\pi_1\pi_2\pi_3)$ is sufficient. If $r = 1$ and $k \geq 4$, then the code $\pi_1\pi_2\pi_1\pi_3\pi_4 \cdots \pi_k(\pi_1 \cdots \pi_k) \cdots (\pi_1 \cdots \pi_k)$ is sufficient. Finally, if $r \geq 2$, then the code $\pi_1 \cdots \pi_r(\pi_1 \cdots \pi_k) \cdots (\pi_1 \cdots \pi_k)$ is sufficient.

We have shown that there must be an instance of a strict inequality at (5.7) in the comparison of $h(K_{\underline{c}^{I-1}})$ with $h(K_{\underline{c}^I})$. It then follows immediately that $c(T_k(n)) = c(K_{\underline{c}^I}) > c(K_{\underline{c}^{I-1}}) \geq c(G)$. $\square$

The proof of Lemma 2.3 has a similar flavour to that of Lemma 5.1. We first prove a preliminary lemma where we evaluate $h_v(2, K_{\underline{c}})$ by considering random codes and then compare $h_v(2, K_{\underline{c}})$ with $h_v(2, K_{\underline{c}'})$. Lemma 2.3 will follow directly from this next lemma. (For what follows we define $R_{A,\underline{b}}$ as in the proof of Lemma 5.1.)

Lemma 5.2. *For $k \geq 3$, suppose $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$ with $0 \neq c_i \leq c_j - 2$. Let $\underline{c}' = (c_1', \ldots, c_k')$ be such that $c_i' = c_i + 1, c_j' = c_j - 1$ and $c_\ell' = c_\ell$ otherwise. Suppose $V_1, \ldots, V_k$ and $V_1', \ldots, V_k'$ are the vertex classes of $K_{\underline{c}}$ and $K_{\underline{c}'}$ and pick some $v \in V_1, v' \in V_1'$. Then*

$$h_v(2, K_{\underline{c}}) \leq \frac{(c_i + 1)c_j}{c_i(c_j - 1)} h_{v'}(2, K_{\underline{c}'}).$$

Proof. Recall that $h_v(2, K_{\underline{c}})$ counts orderings $v_1, \ldots, v_n$ of $V(K_{\underline{c}})$ where $v_1 = v$, $v_2 \in V_2$, and $v_1 \cdots v_n$ is a Hamilton cycle. There is a bijection between such an ordering and the pair $(C, (\pi_i)_{i \in [k]})$ where: C is a code $a_1 \cdots a_n$ on $[k]$ with $a_1 = 1$, $a_2 = 2$ that is in both Q and $P_{\underline{c}}$; and π_i is an ordering of V_i for each i and v is the first vertex in π_1. So if we let $C_{n,k} = a_1 \cdots a_n$ be a random code where each a_i is independently and identically uniformly distributed on $[k]$, we have an expression for $h_v(2, K_{\underline{c}})$:

$$h_v(2, K_{\underline{c}}) = k^n (c_1 - 1)! \left(\prod_{l=2}^{k} (c_l!) \right) \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{c}}, (a_1, a_2) = (1, 2)].$$

By considering the multinomial distribution with parameters n and $\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$ we have

$$\mathbb{P}\left[C_{n,k} \in P_{\underline{c}}\right] = \frac{n!}{\prod_{i=1}^{k}(c_i!)} k^{-n}, \tag{5.11}$$

and so

$$\begin{aligned} h_v(2, K_{\underline{c}}) &= \frac{n!}{c_1} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | C_{n,k} \in P_{\underline{c}}] \\ &= \frac{n!}{c_1} \sum_{A} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}}] \mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}] \end{aligned} \tag{5.12}$$

where $R_{A,\underline{c}}$ is defined as in the proof of Lemma 5.1, and the sum is taken over all $A \in \binom{[n]}{n-(c_i+c_j)}$.

For what follows, we only consider $A \in \binom{[n]}{n-(c_i+c_j)}$ such that $R_{A,\underline{c}} \cap \{(a_1, a_2) = (1,2)\} \neq \emptyset$ as these are the only ones that contribute to (5.12) when considering either $\underline{c}$ and $\underline{c}'$. As in the proof of Lemma 5.1, conditioning on $R_{A,\underline{c}}$, let s_{odd} and s_{even} be the number of $\{i, j\}$ subcodes with respectively odd and even lengths, where we consider the code cyclically. Unlike before, we now require $(a_1, a_2) = (1, 2)$ and so if one of i and j is 1 or 2, one of the subcodes will have a fixed value at a_1 and so a fixed starting letter. Let χ_{even} be the indicator that there is an even length subcode with a fixed first letter. Similarly let χ_{odd} be the indicator that there is an odd length subcode with a fixed first letter and further let $\chi_{\text{odd}}(i)$ and $\chi_{\text{odd}}(j)$ be the indicator that there is an odd length subcode with the first letter having fixed value i and j respectively.

As in Lemma 5.1, by letting $t = \frac{s_{\text{odd}}+c_i-c_j}{2}$ we can now compute $\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}}]$:

$$\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}}] = \frac{2^{s_{\text{even}}-\chi_{\text{even}}} \binom{s_{\text{odd}}-\chi_{\text{odd}}}{t-\chi_{\text{odd}}(i)}}{\binom{c_i+c_j}{c_i}}, \tag{5.13}$$

$$\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}'}] = \frac{2^{s_{\text{even}}-\chi_{\text{even}}} \binom{s_{\text{odd}}-\chi_{\text{odd}}}{t+1-\chi_{\text{odd}}(i)}}{\binom{c_i+c_j}{c_i+1}}. \tag{5.14}$$

Let $b = c_j - c_i \geq 2$. Note that the χ values will be the same when considering both $\underline{c}$ and $\underline{c}'$ and so dividing (5.13) by (5.14) gives

$$\begin{aligned} \frac{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}}]}{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}'}]} &= \frac{c_j(t+1-\chi_{\text{odd}}(i))}{(c_i+1)(s_{\text{odd}}-t-\chi_{\text{odd}}(j))} \\ &= \frac{c_j}{c_i+1} \cdot \frac{s_{\text{odd}}-b+2-2\chi_{\text{odd}}(i)}{s_{\text{odd}}+b-2\chi_{\text{odd}}(j)} \\ &\leq \frac{c_j}{c_i+1} \cdot \frac{s_{\text{odd}}-b+2}{s_{\text{odd}}+b-2}. \end{aligned} \tag{5.15}$$

Note that $\frac{s_{\text{odd}}-b+2}{s_{\text{odd}}+b-2}$ is non decreasing in s_{odd} and $s_{\text{odd}} \le 2c_i + b = 2c_j - b$, so we can bound (5.15) by taking $s_{\text{odd}} = 2c_i + b = 2c_j - b$ to get:

$$\begin{aligned}\frac{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2)|R_{A,\underline{c}}]}{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2)|R_{A,\underline{c}'}]} &\le \frac{c_j}{c_i+1} \cdot \frac{2c_i + b - b + 2}{2c_i + b - b - 2} \\ &= \frac{c_j(c_i+1)}{(c_i+1)(c_j-1)} \\ &= \frac{c_j}{c_j-1}. \end{aligned} \tag{5.16}$$

If we apply inequality (5.16) to (5.12):

$$h_v(2, K_{\underline{c}}) \le \frac{c_j}{c_j-1} \sum_A \left[\frac{n!}{c_1} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2)|R_{A,\underline{c}'}] \cdot \mathbb{P}[R_{A,\underline{c}}|C_{n,k} \in P_{\underline{c}}]\right].$$

Recall that $\mathbb{P}[R_{A,\underline{c}}|C_{n,k} \in P_{\underline{c}}] = \mathbb{P}[R_{A,\underline{c}'}|C_{n,k} \in P_{\underline{c}'}]$, so:

$$\begin{aligned} h_v(2, K_{\underline{c}}) &\le \frac{c_j}{c_j-1} \sum_A \left[\frac{n!}{c_1} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2)|R_{A,\underline{c}'}] \cdot \mathbb{P}[R_{A,\underline{c}}|C_{n,k} \in P_{\underline{c}}]\right] \\ &= \frac{c_1' c_j}{c_1(c_j-1)} \sum_A \left[\frac{n!}{c_1'} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2)|R_{A,\underline{c}'}] \cdot \mathbb{P}[R_{A,\underline{c}'}|C_{n,k} \in P_{\underline{c}'}]\right] \\ &= \frac{c_1' c_j}{c_1(c_j-1)} h_{v'}(2, K_{\underline{c}'}). \end{aligned}$$

Noting that $\frac{c_\ell'}{c_\ell}$ is maximised by $\ell = i$, we get

$$h_v(2, K_{\underline{c}}) \le \frac{(c_i+1)c_j}{c_i(c_j-1)} h_{v'}(2, K_{\underline{c}'}),$$

as required. □

We now apply this result to prove Lemma 2.3.

Proof of Lemma 2.3. Let $k \ge 3$ and $\underline{c} = (c_1, \ldots, c_n) \in \mathbb{N}^k$ and suppose $K_{\underline{c}}$ has vertex classes $V_1, \ldots, V_k$. Further suppose $T_k(n)$ has vertex classes $V_1', \ldots, V_k'$ with $b_i = |V_i'| < |V_j'| = b_j$ only if $c_i \le c_j$ and suppose that $v \in V_1 \cap V_1'$. We will prove by induction on $f(c,b) = \sum_i |c_i - b_i|$ that

$$h_v(2, K_{\underline{c}}) \le h_v(2, T_k(n)) \prod_{i=1}^k e^{\left|\log\left(\frac{b_i}{c_i}\right)\right|}.$$

The base case of $f(c,b) = 0$ follows since $K_{\underline{c}}$ is $T_k(n)$. Suppose that $f(c,b) \ge 1$ and the result holds for smaller values of $f(c,b)$. Note that if $f(c,b) \ne 0$, then since $\sum_i (c_i - b_i) = 0$, there must be i, j such that $c_i \le b_i - 1$ and $c_j \ge b_j + 1$. Let i and j be such that $b_i - c_i$ and

$c_j - b_j$ are maximised. If $b_i = b_j + 1$, we have a contradiction since then $c_i < c_j$, but $b_i > b_j$. This means that $c_j \geq c_i + 2$ and so if we let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_\ell = c_\ell$ otherwise, we may apply Lemma 5.2 to get that

$$\begin{aligned} h_v(2, K_{\underline{c}}) &\leq \frac{(c_i+1)c_j}{c_i(c_j-1)} h_v(2, K_{\underline{c}'}) \\ &= \exp\left\{ \left| \log\left(\frac{c'_i}{c_i}\right) \right| + \left| \log\left(\frac{c'_j}{c_j}\right) \right| \right\} h_v(2, K_{\underline{c}'}). \end{aligned} \tag{5.17}$$

To proceed by induction, we first observe that $f(c', b) < f(c, b)$ and secondly we must check that if $b_r < b_s$, then $c'_r \leq c'_s$. Note that this still holds for $r = i$ and $s = j$ and will still hold if neither $r = i$ nor $s = j$. If $r = i$ and $b_i < b_s$ but $c'_i > c'_s$, then it must be the case that $b_s - c_s > b_i - c_i$, which contradicts our choice of i. Similarly if we have $s = j$, $b_r < b_j$ and $c'_r > c_j$, then we arrive at the similar contradiction that $c_r - b_r > c_j - b_j$. Therefore we may apply the inductive hypothesis to (5.17) to conclude that

$$\begin{aligned} h_v(2, K_{\underline{c}}) &\leq \exp\left\{ \left| \log\left(\frac{c'_i}{c_i}\right) \right| + \left| \log\left(\frac{c'_j}{c_j}\right) \right| \right\} h_v(2, T_k(n)) \prod_{l=1}^{k} e^{\left|\log\left(\frac{b_l}{c'_l}\right)\right|} \\ &= h_v(2, T_k(n)) \prod_{l \neq i,j} e^{\left|\log\left(\frac{b_l}{c_l}\right)\right|} \prod_{l=i,j} \exp\left\{ \left| \log\left(\frac{b_l}{c'_l}\right) \right| + \left| \log\left(\frac{c'_l}{c_l}\right) \right| \right\} \\ &= h_v(2, T_k(n)) \prod_{i=1}^{k} e^{\left|\log\left(\frac{b_i}{c_i}\right)\right|}. \end{aligned}$$

□

We use a more complicated probabilistic argument for the proof of Lemma 2.4. We consider a different version of the random codes we have previously considered.

Proof of Lemma 2.4. Let K be a copy of the Turán graph $T_k(n)$ with vertex classes $V_1, \ldots, V_k$, and fix $b_i = |V_i|$ for each $i \in [k]$. (Note we do not order the sizes of the vertex classes.) Fix $a_1 = 1$, then given a_{i-1} for $i \geq 2$, let a_i be uniformly distributed on $[k] \setminus \{a_{i-1}\}$. Define the code $C^2(b_1, k) = a_1 \cdots a_m$, where $m = \max\{j : |\{i \leq j : a_i = 1\}| = b_1\}$ (in other words, keep track of a random walk on K_k and stop just before the (b_1+1)-th appearance of 1).

Conditional on $m = n$, the code $C^2(b_1, k)$ is uniformly distributed on codes $f_1 \cdots f_n$ in Q that contain b_1 copies of 1 and satisfy $f_1 = 1$. This is equal in distribution to $C_{n,k} = d_1 \cdots d_n$, where each d_i is independently uniformly distributed on $[k]$, conditional on $C_{n,k}$ being in Q, having b_1 copies of 1 and starting with $d_1 = 1$. This conditional equivalence between the two random codes allows us to compute bounds in new ways.

Let W be the number of transitions from 1 to 2 in $C^2(b_1, k)$ – that is $W = |\{j : (a_j, a_{j+1}) = (1, 2)\}|$. Note that any shift of a code in $Q \cap P_{\underline{b}}$ ($a_{M+1} \cdots a_n a_1 \cdots a_M$ for example) will also be in $Q \cap P_{\underline{b}}$. This means that we can shift the code $C^2(b_1, k)$ to each appearance of 1 to get another instance of a code $f_1 \cdots f_n$ in Q, with $f_1 = 1$ containing b_1 appearances of 1.

Thus by symmetry, given W, the probability that $C^2(b_1, k)$ starts with $(a_1, a_2) = (1, 2)$ is $\frac{W}{b_1}$. We seek to show that W is at most $\frac{b_1}{2k}$ with probability asymptotically smaller than the probability that $C^2(b_1, k)$ is in $P_{\underline{b}}$. With this we know that, conditional on the event $\{C^2(b_1, k) \in P_{\underline{b}}\}$, with high probability $W \geq \frac{b_1}{2k}$ and hence by symmetry

$$\mathbb{P}\left[a_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}\right] = \mathbb{E}\left[\frac{W}{b_1}\middle| C^2(b_1, k) \in P_{\underline{b}}\right] \geq \frac{1}{2k}(1 - o(1)).$$

Since each letter after a copy of 1 is independently and uniformly distributed on $\{2, \ldots, k\}$ and there are b_1 copies of 1, W is distributed like a Binomial random variable $\text{Bin}(b_1, \frac{1}{k-1})$. Applying a Chernoff bounds gives:

$$\mathbb{P}\left[W \leq \frac{n}{2k^2}\right] \leq e^{-\frac{n}{8k^2}}. \tag{5.18}$$

Now consider the probability that the code $C^2(b_1, k)$ is of the correct length. Note that the letter directly after a 1 cannot be a 1 but (until the next copy of 1), each subsequent letter is a 1 with probability $\frac{1}{k-1}$ and so removing the letter after each 1 and considering an appearance of a 1 as a *failure*, the variable $m - 2b_1$ is distributed like a Negative Binomial random variable, $\text{NB}(b_1, \frac{k-2}{k-1})$.

$$\begin{aligned}\mathbb{P}[m = n] &= \mathbb{P}\left[\text{NB}\left(b_1, \frac{k-2}{k-1}\right) = n - b_1\right] \\ &= \binom{n - (b_1 + 1)}{n - 2b_1}\left(\frac{k-2}{k-1}\right)^{n-2b_1}\left(\frac{1}{k-1}\right)^{b_1}.\end{aligned}$$

Now an application of de Moivre-Laplace (see [13, VII.3]) tells us that

$$\mathbb{P}[m = n] = \Theta\left(n^{-\frac{1}{2}} \exp\left\{-\frac{(b_1 - \frac{n-b_1}{k-1})^2}{2(n - b_1)\frac{k-2}{(k-1)^2}}\right\}\right). \tag{5.19}$$

Note that $|b_1 - \frac{n}{k}| < 1$, as the size of a vertex class of a copy of the Turán graph $T_k(n)$ and so $|b_1 - \frac{n-b_1}{k-1}| = |\frac{k}{k-1}(b_1 - \frac{n}{k})| < 2$. Putting this into (5.19), we see that

$$\begin{aligned}\mathbb{P}[m = n] &= \Theta\left(n^{-\frac{1}{2}} \exp\left\{-O\left(n^{-1}\right)\right\}\right) \\ &= \Theta\left(n^{-\frac{1}{2}}\right).\end{aligned} \tag{5.20}$$

Next, consider $\mathbb{P}\left[C^2(b_1, k) \in P_{\underline{b}} | m = n\right]$. As mentioned above, conditional on $m = n$, $C^2(b_1, k)$ is distributed like $C_{n,k}$ conditional on being in Q, starting with $d_1 = 1$ and having b_1 copies of 1. By Lemma 5.1, the events $\{C_{n,k} \in P_{\underline{b}}\}$ and $\{C_{n,k} \in Q\}$ are positively correlated and so

$$\begin{aligned}\mathbb{P}[C^2(b_1, k) \in P_{\underline{b}} | m = n] &= \mathbb{P}[C_{n,k} \in P_{\underline{b}} | C_{n,k} \in Q, d_1 = 1,\ b_1 \text{ copies of } 1] \\ &\geq \mathbb{P}[C_{n,k} \in P_{\underline{b}} | C_{n,k} \in Q] \\ &\geq \mathbb{P}[C_{n,k} \in P_{\underline{b}}].\end{aligned}$$

Recalling (5.11) and that $\left|b_i - \frac{n}{k}\right| < 1$ for all i, Stirling's approximation gives

$$\mathbb{P}[C^2(b_1,k) \in P_{\underline{b}} | m = n] = \Omega(n^{-\frac{k}{2}}). \tag{5.21}$$

So combining (5.20) and (5.21) we can conclude

$$\begin{aligned}\mathbb{P}\left[C^2(b_1,k) \in Q \cap P_{\underline{b}}\right] &= \mathbb{P}\left[C^2(b_1,k) \in P_{\underline{b}} | m = n\right] \mathbb{P}[m = n] \\ &= \Omega\left(n^{-\frac{k+1}{2}}\right). \end{aligned} \tag{5.22}$$

We can now complete our proof. We have

$$\begin{aligned} h_v\big(2, T_k(n)\big) &= k^n (b_1 - 1)! \left(\prod_{l=2}^{k} (b_l!)\right) \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{b}}, (d_1, d_2) = (1,2)] \\ &= k^n (b_1 - 1)! \left(\prod_{l=2}^{k} (b_l!)\right) \mathbb{P}[C_{n,k} \in Q, d_1 = 1, |\{j : d_j = 1]| = b_1\} \\ &\cdot \mathbb{P}[C_{n,k} \in P_{\underline{b}}, d_2 = 2 | C_{n,k} \in Q, d_1 = 1, |\{j : d_j = 1\}| = b_1]. \end{aligned}$$

Recall that $C_{n,k} = d_1 \cdots d_n$ given that $C_{n,k} \in Q$ and $d_1 = 1$ and $|\{j : d_j = 1\}| = b_1$ is equal in distribution to $C^2(b_1, k) = a_1 \cdots a_m$ given $m = n$ and so

$$\begin{aligned} h_v\big(2, T_k(n)\big) &= k^n (b_1 - 1)! \left(\prod_{l=2}^{k} (b_l!)\right) \mathbb{P}[C_{n,k} \in Q, d_1 = 1, |\{j : d_j = 1\}| = b_1] \\ &\cdot \mathbb{P}[C^2(b_1,k) \in P_{\underline{b}}, a_2 = 2 | m = n] \\ &= k^n (b_1 - 1)! \left(\prod_{l=2}^{k} (b_l!)\right) \mathbb{P}[C_{n,k} \in Q, d_1 = 1, |\{j : d_j = 1\}| = b_1] \\ &\cdot \mathbb{P}[a_2 = 2 | C^2(b_1,k) \in P_{\underline{b}}, m = n] \cdot \mathbb{P}[C^2(b_1,k) \in P_{\underline{b}} | m = n]. \end{aligned} \tag{5.23}$$

We can bound $\mathbb{P}[d_2 = 2 | C^2(b_1,k) \in P_{\underline{b}}, m = n]$ by conditioning on the value of W as follows:

$$\begin{aligned} \mathbb{P}[a_2 = 2 | C^2(b_1,k) \in P_{\underline{b}}, m = n] &\geq \mathbb{P}\left[a_2 = 2 \middle| C^2(b_1,k) \in P_{\underline{b}}, m = n, W > \frac{n}{2k^2}\right] \\ &- \mathbb{P}\left[W \leq \frac{n}{2k^2} \middle| C^2(b_1,k) \in P_{\underline{b}}, m = n\right] \\ &\geq \frac{n}{2k^2 b_1} - \frac{\mathbb{P}[W \leq \frac{n}{2k^2}]}{\mathbb{P}[C^2(b_1,k) \in P_{\underline{b}}, m = n]}. \end{aligned}$$

By applying (5.18) and (5.22) we get

$$\begin{aligned} \mathbb{P}[a_2 = 2 | C^2(b_1,k) \in P_{\underline{b}}, m = n] &\geq \frac{n}{2k^2 b_1} - O\left(\frac{e^{-\frac{n}{8k^2}}}{n^{-\frac{k+1}{2}}}\right) \\ &= \frac{n}{2k^2 b_1} - o(1). \end{aligned}$$

This means that for sufficiently large n, $\mathbb{P}[a_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}, m = n] \geq \frac{1}{3k}$. Putting this into (5.23), we see

$$\begin{aligned}
h_v\big(2, T_k(n)\big) \geq{} & \frac{k^n (b_1 - 1)!}{3k} \left(\prod_{l=2}^{k} (b_l!) \right) \mathbb{P}[C_{n,k} \in Q, d_1 = 1, |\{j : d_j = 1\}| = b_1] \\
& \cdot \mathbb{P}[C^2(b_1, k) \in P_{\underline{b}} | m = n] \\
={} & \frac{k^n (b_1 - 1)!}{3k} \left(\prod_{l=2}^{k} (b_l!) \right) \mathbb{P}[C_{n,k} \in Q, d_1 = 1, |\{j : d_j = 1\}| = b_1] \\
& \cdot \mathbb{P}[C_{n,k} \in P_{\underline{b}} | C_{n,k} \in Q, d_1 = 1, |\{j : d_j = 1\}| = b_1] \\
={} & \frac{k^n (b_1 - 1)!}{3k} \left(\prod_{l=2}^{k} (b_l!) \right) \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{b}}, d_1 = 1] \\
={} & \frac{k^n}{2n} \left[\prod_{i=1}^{k} (b_i!) \right] \cdot \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{b}}] \cdot \frac{2n \cdot \mathbb{P}[d_1 = 1 | C_{n,k} \in Q \cap P_{\underline{b}}]}{3kb_1} \\
={} & h\big(T_k(n)\big) \cdot \frac{2n \cdot \mathbb{P}[d_1 = 1 | C_{n,k} \in Q \cap P_{\underline{b}}]}{3kb_1}.
\end{aligned}$$

By symmetry, $\mathbb{P}[d_1 = 1 | C_{n,k} \in Q \cap P_{\underline{b}}] = \frac{b_1}{n}$. This completes the proof of the lemma. □

Now we bound below the number of Hamilton cycles in $T_k(n)$ by the number of Hamilton cycles in $T_k(m)$, where $m < n$.

Proof of Lemma 2.5. Let v be a vertex contained in the largest vertex class V_i in $T_k(n)$. Removing v gives $T_k(n-1)$. For each Hamilton cycle $v_1 \cdots v_{n-1}$ in $T_k(n-1)$, we can form a Hamilton cycle in $T_k(n)$ by inserting v between two vertices v_j and v_{j+1}, both not in V_i. For each Hamilton cycle in $T_k(n-1)$, there are at least $(n-1)\frac{k-2}{k}$ spaces where we can insert v and under this construction each Hamilton cycle in $T_k(n)$ will be formed in at most one way. Counting over all Hamilton cycles in $T_k(n-1)$, we get that

$$h(T_k(n)) \geq (n-1)\frac{k-2}{k} h(T_k(n-1)). \tag{5.24}$$

We can apply equation (5.24) inductively to get that for any $i \in [n]$,

$$h(T_k(n)) \geq (n-1)_i \left(\frac{k-2}{k} \right)^i h(T_k(n-i)).$$

□

We now bound the number of cycles in $T_k(n)$ in terms of the number of Hamilton cycles.

Proof of Lemma 2.6. Let I be a subset of $[n]$ with $|I| = r$. Then by Lemma 2.2 and Lemma 2.5, we have

$$\begin{aligned} h(G[I]) &\leq h(T_k(r)) \\ &\leq \left(\frac{k}{k-2}\right)^{n-r} \frac{h(T_k(n))}{(n-1)_{n-r}} \\ &\leq \left(\frac{2k}{k-2}\right)^{n-r} \frac{h(T_k(n))}{(n)_{n-r}}. \end{aligned}$$

Summing over all subsets I, we have

$$\begin{aligned} c(T_k(n)) &\leq \sum_{i=0}^{n-3} \binom{n}{i} \left(\frac{2k}{k-2}\right)^{i} \frac{h(T_k(n))}{(n)_i} \\ &= h(T_k(n)) \sum_{i=0}^{n-3} \frac{1}{i!} \left(\frac{2k}{k-2}\right)^{i} \\ &\leq e^{\frac{2k}{k-2}} h(T_k(n)), \end{aligned}$$

as required. $\square$

Finally, we prove Lemma 2.7.

Proof of Lemma 2.7. Let $n \in \mathbb{N}$ and denote $\lfloor \frac{n}{2} \rfloor$ by t and $\lceil \frac{n}{2} \rceil$ by t'. For $r \geq 2$, the number of cycles of length $2r$ in $T_2(n)$ is

$$\frac{(t)_r\,(t')_r}{2r}.$$

Summing over $r = 2, \ldots, t$ gives

$$\begin{aligned} c(T_2(n)) &= \sum_{r=2}^{t} \frac{(t)_r\,(t')_r}{2r} \\ &= \frac{t!t'!}{2t} \sum_{r=2}^{t} \frac{t}{r(t-r)!(t'-r)!} \\ &\leq \frac{t!t'!}{2t} \sum_{r'=0}^{t-2} \frac{t}{(t-r')r'!r'!}, \end{aligned}$$

where we substituted $r' = t - r$ to obtain the second equality. As $c_{2t}(T_2(n)) = \frac{t!t'!}{2t}$ and $\frac{t}{(t-s)s!}$ is easily bounded by 2, we have

$$\begin{aligned} c(T_2(n)) &\leq 2c_{2t}(T_2(n)) \sum_{r'=0}^{t-2} \frac{1}{r'!} \\ &\leq 2c_{2t}(T_2(n)) \sum_{r'\geq 0} \frac{1}{r'!} = 2e \cdot c_{2t}(T_2(n)). \end{aligned} \tag{5.25}$$

Let $s = \left\lfloor \frac{n-1}{2} \right\rfloor$ and $s' = \left\lceil \frac{n}{2} \right\rceil$. Note that $t = s'$ and $t' = s + 1$, and so

$$\frac{n-2}{2}\frac{s'!s!}{2s} \leq \frac{s}{t} \cdot \frac{s'!s!t'}{2s} = \frac{t!t'!}{2t}. \tag{5.26}$$

Using (5.26) gives

$$c_{2\left\lfloor \frac{n-1}{2} \right\rfloor}(T_2(n-1)) = \frac{s'!s!}{2s} \leq \frac{2}{n-2} \cdot \frac{t!t'!}{2t} = \frac{2}{n-2} c_{2\left\lfloor \frac{n}{2} \right\rfloor}(T_2(n)).$$

As $i = o(n)$, repeatedly applying this bound along with (5.25) gives

$$\begin{aligned} c(T_2(n-i)) &\leq 2e \cdot c_{2\left\lfloor \frac{n-i}{2} \right\rfloor}(T_2(n-i)) \\ &\leq 2e \left(\prod_{j=1}^{i} \frac{2}{n-j-1} \right) c_{2\left\lfloor \frac{n}{2} \right\rfloor}(T_2(n)) \\ &\leq 2e \left(\frac{4}{n} \right)^i c_{2\left\lfloor \frac{n}{2} \right\rfloor}(T_2(n)), \end{aligned}$$

as required. □

6 Conclusion and Open Questions

In this paper we resolve Conjecture 1.1 for sufficiently large n (we do not optimise the value of n given by our approach, as it would still be very large). For triangle-free graphs, Arman, Gunderson and Tsaturian [4] (see also [9]) show that the Turán graph $T_2(n)$ uniquely maximises the number of cycles when $n \geq 141$, but it seems likely that this should hold for all values of n.

Theorem 1.3 only deals with H such that $\chi(H) \geq 3$ and H contains a critical edge. When H does not satisfy these properties, our approach is not feasible as the extremal H-free graph is no longer $T_k(n)$. It is interesting to consider what could be true for such H. For example, it is natural to ask whether it is possible to maximize the number of edges and the number of cycles simulateously (as in Theorem 1.3).

Question 6.1. Let H be a fixed graph. Does $\mathrm{EX}(n; H)$ contain a graph with $m(n; H)$ cycles for sufficiently large n?

As $T_2(n)$ does not contain any odd cycle, Theorem 1.3 implies that for any odd k, $T_2(n)$ is the n-vertex graph with odd girth at least k containing the most cycles. Arman, Gunderson and Tsaturian [4] ask a more general question.

Question 6.2 (Arman, Gunderson, Tsaturian [4]). What is the maximum number of cycles in an n-vertex graph, with girth at least g?

This question seems difficult since comparatively little is known about the maximum number of edges in an graph with girth at least $g \geq 4$.

Another interesting problem was raised by Király [18] who asked for the maximum number of cycles in a graph with m edges can contain (without constraining the number of vertices); he conjectured an upper bound of 1.4^m cycles. In a recent paper Arman and Tsaturian [5] give an upper bound of $8.25 \times 3^{m/3}$ and a lower bound of 1.37^m, and conjecture that their upper bound is correct to within a $(1+o(1))^m$ factor. It would be interesting to consider the effect of adding the additional constraint of forbidding a subgraph. In particular what is the maximum number of cycles that a triangle-free graph with m edges can contain?

A similar problem to that of Király is to maximise the number of cycles in a graph with n vertices and m edges. For $m = \Omega(n^2)$ and n sufficiently large, Arman and Tsaturian [5, Conjecture 6.1] conjecture a maximum of $(1+o(1))^n \left(\frac{2m}{en}\right)^n$ cycles. The current best upper bound is $(1+o(1))^n \left(\frac{2m}{2n}\right)^n$ given in the same paper. We believe that the method used to prove Lemma 3.2 improves this upper bound but does not prove the conjecture.

Another direction of research is to maximise the number of *induced* cycles. Given a graph G, let $m_I(G)$ denote the number of induced cycles in G and let $m_I(n) := \max\{m_I(G) : |V(G)| = n\}$. Morrison and Scott [20] recently determined $m_I(n)$ for n sufficiently large and proved that the extremal graphs are unique. The extremal graphs in question are essentially blow-ups of $C_{n/3}$ and contain many copies of C_4.

It would be interesting to consider what happens to the extremal graphs when we forbid C_4.

Question 6.3. What is $m_I(n; C_4) := \max\{m_I(G) : |V(G)| = n, G \text{ is } C_4\text{-free}\}$?

Acknowledgements. We would like to thank the referees for their careful reading of the manuscript and detailed comments.

References

[1] W. Ahrens, *Ueber das Gleichungssystem einer Kirchhoff'schen galvanischen Stromverzweigung*, Math. Ann. **49** (1897), no. 2, 311–324.

[2] N. Alon and C. Shikhelman, *Many T copies in H-free graphs*, J. Combin. Theory Ser. B **121** (2016), 146–172.

[3] A. Arman, *Maximum number of cycles in graphs and multigraphs*, Ph.D. thesis, The University of Manitoba, http://hdl.handle.net/1993/33030, 2018.

[4] A. Arman, D. S. Gunderson, and Sergei Tsaturian, *Triangle-free graphs with the maximum number of cycles*, Discrete Math. **339** (2016), 699–711.

[5] A. Arman and A. Tsaturian, *The maximum number of cycles in a graph with fixed number of edges*, arXiv:1702.02662v2, preprint, February 2017.

[6] B. Bollobás, *On complete subgraphs of different orders*, Math. Proc. Cambridge Philos. Soc. **79** (1976), no. 1, 19–24.

[7] B. Bollobás, *Extremal Graph Theory*, Academic Press, London, 1978.

[8] B. Bollobás and E. Győri, *Pentagons vs. triangles*, Discrete Math. **308** (2008), no. 19, 4332–4336.

[9] S. Durocher, D. S. Gunderson, P. C. Li, and M. Skala, *Cycle-maximal triangle-free graphs*, Discrete Math. **338** (2015), 274–290.

[10] P. Erdős, *On the number of complete subgraphs contained in certain graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **7** (1962), 459–464.

[11] P. Erdős and M. Simonovits, *A limit theorem in graph theory*, Studia Sci. Math. Hungar **1** (1966), 51–57.

[12] P. Erdős and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091.

[13] W. Feller, *Introduction to Probability Theory and its Applications*, 3rd ed., vol. I, New York: John Wiley & Sons, 1968.

[14] A. Grzesik, *On the maximum number of five-cycles in a triangle-free graph*, J. Combin. Theory Ser. B **102** (2012), no. 5, 1061–1066.

[15] A. Grzesik and B. Kielak, *On the maximum number of odd cycles in graphs without smaller odd cycles*, arXiv:1806.09953v1, preprint, June 2018.

[16] E. Győri and H. Li, *The maximum number of triangles in C_{2k+1}-free graphs*, Combin. Probab. Comput. **21** (2012), no. 1-2, 187–191.

[17] H. Hatami, J. Hladký, D. Král', S. Norine, and A. Razborov, *On the number of pentagons in triangle-free graphs*, J. Combin. Theory Ser. A **120** (2013), 722–732.

[18] Z. Király, *Maximum number of cycles and hamiltonian cycles in sparse graphs*, Tech. report, 2009, `https://web.cs.elte.hu/egres/tr/egres-09-03.ps`, accessed 11 February 2019.

[19] G. Kirchhoff, *Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird*, Annalen der Physik **148** (1847), no. 12, 497–508.

[20] N. Morrison and A. Scott, *Maximising the number of induced cycles in a graph*, J. Combin. Theory Ser. B **126** (2017), 24–61.

[21] A. Roberts and A. Scott, *Stability results for graphs with a critical edge*, Euro. J. Combin. **74** (2018), 27–38.

[22] M. Simonovits, *Extremal graph problems with symmetrical extremal graphs. Additional chromatic conditions*, Discrete Math. **7** (1974), 349–376.

[23] P. Turán, *On an extremal problem in graph theory*, Mat. Fiz. Lapok **48** (1941), 436–452.