# EMBEDDING PROBLEMS AND RAMSEY-TURÁN VARIATIONS IN EXTREMAL GRAPH THEORY

BY

MARYAM SHARIFZADEH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Alexandr V. Kostochka, Chair
Professor József Balogh, Director of Research
Assistant Professor Kay Kirkpatrick
Assistant Professor Theo Molla

# Abstract

In this dissertation, we will focus on a few problems in extremal graph theory. The first chapter consists of some basic terms and tools.

In Chapter 2, we study a conjecture of Mader on embedding subdivisions of cliques. Improving a bound by Mader, Bollobás and Thomason, and independently Komlós and Szemerédi proved that every graph with average degree $d$ contains a subdivision of $K_{\Omega(\sqrt{d})}$. The disjoint union of complete bipartite graph $K_{r,r}$ shows that their result is best possible. In particular, this graph does not contain a subdivision of a clique of order $\omega(r)$. However, one can ask whether their bound can be improved if we forbid such structures. There are various results in this direction, for example Kühn and Osthus proved that their bound can be improved if we forbid a complete bipartite graph of fixed size. Mader proved that that there exists a function $g(r)$ such that every graph $G$ with $\delta(G) \geq r$ and girth at least $g(r)$ contains a $TK_{r+1}$. He also asked about the minimum value of $g(r)$. Furthermore, he conjectured that $C_4$-freeness is enough to guarantee a clique subdivision of order linear in average degree. Some major steps towards these two questions were made by Kühn and Osthus, such as $g(r) \leq 27$ and $g(r) \leq 15$ for large enough $r$. In an earlier result, they proved that for $C_4$-free graphs one can find a subdivision of a clique of order almost linear in minimum degree. Together with József Balogh and Hong Liu, we proved that every $C_{2k}$-free graph, for $k \geq 3$, has such a subdivision of a large clique. We also proved the dense case of Mader's conjecture in a stronger sense.

In Chapter 3, we study a graph-tiling problem. Let $H$ be a fixed graph on $h$ vertices and $G$ be a graph on $n$ vertices such that $h|n$. An *$H$-factor* is a collection of $n/h$ vertex-disjoint copies of $H$ in $G$. The problem of finding sufficient conditions for a graph $G$ to have an $H$-factor has been extensively studied; most notable is the celebrated Hajnal-Szemerédi Theorem which states that every $n$-vertex graph $G$ with $\delta(G) \geq (1-1/r)n$ has a $K_r$-factor. The case $r=3$ was proved earlier by Corrádi and Hajnal. Another type of problems that have been studied over the past few decades are the so-called Ramsey-Turán problems. Erdős and Sós, in 1970, began studying a variation on Turán's theorem: What is the maximum number of edges in an $n$-vertex, $K_r$-free graph $G$ if we add extra conditions to avoid the very

strict structure of Turán graph. In particular, what if besides being $K_r$-free, we also require $\alpha(G) = o(n)$. Since the extremal example for the Hajnal-Szemerédi theorem is very similar to the Turán graph, one can similarly ask how stable is this extremal example. With József Balogh and Theodore Molla, we proved that for an $n$-vertex graph $G$ with $\alpha(G) = o(n)$, if $\delta(G) \geq (1/2 + o(1))n$ then $G$ has a triangle factor. This minimum degree condition is asymptotically best possible. We also consider a fractional variant of the Corrádi-Hajnal Theorem, settling the triangle case of a conjecture of Balogh, Kemkes, Lee, and Young.

In Chapter 4, we first consider a Ramsey-Turán variant of a theorem of Erdős. In 1962, he proved that for any $r > \ell \geq 2$, among all $K_r$-free graphs, the $(r-1)$-partite Turán graph has the maximum number of copies of $K_\ell$. We consider a Ramsey-Turán-type variation of Erdős's result. In particular, we define $\mathrm{RT}(F, H, f(n))$ to be the maximum number of copies of $F$ in an $H$-free graph with $n$-vertices and independence number at most $f(n)$. We study this function for different graphs $F$ and $H$. Recently, Balogh, Hu and Simonovits proved that the Ramsey-Turán function for even cliques experiences a jump. We show that the function $\mathrm{RT}(K_3, H, f(n))$ has a similar behavior when $H$ is an even clique. We also study the sparse analogue of a theorem of Bollobás and Győri about the maximum number of triangles that a $C_5$-free graph can have. Finally, we consider a Ramsey-Turán variant of a function studied by Erdős and Rothschild about the maximum number of edge-colorings that an $n$-vertex graph can have without a monochromatic copy of a given graph.

*To my parents, Zahra Rezaei and Mohsen Sharifzadeh,*
*and my grandmother, Sadrijan Rezaei.*

# Acknowledgments

First, I would like to express my deepest appreciation to my advisor József Balogh. I have been so fortunate to have the opportunity to learn from him. Without his guidance and persistence, this dissertation would not have been possible.

I am grateful to Sasha Kostochka and Theo Molla for many valuable and fruitful discussions. I also wish to thank Kay Kirkpatrick for being on my dissertation committee.

My sincere thanks also goes to many colleagues who provided me an opportunity to work with and learn from them. In particular, I am grateful to Hong Liu for encouragement and lots of stimulating discussions.

I am always greatly indebted to my parents and my brother for their unconditional love and support.

# Table of Contents

# Chapter 1

# Introduction and preliminaries

Extremal graph theory is one of the main branches of graph theory that has experienced significant developments in the past few decades. Given a family of graphs $\mathcal{G}$, denote $\lambda$ a graph invariant and $\mathcal{P}$ a graph property. Let $\mathcal{G}' \subseteq \mathcal{G}$ consists of all $G \in \mathcal{G}$ that satisfy $\mathcal{P}$. Two meta-questions in this branch of graph theory are the followings:

(1) Determine the minimum (maximum) value of $\lambda$ among all $G \in \mathcal{G}'$, and describe the structure of such $G$.

(2) Determine the minimum (maximum) value $\lambda_1$ such that for all $G \in \mathcal{G}$, if $\lambda(G) \geq \lambda_1$ ($\lambda(G) \leq \lambda_1$) then $G \in \mathcal{G}'$.

In this dissertation, in Chapter 2, we will study an embedding problem of type (1). In Chapter 3 and Chapter 4, we will consider a Ramsey-Turán-type variation of some classical results. The results in Chapter 3 are of type (2) and in Chapter 4 are of type (1). Probabilistic techniques have been proved to be particularly powerful for attacking such problems. Among other tools, we will make use of these techniques to study these questions. For the rest of this chapter, we will introduce some basic definitions in Section 1.1, then in Section 1.2, we will go over some inequalities and bounds on random variables. At the end, in Section 1.3, we will state Szemerédi's regularity lemma, which is one of the most powerful tools in extremal combinatorics.

## 1.1 Basic definitions

In this section, we will define some basic terminology which will be used throughout this dissertation. A graph is a combinatorial object that has been used in many areas of mathematics and also any other field where discrete models play crucial roles such as: number theory, information theory, computer science and social sciences. A *graph* $G = (V, E)$ is an ordered pair consisting of a set of *vertices* or *nodes*, $V$, and a set of *edges* or *relations*, $E$, where the set $E$ is a set of pairs of vertices, for example, $V(G) = \{v_1, v_2, \ldots, v_5\}$ and

$E(G) = \{v_1v_3, v_1v_5, v_4v_3, v_5v_2\}$. We say two vertices $u, v \in V(G)$ are *adjacent*, if $uv \in E(G)$. Denote $d(v)$ the number of vertices $v' \in V(G)$ such that $v$ and $v'$ are adjacent.

For the rest of this section, let $G = (V, E)$ be an $n$-vertex graph. An $h$-vertex graph $H$ is a *subgraph* of $G$, $H \subseteq G$, if the following is true. We can assign a unique vertex of $G$ to every vertex of $H$, such that two different vertices of $H$ are assigned to two different vertices of $G$. In addition, if for two vertices of $H$, $u$ and $v$, $uv$ is an edge in $H$ then the two vertices assigned to $u$ and $v$ also form an edge in $G$. We say that $G$ is *$H$-free* if $H \nsubseteq G$, i.e. $H$ is not a subgraph of $G$.

For $U \subseteq V$ and $v \in V$, denote $d_U(v)$ the degree of $v$ in $U$ and $N_U(v)$ the set of vertices $u \in U$ such that $vu \in E(G)$. Also, define $G[U]$ to be the subgraph of $G$ induced on the vertex set $U$, i.e. the subgraph consisting of all the edges with both ends in $U$. Denote $E(G[U])$ the edge set of $G$ with both ends in $U$. Define $U \subseteq V$ to be an *independent set* of $G$ if and only if $E(G[U]) = \emptyset$. In other words, an independent set is a subset of vertices in which no two vertices form an edge.

The *chromatic number* of $G$, denoted by $\chi(G)$, is the minimum number of colors needed to color the vertex set of $G$ such that it satisfies the following property. For every edge of $G$, $uv$, the color assigned to $u$ and $v$ are different.

The *complete* $k$-vertex graph or *clique* of order $k$, denoted by $K_k$, is a graph where $E(K_k) = \{uv : \text{for all } u, v \in V(K_k)\}$. We say that $G$ is a *$k$-partite* graph if $V(G)$ can be partitioned into $k$ parts, $V(G) = V_1 \cup \ldots \cup V_k$, such that $G[V_i]$ is an independent set for all $1 \leq i \leq k$. If, in addition, for all pairs of vertices $u \in V_i$ and $v \in V_j$ with $i \neq j$, $uv \in E(G)$, then $G$ is a *complete $k$-partite* graph. The Turán graph, named after the Hungarian mathematician Pál Turán, is an $n$-vertex complete $k$-partite graph, denoted by $T_k(n)$, where all the partite sets have size $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$.

We will also use the following two graphs. In 1995, Kim [35] proved the existence of a triangle-free graph on $n$ vertices with independence number $O(\sqrt{n \log n})$. Hence, Kim's graph has $O(n^{3/2} \log^{1/2} n)$ edges.

In 1976, Bollobás and Erdős [10] constructed an $n$-vertex $K_4$-free graph with independence number $o(n)$ and $(\frac{1}{8} + o(1))n^2$ edges, which we will denote by $\mathrm{BE}(n)$. We follow the explanation in [56] to present their construction. For a constant $\varepsilon > 0$, and large enough integers $d$ and $n_0$, let $n > n_0$ and $\mu = \varepsilon/\sqrt{d}$. Next, partition the high-dimensional unit sphere $\mathbb{S}^d$ into $n/2$ domains of equal measure and diameter (maximum distance) less than $\mu/2$, $D_1, \ldots, D_{n/2}$. For every $1 \leq i \leq n/2$, choose a vertex $x_i \in D_i$ and $y_i \in D_i$. Let $V(\mathrm{BE}(n)) = X \cup Y$, where $X = \{x_1, \ldots, x_{n/2}\}$ and $Y = \{y_1, \ldots, y_{n/2}\}$. For every $x, x' \in X$ and $y, y' \in Y$,

(1) let $xy \in E(\mathrm{BE}(n))$ if their distance is less than $\sqrt{2} - \mu$,

(2) let $xx' \in E(\mathrm{BE}(n))$ if their distance is more than $2 - \mu$,

(3) let $yy' \in E(\mathrm{BE}(n))$ if their distance is more than $2 - \mu$.

Note that the number of edges with both ends in $X$ or $Y$ is $o(n^2)$.

## 1.2 Concentration inequalities

One of the main tools in probabilistic arguments are concentration inequalities. These inequalities give bounds on the probability that a random variable is far from its expected value. First we will state Markov's inequality.

**Theorem 1.2.1.** *Let $X$ be a non-negative random variable. For any constant $a > 0$*

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}X}{a}.$$

Often, the bound given by the Markov's inequality is not strong enough. In some applications, we are working with a random variable that can be written as a sum of independent binary random variables, and therefore we can use Chernoff bound which gives us exponential estimates.

**Theorem 1.2.2.** *Let $X_1, \ldots, X_n$ be independent random variables with values in $\{0, 1\}$, $X = X_1 + \ldots + X_n$, and $\mu = \mathbb{E}(X)$. For every $c \in (0, 1)$, we have*

$$\mathbb{P}(X \geq (1+c)\mu) \leq e^{-\frac{c^2\mu}{3}} \quad \text{and} \quad \mathbb{P}(X \leq (1-c)\mu) \leq e^{-\frac{c^2\mu}{2}}.$$

In particular, Theorem 1.2.2 can be applied when $X$ is a binomial random variable. There are situations where the corresponding random variable can be written as a sum of binary random variables, which are not independent. The following variation of Chernoff bound will be applied if we are working with a hypergeometric random variable, see e.g. Corollary 2.3, Theorem 2.8 and Theorem 2.10 in [30].

**Theorem 1.2.3.** *Let $X$ be a hypergeometric random variable with $\mathbb{E}(X) = \mu$. For any constant $0 < \lambda \leq 3/2$,*

$$\mathbb{P}(|X - \mu| \geq \lambda\mu) \leq 2e^{-\frac{\lambda^2}{3}\mu}.$$

Next theorem, which is known as the "independent bounded differences inequality" can be used in more general cases, (see [51]).

**Theorem 1.2.4.** *Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a family of independent random variables with $X_k$ taking values in a set $A_k$ for each $k$. Suppose that the real-valued function $f$ defined on $\prod A_k$ satisfies $|f(\mathbf{x}) - f(\mathbf{x}')| \leq \sigma_k$ whenever the vectors $\mathbf{x}$ and $\mathbf{x}'$ differ only in the $k$-th coordinate. Let $\mu$ be the expected value of the random variable $f(\mathbf{X})$. Then for any $t \geq 0$,*

$$\mathbb{P}(|f(\mathbf{X}) - \mu| \geq t) \leq 2e^{-2t^2/\sum \sigma_k^2}.$$

## 1.3 Regularity lemma

Szemerédi regularity lemma has proved to be one of the most powerful tools of extremal combinatorics. In 1975, it was originally introduced by Szemerédi [59] to prove a long-standing conjecture of Erdős and Turán [25] about arithmetic progressions in subsets of integers. The lemma states that the vertex set of any large graph can be partitioned into constant many parts of almost equal size, such that the distribution of edges in between almost all parts is random-like. We first need some definitions.

**Definition 1.3.1.** For every $n$-vertex graph $G = (V, E)$ and $X, Y \subseteq V$, denote

$$E(X, Y) = \{uv \in E(G) : u \in X \text{ and } v \in Y\}, \quad e(X, Y) = |E(X, Y)|,$$

and

$$d(X, Y) = \frac{e(X, Y)}{|X||Y|}.$$

For a constant $\varepsilon > 0$, the pair $(X, Y)$ is $\varepsilon$-regular, if for every $X' \subseteq X$ and $Y' \subseteq Y$ satisfying $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$, we have

$$|d(X, Y) - d(X', Y')| \leq \varepsilon.$$

A partition of $V(G) = V_1 \cup \ldots \cup V_t$ is $\varepsilon$-regular if for all $1 \leq i, j \leq t$, $||V_i| - |V_j|| \leq 1$, and all but at most $\varepsilon t^2$ pairs of $(V_i, V_j)$ are $\varepsilon$-regular.

Now, we can state the lemma formally.

**Theorem 1.3.2** (Szemerédi's regularity lemma)**.** *For every $\varepsilon > 0$ and positive integer $m$, there exists an $n_0$ and $M$, such that every $n$-vertex graph $G$ with $n > n_0$ has an $\varepsilon$-regular partition $V_1 \cup \ldots \cup V_t$ with $m < t < M$.*

# Chapter 2

# Embedding subdivisions of cliques

Mader conjectured that every $C_4$-free graph has a subdivision of a clique of order linear in its average degree. We show that every $C_6$-free graph has such a subdivision of a large clique.

We also prove the dense case of Mader's conjecture in a stronger sense, i.e., for every $c$, there is a $c'$ such that every $C_4$-free graph with average degree $cn^{1/2}$ has a subdivision of a clique $K_\ell$ with $\ell = \lfloor c'n^{1/2} \rfloor$ where every edge is subdivided exactly 3 times.

## 2.1 Introduction

A *subdivision* of a clique $K_\ell$, denoted by $TK_\ell$, is a graph obtained from $K_\ell$ by subdividing each of its edges into internally vertex-disjoint paths. Bollobás and Thomason [12], and independently Komlós and Szemerédi [39] proved the following celebrated result.

**Theorem 2.1.1.** *Every graph of average degree $d$ contains a subdivision of a clique of order $\Omega(\sqrt{d})$.*

Theorem 2.1.1 is best possible: the disjoint union of $K_{d,d}$'s contains no subdivision of $K_\ell$ with $\ell \geq \sqrt{8d}$ (observed first by Jung [31]).

Mader [49] conjectured that if a graph is $C_4$-free, then one can find a subdivision of a much larger clique, of order linear in its average degree. Two major steps towards this conjecture were made by Kühn and Osthus: in [46], they showed that if the graph $G$ has girth at least 15 and large average degree, then the conjecture is true in a stronger sense: a subdivision of $K_{\delta(G)+1}$ is guaranteed; in [44], they showed that one can find a subdivision of a clique of order almost linear, $\Omega(d/\log^{12} d)$, in any $C_4$-free graph with average degree $d$.

Extending ideas in [38] and [39], we prove that every $C_6$-free graph has such a subdivision of a large clique.

**Theorem 2.1.2.** *Let $G$ be a $C_6$-free graph with average degree $d$. Then a $TK_\ell$ is a subgraph of $G$ with $\ell = \lfloor cd \rfloor$ for some small positive constant $c$ independent of $d$.*

Similar proof gives the following result, whose proof is omitted.

**Theorem 2.1.3.** *Let $G$ be a $C_{2k}$-free graph with $k \geq 3$ and average degree $d$. Then a $TK_\ell$ is a subgraph of $G$ with $\ell = \lfloor cd \rfloor$ for some small positive constant $c$ independent of $d$.*

It is known that any $C_4$-free $n$-vertex graph has at most $O(n^{3/2})$ edges (see [40]). Our next result verifies the dense case of Mader's conjecture in a stronger sense.

**Theorem 2.1.4.** *For every $c > 0$ there is a $c' > 0$ such that the following holds. Let $G$ be a $C_4$-free $n$-vertex graph with $cn^{3/2}$ edges. Then $G$ contains a $TK_\ell$ with $\ell = \lfloor c'n^{1/2} \rfloor$, in which every edge of the $K_\ell$ is subdivided exactly 3 times.*

Theorem 2.1.4 can also be viewed as an extension of the following result of Alon, Krivelevich and Sudakov [2] for $C_4$-free graphs. Settling a question of Erdős [18], they showed, using the dependent random choice lemma, that if the average degree of a graph is of order $\Omega(n)$, then there is a $TK_\ell$ with $\ell = \Omega(n^{1/2})$, in which every edge of the $K_\ell$ is subdivided exactly once.

**Notation.** For a vertex $v$, denote by $S(v, i)$ the $i$-th sphere around $v$, i.e., the set of vertices of distance $i$ from $v$ and denote by $B(v, r)$ the ball of vertices of radius $r$ around $v$, so $B(v, r) = \cup_{i \leq r} S(v, i)$. For a set $X \subseteq V(G)$, denote by $\Gamma(X)$ the external neighborhood of $X$, that is $\Gamma(X) := N(X) \setminus X$. Denote by $d(G)$ the average degree of $G$ and for $S \subseteq V(G)$ denote by $d(S)$ the average degree of the induced subgraph $G[S]$. For a set of vertices $S$, denote by $N_i(S)$ the $i$-th common neighborhood of $S$, i.e., vertices of distance exactly $i$ from every vertex in $S$. For a set $B \subseteq V(G)$, let $\Delta(B) := \max_{v \in B} d_G(v)$ and $\delta(B) := \min_{v \in B} d_G(v)$.

**Outline of the chapter.** In Section 2.2, we will introduce some tools and theorems that will be used later in proofs of main results. The proof of Lemma 2.2.4 will be given in Section 2.3 as well as the reduction of Theorem 2.1.2 to Theorem 2.2.5. The proof of Theorem 2.2.5 will be divided into two parts according to the range of $d$: the dense case when $d \geq \log^{14} n$ will be handled in Section 2.4, and the sparse case when $d < \log^{14} n$ in Section 2.5. The proof of Theorem 2.1.4 will be given in Section 2.6. In Section 2.7, we will give some concluding remarks.

We will omit floors and ceilings signs when they are not crucial.

## 2.2 Preliminaries

For any graph $G$, there is a bipartite subgraph $G'$ such that $e(G') \geq e(G)/2$. We shall use a result of Györi [28] which states that every bipartite $C_6$-free graph has a $C_4$-free subgraph with at least half of its edges. So having a loss of factor of 4 in the average degree, we may assume that our $C_6$-free graph is bipartite and also $C_4$-free. Following Komlós and Szemerédi [38], we introduce the following concept.

**$(\varepsilon_1, t)$-expander:** For $\varepsilon_1 > 0$ and $t > 0$, let $\varepsilon(x)$ be the function as follows:

$$\varepsilon(x) = \varepsilon(x, \varepsilon_1, t) := \begin{cases} 0 & \text{if } \ x < t/5 \\ \varepsilon_1/\log^2(15x/t) & \text{if } x \geq t/5. \end{cases} \tag{2.1}$$

For the sake of brevity, on $\varepsilon(x)$ we do not write the dependency of $\varepsilon_1$ and $t$ when it is clear from the context. Note that $\varepsilon(x)\cdot x$ is increasing for $x \geq t/2$. A graph $G$ is an $(\varepsilon_1, t)$-*expander* if $|\Gamma(X)| \geq \varepsilon(|X|) \cdot |X|$ for all subsets $X \subseteq V$ of size $t/2 \leq |X| \leq |V|/2$.

Komlós and Szemerédi [38, 39] showed that every graph $G$ contains an $(\varepsilon, t)$-expander that is almost as dense as $G$.

**Theorem 2.2.1.** *Let $t > 0$, and choose $\varepsilon_1 > 0$ sufficiently small (independent of $t$) so that $\varepsilon = \varepsilon(x)$ defined in (2.1) satisfies $\int_1^\infty \frac{\varepsilon(x)}{x} dx < \frac{1}{8}$. Then every graph $G$ has a subgraph $H$ with $d(H) \geq d(G)/2$ and $\delta(H) \geq d(H)/2$, which is an $(\varepsilon_1, t)$-expander.*

**Remark:** The subgraph $H$ might be much smaller than $G$. For example if $G$ is a vertex-disjoint collection of $K_{d+1}$'s, then $H$ will be just one of the $K_{d+1}$'s.

We will use the following version of Theorem 2.2.1.

**Corollary 2.2.2.** *There exists $\varepsilon_0$ with $0 < \varepsilon_0 < 1$ such that for every $0 < \varepsilon_1 \leq \varepsilon_0$, $\varepsilon_2 > 0$ and every graph $G$, there is a subgraph $H \subseteq G$ with $d(H) \geq d(G)/2$ and $\delta(H) \geq d(H)/2$ which is an $(\varepsilon_1, \varepsilon_2 d(H)^2)$-expander.*

*Proof.* Let $G' \subseteq G$ be a subgraph maximizing $d(G')$ and define $t' := \varepsilon_2 d(G')^2/4$. If $\varepsilon_0$ is sufficiently small, then for any $\varepsilon_1 \leq \varepsilon_0$, applying Theorem 2.2.1 yields a $(4\varepsilon_1, t')$-expander $H \subseteq G'$ with $d(G')/2 \leq d(H) \leq d(G')$ and $\delta(H) \geq d(H)/2$. Define $t := \varepsilon_2 d(H)^2$. Since $d(G')/2 \leq d(H) \leq d(G')$, we have $t' \leq t \leq 4t'$. A simple calculation shows that for every $x \geq t/2$,

$$\frac{4\varepsilon_1}{\log^2(15x/t')} \geq \frac{\varepsilon_1}{\log^2(15x/t)}.$$

Hence $H$ is an $(\varepsilon_1, t)$-expander as desired. □

Every $(\varepsilon_1, t)$-expander graph has the following robust "small diameter" property (see Corollary 2.3 in [39]):

**Corollary 2.2.3.** *If $G$ is an $(\varepsilon_1, t)$-expander, then any two vertex sets, each of size at least $x \geq t$, are of distance at most*

$$diam := diam(n, \varepsilon_1, t) = \frac{2}{\varepsilon_1} \log^3(15n/t),$$

*and this remains true even after deleting $x\varepsilon(x)/4$ arbitrary vertices from $G$.*

By Corollary 2.2.2, we may assume, when proving Theorem 2.1.2, that $G$ is a bipartite, $\{C_4, C_6\}$-free, $(\varepsilon_1, t)$-expander graph with average degree $d$, $\delta(G) \geq d/2$ and $t = \varepsilon_2 d^2$ for some $\varepsilon_1 \leq \varepsilon_0$ and $\varepsilon_2 > 0$. Indeed, instead of $G$ we might work in a still dense subgraph $H$ of it, having the properties listed before and by resetting $d := d(H) \geq d(G)/2$ it suffices to find in $H$ a $TK_\ell$ with $\ell = \Omega(d(H))$. The next lemma finds in $G$ a "nice" subgraph with "bounded" maximum degree.

**Lemma 2.2.4.** *Let $0 < \varepsilon_1 < 1$ and $\varepsilon_2 > 0$. Let $G$ be an $n$-vertex bipartite, $C_4$-free, $(\varepsilon_1, \varepsilon_2 d^2)$-expander graph with average degree $d$ and $\delta(G) \geq d/2$. Then either $G$ contains a subdivision of a clique of order linear in $d$, or $G$ has a $C_4$-free subgraph $G'$ with average degree $d(G') \geq d/2$ and $\delta(G') \geq d(G')/4$, that is $(\varepsilon_1/8, 4\varepsilon_2 d(G')^2)$-expander. Furthermore, $G'$ has at least $n/2$ vertices and $\Delta(G') \leq d(G') \log^8(|V(G')|/d(G')^2)$.*

Note that we do not use the $C_6$-freeness of $G$ in Lemma 2.2.4. Using Lemma 2.2.4, to prove Theorem 2.1.2, it will be sufficient to show Theorem 2.2.5 below.

**Theorem 2.2.5.** *Let $0 < \varepsilon_1 \leq \varepsilon_0$ and $\varepsilon_2 > 0$, where $\varepsilon_0$ is the constant from Corollary 2.2.2. Let $G$ be an $n$-vertex bipartite, $\{C_4, C_6\}$-free, $(\varepsilon_1, \varepsilon_2 d^2)$-expander graph with average degree $d$, $\delta(G) \geq d/4$ and $\Delta(G) \leq d \log^8 n$. Then $G$ contains a $TK_{\ell/2}$ for $\ell = cd$ for some constant $c > 0$ independent of $d$.*

## 2.3 Reduction to "bounded" maximum degree

Let $G$ be an $n$-vertex bipartite $C_4$-free $(\varepsilon_1, \varepsilon_2 d^2)$-expander graph with average degree $d$ and $\delta(G) \geq d/2$.

In this section, we will show that we can transform $G$ into a subgraph $G'$ with $d(G') \geq d/2$, $\delta(G') \geq d(G')/4$ and $\Delta(G') \leq d(G') \log^8(|V(G')|/d(G')^2)$, where $G'$ is an $(\varepsilon_1/8, 4\varepsilon_2 d(G')^2)$-expander. For simplicity, throughout this section, define

$$t := \varepsilon_2 d^2 \qquad \text{and} \qquad t' := 4\varepsilon_2 d(G')^2.$$

To prove Lemma 2.2.4, we shall use the following two lemmas: Lemmas 2.3.1 and 2.3.2.

Choose a constant $c < \frac{1}{24000}$ such that $c \ll \varepsilon_1$. Set the parameters as follows:

$$\ell = cd, \quad m = \log \frac{15n}{t}, \quad \Delta = \frac{dm^8}{600}, \quad \Delta' = dm^4, \quad \varepsilon(n, \varepsilon_1, t) = \frac{\varepsilon_1}{m^2}, \quad diam = \frac{2m^3}{\varepsilon_1}.$$

Note that $d$ has to be sufficiently large (say $d > 1/c$) so that $\ell \geq 1$.

If $m \leq 1/c^2$, then $d \geq e^{-1/2c^2} n^{1/2}$, and we can apply Theorem 2.1.4 to get a subdivision of a clique of order linear in $d$. Thus we may assume that $1/m \ll c \ll \varepsilon_1$. By the same argument, we may also assume that $d\Delta \leq n$ and $n/d^2 \gg 1/\varepsilon_2$.

Let $L \subseteq V(G)$ be the set of all vertices of degree at least $\Delta$.

**Lemma 2.3.1.** *We can find in $G$ either a $TK_{\ell/2}$, or $|L| \leq \ell$ and $G' := G[V \setminus L]$ has maximum degree at most $\Delta$.*

*Proof.* Indeed, if $|L| \geq \ell$, then we can choose a subset $L' \subseteq L$ of exactly $\ell$ vertices, say $L' := \{v_1, \ldots, v_\ell\}$. We shall build a copy of $TK_{\ell/2}$ using a subset of these high-degree vertices from $L'$ as core vertices.

First we choose for each vertex $v_i$, $S_1(v_i) \subseteq S(v_i, 1)$ and $S_2(v_i) \subseteq S(v_i, 2)$ such that:

(i) all $S_1(v_i)$'s are pairwise disjoint, and each $S_1(v_i)$ is disjoint from $L'$ and of size $\Delta/2$;

(ii) every $S_2(v_i)$ is disjoint from $\bigcup_{j=1}^{\ell} S_1(v_j) \cup L'$, and each $S_2(v_i)$ is of size $d\Delta/5$;

(iii) for every $1 \leq i \leq \ell$, each vertex in $S_1(v_i)$ has at most $d/2$ neighbors in $S_2(v_i)$.

We can indeed select such sets:

For (i), since $G$ is $C_4$-free, for any $v_i$, every other $v_j$ with $j \neq i$ has at most one neighbor in $S(v_i, 1)$. Since $|S(v_i, 1)| - 2\ell \geq \Delta - 2\ell \geq \Delta/2$, we can remove these neighbors of $v_j$'s and $L'$ from $S(v_i, 1)$ and then choose exactly $\Delta/2$ vertices for $S_1(v_i)$.

For (ii) and (iii), recall that $G$ is bipartite and $\delta(G) \geq d/2$. Thus we can choose, for each vertex in $S_1(v_i)$, exactly $d/2 - 1$ vertices in $S(v_i, 2)$. Since $G$ is $C_4$-free, for a given $v_i$, all chosen vertices should be distinct. Thus we have chosen at least $(d/2 - 1)(\Delta/2) \geq 100\ell\Delta \geq 100 \left| \bigcup_{j=1}^{\ell} S_1(v_j) \right|$ vertices, simply discard those vertices which are in $\bigcup_{j=1}^{\ell} S_1(v_j) \cup L'$ and then choose $d\Delta/5$ vertices for $S_2(v_i)$. Clearly $S_2(v_i)$ satisfies both (ii) and (iii).

We now describe the greedy algorithm that we use to connect the vertices $v_i$'s. Denote by $B_1(v_i) := S_1(v_i) \cup \{v_i\}$ and by $B_2(v_i) := B_1(v_i) \cup S_2(v_i)$.

**Greedy Algorithm:** We try to connect these $\ell$ core vertices pair by pair in an arbitrary order. For the current pair of core vertices $v_i, v_j$, we try to connect $B_2(v_i)$ and $B_2(v_j)$ using a shortest path of length at most *diam* and then exclude all the internal vertices in this path from further connections. We need to justify that such a short path exists.

Suppose we have already connected some pairs using paths of length at most *diam*. We will exclude all previously used vertices from $B_1(v_i) \cup B_1(v_j)$ and also those vertices from

$S_2(v_i)$, $S_2(v_j)$ adjacent to removed vertices from $S_1(v_i)$ or $S_1(v_j)$. Formally, let $U$ be the set of vertices used in previous connections and denote by $U_i := U \cap S_1(v_i)$ and by $U_j := U \cap S_1(v_j)$. Define $N := (\Gamma(U_i) \cap S_2(v_i)) \cup (\Gamma(U_j) \cap S_2(v_j))$. Then the set of vertices excluded is $U \cup N$. First we bound the size of $U$, it is at most

$$\ell^2 \cdot diam \le c^2 d^2 \cdot \frac{2m^3}{\varepsilon_1} \le cd^2 m^3,$$

as there are at most $\ell^2$ pairs of core vertices and for each connection, the length of a path is bounded by *diam*.

Call a core vertex $v_i$ bad, if more than $\Delta'$ vertices from $S_1(v_i)$ are used in previous connections. During the connections, we discard a core vertex when it becomes bad. We discard in total at most $\ell/2$ core vertices. Indeed, we have used at most $\ell^2 \cdot diam$ vertices. Since by (i), $S_1(v_i)$'s are pairwise disjoint, each bad core vertex, by definition, uses at least $\Delta'$ of them. Thus the number of discarded bad core vertices is at most

$$\frac{\ell^2 \cdot diam}{\Delta'} \le \frac{cd^2m^3}{dm^4} = \frac{cd}{m} \ll \frac{\ell}{2}.$$

Hence there are at least $\ell/2$ core vertices survive the entire process.

Recall that by (iii), each vertex in $U_i$ (or $U_j$ resp.) has at most $d/2$ neighbors in $S_2(v_i)$ (or $S_2(v_j)$ resp.). Note that every survived core vertex is not bad, namely $|U_i| \le \Delta'$. Thus $|N| \le \Delta' \cdot d/2 = d^2m^4/2$. Hence the total number of vertices we exclude from $B_2(v_i)$ (or $B_2(v_j)$ resp.) is at most

$$\ell^2 \cdot diam + |N| \le cd^2m^3 + \frac{1}{2}d^2m^4 \le d^2m^4.$$

After excluding these vertices, we still have at least

$$|S_2(v_i)| - \ell^2 \cdot diam - |N| \ge \frac{d\Delta}{5} - d^2m^4 \ge \frac{d\Delta}{10}$$

vertices left in $S_2(v_i)$, the same holds for $S_2(v_j)$. Recall that, when $x \ge t/2$, $\varepsilon(x, \varepsilon_1, t)$ is decreasing and $x\varepsilon(x, \varepsilon_1, t)$ is increasing. So we have that the number of vertices we are allowed to exclude, by Corollary 2.2.3, is at least

$$\frac{1}{4} \cdot \frac{d\Delta}{10} \cdot \varepsilon\left(\frac{d\Delta}{10}, \varepsilon_1, t\right) \ge \frac{d\Delta}{40} \cdot \varepsilon(n, \varepsilon_1, t) \ge \frac{d^2m^8}{24000} \cdot \frac{\varepsilon_1}{m^2} = \frac{\varepsilon_1 d^2 m^6}{24000} \gg d^2m^4,$$

where the last inequality follows from $1/m \ll c \ll \varepsilon_1$ and $c < \frac{1}{24000}$. Thus the exclusion of these vertices will not affect the robust small diameter property between $B_2(v_i)$'s. So the $\ell/2$ remaining core vertices can be connected to form a $TK_{\ell/2}$. □

Given that $c$ is sufficiently small and now we can assume $|L| \leq \ell$, we have that $|V(G')| \geq n - \ell \geq n/2$. Note that $d(G') \geq \frac{2(dn/2-\ell n)}{n} = d - 2\ell \geq d/2$, thus $t' \geq t$. On the other hand, $G' = G[V \setminus L]$ and $L$ consists of vertices of degree at least $\Delta \gg d$, thus $d(G') \leq \frac{nd-|L|\Delta/2}{n-|L|} \leq d$. Hence $t' \leq 4t$ and $\delta(G') \geq \delta(G) - \ell \geq d/2 - \ell \geq d(G')/4$.

**Lemma 2.3.2.** *The obtained graph $G'$ is an $(\varepsilon_1/8, t')$-expander.*

*Proof.* Recall that $t \leq t' \leq 4t$. Since $G$ is an $(\varepsilon_1, t)$-expander, for any set $X$ in $G'$ of size $x \geq t'/2 \geq t/2$, it is easy to check that

$$\begin{aligned} |\Gamma_G(X)| &\geq x \cdot \varepsilon(x, \varepsilon_1, t) = x \cdot \frac{\varepsilon_1}{\log^2(15x/t)} \geq x \cdot \frac{\varepsilon_1/4}{\log^2(15x/t')} = x \cdot \varepsilon(x, \varepsilon_1/4, t') \\ &\geq \frac{t'}{2} \cdot \varepsilon\left(\frac{t'}{2}, \frac{\varepsilon_1}{4}, t'\right) = \frac{\varepsilon_1 t'}{8\log^2(7.5)} \gg \ell \geq |L|. \end{aligned}$$

Hence $|\Gamma_{G'}(X)| \geq |\Gamma_G(X)| - |L| \geq x\varepsilon(x, \varepsilon_1/4, t') - \ell \geq \frac{1}{2}x\varepsilon(x, \varepsilon_1/4, t') = x\varepsilon(x, \varepsilon_1/8, t')$. □

Recall that $1/\varepsilon_2 \ll n/d^2 \leq 2|V(G')|/d(G')^2$, the maximum degree of $G'$ is at most

$$\Delta = \frac{dm^8}{600} \leq \frac{d(G')}{300} \cdot \log^8 \frac{30|V(G')|}{\varepsilon_2 d(G')^2} \leq \frac{d(G')}{300}\left(2\log\frac{|V(G')|}{d(G')^2}\right)^8 \leq d(G')\log^8\frac{|V(G')|}{d(G')^2}.$$

Slightly abusing the notation, we work in the future only with $G'$. We will rename $G'$ as $G$, relabelling $n = |V(G')|$ and $d = d(G')$, and by changing $\varepsilon_1$ to $\varepsilon_1/8$ and $\varepsilon_2$ to $4\varepsilon_2$, we assume that $G$ is $(\varepsilon_1, \varepsilon_2 d^2)$-expander and its maximum degree is at most $d\log^8(n/d^2)$. This completes the reduction step, i.e., to prove Theorem 2.1.2 it is sufficient to prove Theorem 2.2.5.

## 2.4 Dense case of Theorem 2.2.5

In this section, we prove the following lemma, which covers the dense case of Theorem 2.2.5.

**Lemma 2.4.1.** *Let $0 < \varepsilon_1 \leq \varepsilon_0$ and $\varepsilon_2 > 0$, where $\varepsilon_0$ is the constant from Corollary 2.2.2. Let $G$ be an $n$-vertex bipartite, $\{C_4, C_6\}$-free, $(\varepsilon_1, \varepsilon_2 d^2)$-expander graph with average degree $d \geq \log^{14} n$, $\delta(G) \geq d/4$ and $\Delta(G) \leq d\log^8 n$. Then $G$ contains a $TK_{\ell/2}$ for $\ell = cd$ for some constant $c > 0$ independent of $d$.*

Let $G$ be a graph satisfying the conditions in Lemma 2.4.1. Choose a constant $c > 0$ such that $c \ll \varepsilon_1$ and set $\ell = cd$. In addition, set the parameters in this section as follows:

$$\Delta = d\log^8 n, \quad \Delta'' = d\log^{13} n, \quad b = \frac{d}{\log^9 n}, \quad diam = \frac{2}{\varepsilon_1}\log^3\left(\frac{15n}{\varepsilon_2 d^2}\right) \le \frac{1}{c}\log^3 n.$$

Note that $\Delta \gg d \gg b$, $\Delta'' = o(d^2)$, and $\ell/b \le d/b = \log^9 n$.

We will first find $\ell$ vertices, $v_1, \dots, v_\ell$ serving as core vertices, along with some sets $B_3(v_i) \subseteq B(v_i, 3)$. We then connect all core vertices by linking $B_3(v_i)$'s using a greedy algorithm. Similarly to the proof in Section 2.3, we might discard few core vertices during the process.

### 2.4.1 Choosing core vertices and building $B_3(v_i)$

We will select $\ell$ vertices $v_1, \dots, v_\ell$ in $\ell/b$ steps to serve as core vertices.

**Stage 1:** We choose core vertices $v_1, \dots, v_\ell$ and the sets $B_2(v_i)$'s.

In each step, we choose a block of vertices consisting of: $b$ core vertices and for each core vertex $v_i$ a set $B_2(v_i) := S_1(v_i) \cup S_2(v_i) \cup \{v_i\}$, where $S_1(v_i) \subseteq S(v_i, 1)$ and $S_2(v_i) \subseteq S(v_i, 2)$ with the following properties:

(i) $S_1(v_i)$'s are pairwise disjoint for all $1 \le i \le \ell$ and $|S_1(v_i)| = d/2$.

(ii) For every $i$, $|S_2(v_i)| = d^2/10$.

(iii) Every vertex $w \in S_1(v_i)$ has at most $d/4$ neighbors in $S_2(v_i)$.

(iv) Inside each block, the sets $B_2(v_i)$'s are pairwise disjoint.

(v) Every $S_2(v_i)$ is disjoint from $\cup_{j=1}^{\ell} S_1(v_j)$.

(vi) For every $i \ne j$, $v_i \notin B_2(v_j)$.

To achieve this, we first choose a core vertex $v_i$ with sets $S_1(v_i)$ of size $d/2$ and $S'_2(v_i) \subseteq S(v_i, 2)$ of size $d^2/8 - d/2$ for all $i \le \ell$. We then choose $S_2(v_i) \subseteq S'_2(v_i)$. Suppose we have chosen some core vertices $v_1, v_2, \dots, v_{i-1}$ and sets $S_1(v_j)$ and $S'_2(v_j)$'s for $j \le i-1$. Denote by $D$ the current block and let $B_1(v_j) := S_1(v_j) \cup \{v_j\}$, $j \le i-1$. To choose the next core vertex $v_i$, we will exclude $\{\bigcup_{j \le i-1} B_1(v_j)\} \cup \{\bigcup_{v_k \in D} S'_2(v_k)\}$. The number of excluded vertices is at most

$$\sum_{j \le i} |B_1(v_j)| + b \cdot \max_{v_k \in D} |S'_2(v_k)| \le \ell d + b \cdot d^2/2 \le b \cdot d^2.$$

The number of the edges incident to the excluded vertices is at most

$$\Delta \cdot b \cdot d^2 = \frac{d^4}{\log n} \ll \frac{dn}{2} = e(G),$$

the last inequality holds since $G$ is $C_6$-free and therefore $d = O(n^{1/3})$ (see [13]). Thus, we can easily find in $G$, excluding these vertices, a subgraph $G'$ with average degree at least $d/2$ and minimum degree at least $d/4$. We then choose $v_i$ to be any vertex in $G'$ of degree at least $d/2$. Choose $d/2$ neighbors of $v_i$ to be $S_1(v_i)$. Since $G$ is bipartite, for each vertex $u \in S_1(v_i)$, we can choose $d/4 - 1$ neighbors of $u$ not in $B_1(v_i)$. Again, by $C_4$-freeness, we have chosen $d^2/8 - d/2$ different vertices. Denote the resulting set $S'_2(v_i)$. Note that in the process above, for any $i > j$, the set $S_1(v_i)$ is chosen after $S'_2(v_j)$. Thus when choosing $S_1(v_i)$, vertices in $S'_2(v_j)$ could be included if $v_i$ is in a different block from $v_j$. Since $|S'_2(v_i) \setminus \cup_{j\le\ell} S_1(v_j)| \ge |S'_2(v_i)| - \ell \cdot d \ge d^2/10$, we choose a subset of $S'_2(v_i) \setminus \cup_{j\le\ell} S_1(v_j)$ of size exactly $d^2/10$ to be $S_2(v_i)$.

**Stage 2:** For each $1 \le i \le \ell$, choose $S_3(v_i)$ of size $d^3/50$ and $B_3(v_i)$.

For each vertex in $S_2(v_i)$, since $G$ is bipartite and $C_4$-free, we can choose $d/4 - 1$ of its neighbors not in $S_1(v_i) \cup S_2(v_i)$ and denote the resulting set $S'_3(v_i)$. Since $G$ is $C_6$-free, $|S'_3(v_i)| = |S_2(v_i)| \cdot (d/4 - 1) = d^3/40 - d^2/10$. Delete from $S'_3(v_i)$ any vertex in $\bigcup_{1\le j\le\ell} B_1(v_j)$. Since we delete at most $d^2$ vertices, we can choose a subset of size $d^3/50$ to be $S_3(v_i)$. Let $B_3(v_i) := B_2(v_i) \cup S_3(v_i)$.

### 2.4.2 Connecting core vertices

**Greedy Algorithm:** Now we will connect the $\ell$ core vertices pair by pair in an arbitrary order. For each pair $v_i$ and $v_j$, we will connect them with a path of length at most *diam* avoiding $\bigcup_{p\ne i,j} B_1(v_p)$.

**(I) Discard bad core vertices:**

Call a core vertex $v_i$ bad, if we use more than $\Delta''$ vertices from $S_2(v_i)$. Discard a core vertex as soon as it becomes bad. During the entire process, we use at most $\ell^2 \cdot diam$ vertices from previous connections. Since $B_2(v_i)$'s are pairwise disjoint inside each block, each of the excluded vertices can appear in at most $\ell/b$ many $S_2(v_i)$'s. Hence, the number of bad core vertices is at most:

$$\frac{\ell^2 \cdot diam \cdot (\ell/b)}{\Delta''} \le \frac{d^2 \cdot diam \cdot (\ell/b)}{d \log^{13} n} \le \frac{d \log^3 n \cdot \ell}{cb \log^{13} n} = \frac{\ell}{c \log n} \ll \ell/2.$$

**(II) Cleaning before connection:**

Assume that we have already connected some pairs of core vertices, and now we want to connect $v_i$ and $v_j$. Before we start connecting them, clean $B_3(v_i)$ (do the same for $B_3(v_j)$) in the following way. Notice that we have used in previous connections at most $\ell$ vertices in $S_1(v_i)$, at most $\Delta''$ vertices in $S_2(v_i)$ and at most $\ell^2 \cdot diam$ vertices in $S_3(v_i)$, since vertices in $S_1(v_i)$ were only used when connecting $v_i$ to other core vertices and $v_i$ is not bad. Also, delete those vertices that are no longer available, i.e., those adjacent to used ones. Call the resulting set $B_3'(v_i)$. Since every vertex in $S_k(v_i)$ for $k \in \{1, 2\}$ has at most $d/4$ neighbors in $S_{k+1}(v_i)$, we have deleted at most $\ell(1 + d/4 + d^2/16) + \Delta''(1 + d/4) + \ell^2 \cdot diam \ll d^3/100$ vertices. Thus $|B_3'(v_i)| \geq |B_3(v_i)| - d^3/100 \geq d^3/100$.

**(III) Connecting core vertices:**

We will connect $v_i$ and $v_j$ by a shortest path from $B_3'(v_i)$ to $B_3'(v_j)$ avoiding $\bigcup_{p \neq i,j} B_1(v_p)$ which is of size at most $d^2$. This path has length at most $diam$ if we do not break the robust diameter property. We then exclude this path for further connections. The number of excluded vertices from previous paths and from $\bigcup_{p \neq i,j} B_1(v_p)$ is at most $\ell^2 \cdot diam + d^2 \leq d^2 \log^3 n$. On the other hand, the number of vertices we are allowed to exclude without breaking the robust small diameter among $B_3'(v_i)$'s is

$$\frac{1}{4}|B_3'(v_i)|\varepsilon(|B_3'(v_i)|) \geq \frac{d^3}{400}\varepsilon(n) \geq \frac{\varepsilon_1 d^3}{400 \log^2 n} \gg d^2 \log^3 n.$$

Thus the robust diameter property is guaranteed during the entire process.

This completes the proof of Lemma 2.4.1, hence the dense case of Theorem 2.2.5.

## 2.5 Sparse case of Theorem 2.2.5

In this section, we will prove the sparse case of Theorem 2.2.5. Throughout this section $G$ will be a sparse graph satisfying the conditions in Theorem 2.2.5, i.e., an $n$-vertex bipartite $\{C_4, C_6\}$-free $(\varepsilon_1, \varepsilon_2 d^2)$-expander graph, with average degree $d \leq \log^{14} n$, $\delta(G) \geq d/4$ and $\Delta(G) \leq d \log^8 n$. We always use $n$ for $|V(G)|$ and $d$ for $d(G)$. Inspired by an idea from [43] together with a random sparsening trick, we will show that in the sparse case, either we can find in $G$ a 1-subdivision (i.e., each edge is subdivided once) of some graph $H$ with $d(H) = \Omega(d^2)$, or there is a sparse and "almost regular" expander subgraph $G_1$ in $G$. In the first case, we apply Theorem 2.1.1 to find a subdivision of $K_\ell$ in $H$, hence in $G$, with $\ell = \Omega(\sqrt{d(H)}) = \Omega(d)$. For the second case, we use the following result of Komlós and Szemerédi (Theorem 3.1 in [38]).

**Theorem 2.5.1.** *If $F$ is an $(\varepsilon_1, d(F))$-expander satisfying $d(F)/2 \leq \delta(F) \leq \Delta(F) \leq 72(d(F))^2$ and $d(F) \leq \exp\{(\log|V(F)|)^{1/8}\}$, then $F$ contains a copy of $TK_\ell$ with $\ell = \Omega(d(F))$.*

The following lemma will be useful.

**Lemma 2.5.2.** *Let $F = (X \cup Y, E)$ be a bipartite $C_4$-free graph. If $|X| = \Omega(d^2|Y|)$ and $\frac{e(F)}{|X|} = \Omega(\Delta(X))$, then $F$ contains a copy of $TK_\ell$ with $\ell = \Omega(d)$.*

*Proof.* In $F$, we call a path of length 2 with endpoints in $Y$ a *hat*. By the convexity of the function $f(x) = \binom{x}{2}$, we have that the total number of hats in $F$ is at least

$$\sum_{v \in X} \binom{\deg(v)}{2} \geq \frac{|X|}{3} \cdot \left(\frac{e(F)}{|X|}\right)^2.$$

By the pigeonhole principle, there exists a collection of hats $\mathcal{H}$ with distinct midpoints of size

$$|\mathcal{H}| \geq \frac{|X|}{3(\Delta(X))^2} \cdot \left(\frac{e(F)}{|X|}\right)^2 = \Omega(|X|) = \Omega(d^2|Y|).$$

Define a graph $H$ on vertex set $Y$, where two vertices $y, y' \in Y$ are adjacent if there is a hat in $\mathcal{H}$ with $y, y'$ as endpoints. Note that since $F$ is $C_4$-free, any two hats have different sets of endpoints. Hence, each hat in $\mathcal{H}$ gives rise to a distinct edge in $H$. Thus

$$d(H) = \frac{2e(H)}{|Y|} = \frac{2|\mathcal{H}|}{|Y|} = \Omega(d^2).$$

Since the hats in $\mathcal{H}$ have distinct midpoints, there is a 1-subdivision of $H$ in $F$ with core vertices in $Y$ and hats in $\mathcal{H}$ served as subdivided edges. We then apply Theorem 2.1.1 to find a subdivision of $K_\ell$ in $H$, hence in $F$, with $\ell = \Omega(\sqrt{d(H)}) = \Omega(d)$. $\square$

Let $B := \{v \in V(G) : \deg_G(v) \geq d^3\}$ and $A := V(G) \setminus B$. Note that $|B| \leq \frac{d \cdot |V(G)|}{d^3} = \frac{n}{d^2}$, hence $|A| = |V(G)| - |B| \geq \frac{9n}{10}$. We first show that we may assume that there is a $G' \subseteq G$ with $|V(G')| = \Omega(n)$, $d(G') = \Theta(d)$ and $\Delta(G') \leq d^3$.

**Lemma 2.5.3.** *We can find in $G$ either a $TK_\ell$ with $\ell = \Omega(d)$, or there is a $G' \subseteq G$ with $|V(G')| \geq 9n/10$, $d/20 \leq d(G') \leq d$ and $\Delta(G') \leq d^3$. In the later case, there is a set $A' \subseteq V(G')$ such that $|A'| \geq |V(G')|/2$ and for any $v \in A'$, $\deg_{G'}(v) \geq d/10$.*

*Proof.* Define $G' := G[A]$, $A' := \{v \in A : \deg_{G'}(v) \geq d/10\}$ and $A'' := A \setminus A'$. We distinguish two cases based on the sizes of $A'$ and $A''$.

**Case 1:** Assume $|A''| \geq |A|/2$. Then $|A''| \geq 9n/20 = \Omega(d^2|B|)$. Note that, by the definition of $A''$, for any $a \in A''$, we have $\deg_{G[A'',B]}(a) \geq \delta(G) - \deg_{G'}(a) \geq d/4 - d/10 \geq d/10$. We bound in $G[A'', B]$ the degree of vertices in $A''$ as follows: for each $a \in A''$ with more than $d$ edges to $B$, keep exactly $d$ of them and delete the rest. Let the resulting graph be $G''$. Then in $G''$, $\Delta(A'') \leq d$, hence $\frac{e(G'')}{|A''|} \geq \delta(A'') \geq d/10 = \Omega(\Delta(A''))$. Applying Lemma 2.5.2 to $G''$ gives the first alternative of the conclusion of Lemma 2.5.3.

**Case 2:** Assume $|A'| \geq |A|/2$. The graph $G'$ was obtained from $G$ by removing vertices of degree at least $d^3$ (which were in $B$), thus $d(G') \leq d$. On the other hand, by the definition of $A'$, we have $d(G') \geq \frac{|A'| \cdot d/10}{|A|} \geq d/20$ and $\Delta(G') \leq d^3$ as desired. □

From now on, we will work only in $G' = G[A]$ with the properties listed in Lemma 2.5.3. For the rest of the proof in this section, we fix sufficiently large constants $C' \ll C \ll K$ and a small constant $c_0 \leq \frac{1}{1000}$.

Let $W := \{v \in V(G') : \deg_{G'}(v) \geq c_0 d^2\}$, and $U := V(G') \setminus W$. Note that $|W| \leq \frac{d(G') \cdot |V(G')|}{c_0 d^2} \leq \frac{n}{c_0 d}$, hence $|U| = |A| - |W| \geq \frac{4n}{5}$.

**Lemma 2.5.4.** *We can find in $G'$ either a $TK_\ell$ with $\ell = \Omega(d)$, or there exist vertex sets $U_0 \subseteq U$ and $W_0 \subseteq W$ with $|U_0| \geq |U|/6$ and $|W_0| \leq 2C|W|/d$ such that $G'[U_0, W_0]$ has at least $C'|U_0|$ edges and every vertex in $U_0$ has degree at most $K$ in $G'[U_0, W_0]$.*

We first show how Lemma 2.5.4 completes the proof of the sparse case of Theorem 2.2.5. Let $U_0, W_0$ be sets with properties listed in Lemma 2.5.4. Note that $|U_0| = \Omega(d^2|W_0|)$. Denote by $G_0 := G'[U_0, W_0]$. Recall that $\Delta(U_0) = K = O(1)$, thus $\frac{e(G_0)}{|U_0|} \geq C' = \Omega(\Delta(U_0))$. Applying Lemma 2.5.2 to $G_0$ gives a copy of $TK_\ell$ with $\ell = \Omega(d)$. This completes the proof of the sparse case of Theorem 2.2.5.

*Proof of Lemma 2.5.4.* Recall that $A' \subseteq V(G')$ consists of vertices of degree at least $d/10$ in $G'$. Define $U' := \{v \in A' \cap U : \deg_{G'[U,W]}(v) \geq d/20\}$ and $U'' := \{A' \cap U\} \setminus U'$. By Lemma 2.5.3, $|A'| \geq \frac{|V(G')|}{2} = \frac{|U|+|W|}{2}$. Thus $|U'| + |U''| = |A' \cap U| \geq |A'| - |W| \geq \frac{|U|-|W|}{2} \geq \frac{2|U|}{5}$. We distinguish two cases based on the sizes of $U'$ and $U''$.

**Case 1:** $|U''| \geq |U|/5$. Note that for every $v \in U''$, by the definition of $U''$,

$$\deg_{G'[U]}(v) = \deg_{G'}(v) - \deg_{G'[U,W]}(v) \geq \frac{d}{10} - \frac{d}{20} = \frac{d}{20}.$$

Thus $d(G'[U]) \geq \frac{d/20 \cdot |U''|}{|U|} \geq d/100$ and by the definition of $U$ we have $\Delta(G'[U]) \leq c_0 d^2$. Then we apply Corollary 2.2.2 to $G'[U]$ and let $G_1$ be the resulting $(\varepsilon_1, \varepsilon_2 d(G_1)^2)$-expander subgraph with $\varepsilon_2 < 1/1000$, $d(G_1) \geq d(G'[U])/2 \geq d/200$, $\delta(G_1) \geq d(G_1)/2$ and $\Delta(G_1) \leq \Delta(G'[U]) \leq c_0 d^2$. Let $n_1 := |V(G_1)|$.

If $d(G_1) \geq \exp\{(\log n_1)^{1/8}\}$, then we apply Lemma 2.2.4 to $G_1$. Then either we have a copy of $TK_\ell$ with $\ell = \Omega(d)$, in which case we are done, or we obtain a subgraph $G_2 \subseteq G_1$ with $d(G_2) \geq d(G_1)/2 \geq d/400$, $\delta(G_2) \geq d(G_2)/4$ and $\Delta(G_2) \leq d(G_2) \log^8 \frac{|V(G_2)|}{d(G_2)^2}$, which is an $(\varepsilon_1/8, 4\varepsilon_2 d(G_2)^2)$-expander. Since $|V(G_2)| \leq n_1$, we have that $d(G_2) \geq d(G_1)/2 \gg \log^{14}|V(G_2)|$. Applying Lemma 2.4.1 to $G_2$ gives a $TK_\ell$ with $\ell = \Omega(d(G_2)) = \Omega(d)$.

We may now assume that $d(G_1) \leq \exp\{(\log n_1)^{1/8}\}$. We want to apply Theorem 2.5.1 to $G_1$ to get a $TK_\ell$ with $\ell = \Omega(d(G_1)) = \Omega(d)$. Recall that $d(G_1)/2 \leq \delta(G_1) \leq \Delta(G_1) \leq c_0 d^2 \leq 72 d(G_1)^2$, where the last inequality follows from $d(G_1) \geq d/200$ and $c_0 \leq 1/1000$. It suffices to check that $G_1$ is an $(\varepsilon_1, d(G_1))$-expander.

**Claim 2.5.5.** The graph $G_1$ is an $(\varepsilon_1, d(G_1))$-expander.

*Proof.* Recall that $G_1$ is bipartite, $C_4$-free and $(\varepsilon_1, \varepsilon_2 d(G_1)^2)$-expander. For any set $X$ of size $x \geq \varepsilon_2 d(G_1)^2/2$, $|\Gamma(X)| \geq x \cdot \varepsilon(x, \varepsilon_1, \varepsilon_2 d(G_1)^2) \geq x \cdot \varepsilon(x, \varepsilon_1, d(G_1))$, as $\varepsilon(x, \varepsilon_1, t)$ is an increasing function in $t$.

It is known that in $C_4$-free bipartite graphs of minimum degree $k$, any set of size at most $k^2/500$ expands by a rate of at least 2 (see e.g. Lemma 2.1 in [57]). Recall that $\delta(G_1) \geq d(G_1)/2$ and $\varepsilon_2 \leq 1/1000$, so $\varepsilon_2 d(G_1)^2/2 \leq 2\varepsilon_2 \delta(G_1)^2 \leq \frac{\delta(G_1)^2}{500}$. Since $\varepsilon(x, \varepsilon_1, d(G_1))$ is a decreasing function in $x$, for any $x \geq d(G_1)/2$, $\varepsilon(x, \varepsilon_1, d(G_1)) \leq \varepsilon(d(G_1)/2, \varepsilon_1, d(G_1)) = \frac{\varepsilon_1}{\log^2(7.5)} < 2$. Thus for any set $X$ of size $d(G_1)/2 \leq x \leq \varepsilon_2 d(G_1)^2/2 \leq \frac{\delta(G_1)^2}{500}$, we have $|\Gamma(X)| \geq 2x \geq x \cdot \varepsilon(x, \varepsilon_1, d(G_1))$ as desired. □

This gives the first alternative of the conclusion of Lemma 2.5.4.

**Case 2:** $|U'| \geq |U|/5 \geq \frac{4n/5}{5} \geq n/7$. Recall that $|W| \leq \frac{n}{c_0 d}$. Consider the subgraph $G_3 := G'[U', W]$, by deleting extra edges, we may assume that each vertex in $U'$ has degree at most $d$ in $W$. Then by the definition of $U'$, we have

$$\frac{d}{11} \leq \frac{2|U'| \cdot d/20}{|U'| + |W|} \leq d(G_3) \leq \frac{2|U'| \cdot d}{|U'| + |W|} \leq 2d.$$

Set $p := C/d$. We will choose a random subset $W_0 \subseteq W$, in which each element of $W$ is included with probability $p$ independent of each other. We then choose some $U_0 \subseteq U'$ consisting of vertices of degree at most $K$ in $W_0$. We will show that with positive probability, $W_0$ and $U_0$ have the desired properties. For simplicity, we define $G_4 := G_3[U', W_0]$.

We may assume that $|W| \geq \frac{n}{d^2}$, since otherwise $|U'| = \Omega(d^2|W|)$ and $\frac{e(G_3)}{|U'|} \geq \delta(U') \geq d/20 = \Omega(\Delta(U'))$. Then applying Lemma 2.5.2 to $G_3$ yields a $TK_\ell$ with $\ell = \Omega(d)$. Note that $\mathbb{E}|W_0| = p|W|$, by Chernoff's Inequality, w.h.p. $|W_0| \leq 2\mathbb{E}|W_0| = 2C|W|/d$. As mentioned above, we will delete vertices from $U'$ with degree more than $K$ in $W_0$ to form $U_0$. It suffices to show that w.h.p.

(i) $e(G_4) \geq 2C'|U'|$;

(ii) the number of vertices deleted (i.e., $U' \setminus U_0$) is at most $|U'|/10$ and the number of edges deleted (from $G_4$ to form $G_3[U_0, W_0] = G'[U_0, W_0]$) is at most $C'|U'|$.

It then follows that $|U_0| \geq 9|U'|/10 \geq |U|/6$ and the number of edges in $G_0 = G'[U_0, W_0]$ is at least $e(G_4) - C'|U'| \geq C'|U'| \geq C'|U_0|$ as desired.

For (i), recall that by Lemma 2.5.3, $\Delta(G_3) \leq d^3$. For each vertex $v_i \in W$, define a random variable $X_i$ taking value $\deg_{G_3}(v_i)$ if $v_i \in W_0$ and 0 otherwise. Then $e(G_4) = \sum_{i \leq |W|} X_i$ and

$$\mathbb{E}(e(G_4)) = \sum_{i \leq |W|} \mathbb{E}X_i = \sum_{v_i \in W} p \cdot \deg_{G_3}(v_i) = p \cdot e(G_3) \geq \frac{C}{d} \cdot \frac{d}{20} \cdot |U'| \geq 4C'|U'|.$$

Recall that $\frac{n}{d^2} \leq |W| \leq \frac{n}{c_0 d}$ and $d \leq \log^{14} n$. Using the independent bounded differences inequality 1.2.3 with $f(\mathbf{X}) = \sum X_i$, $\sigma_i = d^3$ and $t = \mathbb{E}(e(G_4))/2 \geq 2C'|U'| \geq \frac{2C'n}{7} \geq \frac{2C'}{7} \cdot c_0 d|W| \geq c_0 d|W|$, we have that

$$\mathbb{P}\left[e(G_4) \leq \frac{1}{2}\mathbb{E}(e(G_4))\right] \leq 2e^{-\frac{2(c_0 d|W|)^2}{d^6|W|}} = e^{-c_0^2|W|/d^4} \leq e^{-c_0^2 n/d^6} \to 0.$$

For (ii), for each $u_i \in U'$, we define a random variable $Y_i := \deg_{G_4}(u_i)$. Note that for any two vertices $u_i, u_j \in U'$, if they have no common neighbor in $W$, then $Y_i$ and $Y_j$ are independent. Define an auxiliary dependency graph $F$ on vertex set $\{Y_i\}_{i=1}^{|U'|}$, in which $Y_i$ and $Y_j$ are adjacent if and only if they are not independent. Since in $G_3$ every vertex in $U'$ has degree at most $d$ and every vertex in $W$ has degree at most $d^3$, it follows that $\Delta(F) \leq d^4$ and by Brook's theorem that $\chi(F) \leq d^4 + 1$. Thus we can partition $U'$ into $d^4 + 1$ classes, say $U' := Z_0 \cup Z_1 \cup \ldots \cup Z_{d^4}$, such that $Y_i$'s corresponding to vertices in the same class are independent. First we discard classes of size smaller than $n/d^6$, the number of vertices we delete at this step is at most $\frac{n}{d^6} \cdot (d^4 + 1) \ll |U'|$. Thus we may assume that each class is of size at least $n/d^6$. Fix a class $Z_j$, for every $v \in Z_j$ and every $i \geq K \gg C$,

$$\mathbb{P}[\deg_{G_4}(v) = i] = \binom{\deg_{G_3}(v)}{i} p^i (1-p)^{\deg_{G_3}(v) - i} \leq \frac{d^i}{i!} \cdot \frac{C^i}{d^i} \leq e^{-i \log i/2} := q_i.$$

For each $1 \leq i \leq d$, let $N_i$ ($N_{\geq i}$ resp.) be the number of vertices in $Z_j$ of degree $i$ (at least $i$ resp.) in $W_0$. Then $\mathbb{E}N_i \leq |Z_j|q_i$. For each $i \leq \log^2 d$, by Chernoff's Inequality and recall that $d \le \log^{14} n$, we have

$$\mathbb{P}[N_i \ge 2\mathbb{E}N_i] < \exp\{-|Z_j|q_i/3\} \ll \exp\left\{-\frac{n}{d^6} \cdot e^{-\log^3 d}\right\} \ll \exp\left\{-\frac{n}{e^{(\log\log n)^4}}\right\}. \qquad (2.2)$$

Note that for any $v \in Z_j$, $\mathbb{P}[\deg_{G_4}(v) \ge \log^2 d] \le \sum_{i=\log^2 d}^{d} q_i \ll e^{-\log^2 d}$. It follows that

$$\mathbb{P}[N_{\ge \log^2 d} \ge 2\mathbb{E}N_{\ge \log^2 d}] \ll \exp\left\{-|Z_j| \cdot e^{-\log^2 d}\right\} \ll \exp\left\{-\frac{n}{e^{(\log\log n)^3}}\right\}. \qquad (2.3)$$

By (2.2), (2.3) and the union bound, the probability that there exists a class $Z_j$ in which either $N_{\ge \log^2 d} \ge 2\mathbb{E}N_{\ge \log^2 d}$ or for some $i \le \log^2 d$, $N_i \ge 2\mathbb{E}N_i$ is at most

$$(d^4+1) \cdot \left(\log^2 d \cdot \mathbb{P}[N_i \ge 2\mathbb{E}N_i] + \mathbb{P}[N_{\ge \log^2 d} \ge 2\mathbb{E}N_{\ge \log^2 d}]\right) \to 0.$$

Note that $\sum_{K \le i \le \log^2 d} \mathbb{E}N_i \le \sum_{K \le i \le \log^2 d} q_i|Z_j| \ll e^{-K}|Z_j|$. Thus w.h.p. the number of vertices deleted is at most

$$\sum_j ((2 \sum_{K \le i \le \log^2 d} \mathbb{E}N_i + 2\mathbb{E}N_{\ge \log^2 d}) \cdot |Z_j|) \ll \sum_j (e^{-K} + e^{-\log^2 d}) \cdot |Z_j| < 2e^{-K}|U'| \ll |U'|.$$

The number of edges incident to vertices deleted in $Z_j$ is at most

$$\sum_{K \le i \le \log^2 d} (2q_i|Z_j| \cdot i) + (\sum_{i=\log^2 d}^{d} 2q_i|Z_j|) \cdot d \ll (e^{-K} + d \cdot e^{-\log^2 d}) \cdot |Z_j| < 2e^{-K}|Z_j|.$$

Recall that every vertex in $U'$ has degree at most $d$ in $W$ and that $|U'| \ge n/7$. Then summing over all classes, the total number of edges deleted is at most

$$\sum_{|Z_j| \ge n/d^6} 2e^{-K}|Z_j| + \sum_{|Z_k| \le n/d^6} d \cdot |Z_k| \le 2e^{-K}|U'| + (d^4+1) \cdot d \cdot \frac{n}{d^6} \ll |U'|.$$

□

## 2.6 Proof of Theorem 2.1.4

In this section, we will prove Theorem 2.1.4 using a variation of the Dependent Random Choice Lemma (see survey [27] for more details on the method of dependent random choice). The following lemma roughly says that in a dense $C_4$-free graph one can find a set in which every small subset has a large second common neighborhood.

**Lemma 2.6.1.** *Let $G = (A \cup B, E)$ be a $C_4$-free bipartite graph on $n$ vertices with $cn^{3/2}$ edges and $|A| = |B| = \frac{n}{2}$, where $n > 1/c^{20}$. If there exist positive integers $a$, $m$, $r$ and $t$ such that*

$$c^{2t}n - \binom{n}{r}\left(\frac{m}{n/2}\right)^t \geq a, \tag{2.4}$$

*then there exists $U \subseteq A$ with at least $a$ vertices such that for every $r$-subset $S \subseteq U$, $|N_2(S)| \geq m$.*

*Proof.* First notice that

$$\begin{aligned}\sum_{v\in A}|N_2(v)| &= \sum_{v\in B}(d(v)-1)d(v) = \sum_{v\in B}d(v)^2 - \sum_{v\in B}d(v) \geq \frac{n}{2}\left(\frac{\sum_{v\in B}d(v)}{n/2}\right)^2 - e(G)\\ &= \frac{n}{2}(2cn^{1/2})^2 - cn^{3/2} \geq c^2n^2.\end{aligned}$$

Pick a set $T \subseteq A$ of $t$ vertices uniformly at random with repetition. Let $W := N_2(T) \subseteq A$ and put $X := |W|$. Then by the linearity of expectation and $t \geq 1$, we have

$$\begin{aligned}\mathbb{E}[X] &= \sum_{v\in A}\mathbb{P}(v \in N_2(T)) = \sum_{v\in A}\left(\frac{|N_2(v)|}{n/2}\right)^t = \left(\frac{2}{n}\right)^t\cdot\frac{n}{2}\cdot\left(\frac{1}{n/2}\sum_{v\in A}|N_2(v)|^t\right)\\ &\geq \left(\frac{n}{2}\right)^{1-t}\cdot\left(\frac{\sum_{v\in A}|N_2(v)|}{n/2}\right)^t \geq \left(\frac{n}{2}\right)^{1-t}\cdot(2c^2n)^t = 2^{2t-1}c^{2t}n \geq c^{2t}n.\end{aligned}$$

Let $Y$ be the random variable counting the number of $r$-sets in $W$ that have fewer than $m$ common second neighbors. The probability for a fixed such $r$-set $S$ to be in $W$ is at most $\left(\frac{m}{n/2}\right)^t$. There are at most $\binom{n}{r}$ $r$-sets, hence

$$\mathbb{E}[X - Y] \geq c^{2t}n - \binom{n}{r}\left(\frac{m}{n/2}\right)^t \geq a.$$

Thus there exists a choice of $T$, such that $X - Y \geq a$. Delete one vertex from $X$ for each such "bad" $r$-set from $W$, and the resulting set $U$ has the desired property. □

**Claim 2.6.2.** When proving Theorem 2.1.4, we may assume that $G$ is bipartite on $A \cup B$ with $|A| = |B| = n/2$, $d(G) = d$ and all vertices in $B$ have degree smaller than $30d$.

*Proof.* We may assume that for any $H \subseteq G$, $d(H) \leq d$, otherwise we can work in $H$ instead. Let $X \subseteq V$ be the set of vertices of degree at least $10d$, thus $|X| \leq n/10$. Let $Y = V \setminus X$. Since $d(G[X]) \leq d$, we have $e(G[X]) \leq d|X|/2 \leq e(G)/10$. Take an $\frac{n}{2}$-subset $B$ of $Y$

uniformly at random and call $V \setminus B = A$. Then we have,

$$\mathbb{E}(e(G[A,B])) \geq 0.4[e(G[Y]) + e(G[X,Y])] = 0.4[e(G) - e(G[X])] \geq 0.36e(G).$$

Therefore there exists a choice of $A, B$ such that $e(G[A,B]) \geq 0.36e(G)$. Hence we can replace $G$ by $G' := G[A,B]$, and every vertex in $B$ has degree less than $10d \leq 10\cdot(d(G')/0.36) < 30d(G')$. $\square$

*Proof of Theorem 2.1.4.* Assume $G$ satisfies the conditions of Claim 2.6.2 and apply Lemma 2.6.1 to $G$ with the following parameters:

$$a = \frac{c^6 n^{1/2}}{240}, \quad r = 2, \quad t = \frac{\log n}{4\log(1/c)}, \quad m = \frac{c^6 n}{2}.$$

In order to prove that (2.4) is satisfied, we shall prove $2\binom{n}{2}\left(\frac{m}{n/2}\right)^t \leq c^{2t}n$ and $c^{2t}n \geq 2a$. Indeed,

$$2\binom{n}{2}\left(\frac{m}{n/2}\right)^t \leq c^{2t}n \quad \Leftarrow \quad n \leq \left(\frac{c^2 n/2}{m}\right)^t = \left(\frac{1}{c}\right)^{4t} \Leftarrow \log n \leq 4t \cdot \log\frac{1}{c} = \log n.$$

On the other hand, we have

$$c^{2t}n \geq 2a = \frac{c^6 n^{1/2}}{120} \Leftrightarrow \frac{120 n^{1/2}}{c^6} \geq \left(\frac{1}{c}\right)^{2t} \Leftarrow \log 120 + \frac{1}{2}\log n + 6\log\frac{1}{c} \geq 2t\log\frac{1}{c} = \frac{1}{2}\log n.$$

Thus there exists $U \subseteq A$ of size at least $a = \frac{c^6 n^{1/2}}{240}$ such that for every pair of vertices $S \subseteq U$, $|N_2(S)| \geq m = c^6 n/2$.

We embed a copy of $TK_\ell$ with $\ell = a = c^5 d/480$ greedily as follows: first embed all the core vertices arbitrarily to $U$. Then we connect all pairs of core vertices one by one, in an arbitrary order, with internally vertex-disjoint paths of length 4. Fix a pair of vertices $S \subseteq U$. For every vertex $v$ in $N_2(S)$, call $C(v) := N(v) \cap \Gamma(S)$ its *connector set* and call $v$ "bad" if $|C(v)| = 1$. Since $G$ is $C_4$-free, $|N_1(S)| \leq 1$, so there are at most $\Delta(B) \leq 30d$ bad vertices in $N_2(S)$. Any vertex $v \in N_2(S)$ that is not bad has $|C(v)| = 2$. When connecting $S$, we will exclude from $N_2(S)$ the following vertices: (i) bad vertices (if they exist); (ii) vertices in $U$; (iii) vertices that were already used in previous connections; (iv) vertices whose connector set was used. It follows immediately that if there is a vertex left in $N_2(S)$, then together with its connector set, we can connect $S$.

For (i) and (ii), recall that there are at most $30d$ bad vertices and $|U| \leq \ell$. For (iii), there are at most $\binom{\ell}{2}$ such vertices, one for each pair of core vertices. Thus there are at least

$m - 30d - \ell - \binom{\ell}{2} \geq c^6n/2 - 60cn^{1/2} - \ell^2 \geq c^6n/4$ many vertices left in $N_2(S)$.

For (iv), we say that two vertices in $N_2(S)$ have no *conflict* with each other if their connector sets are disjoint. Notice that every vertex $v$ in $N_2(S)$ that is not bad can have a conflict with at most $|C(v)| \cdot \Delta(B) = 2\Delta(B) \leq 60d$ vertices. Thus we can find at least

$$\frac{c^6n/4}{2\Delta(B)} \geq \frac{c^6n}{240d} = \frac{c^6n}{480cn^{1/2}} = \frac{c^5n^{1/2}}{480} \geq 2\ell$$

not-previously-used vertices in $N_2(S)$ that are pairwise conflict-free. Again since $G$ is $C_4$-free, any other core vertex in $U \setminus S$ can be adjacent to connector sets of at most 2 vertices in $N_2(S)$. Thus there are at least $2\ell - 2(\ell - 2) = 4$ vertices available in $N_2(S)$ to connect the pair of vertices in $S$. □

## 2.7 Concluding Remarks

The proof of Theorem 2.1.3 is almost identical to the proof of Theorem 2.1.2. The only differences is to generalize Lemma 2.4.1 to $\{C_4, C_{2k}\}$-free graphs for any $k \geq 4$. First we need a result of Kühn and Osthus [42], which finds a $C_4$-free subgraph $G'$ in a $C_{2k}$-free graph $G$ for $k \geq 4$ such that $d(G') = \Omega(d(G))$. Then after cleaning $S_1(v_i)$ and $S_2(v_i)$(as in Section 2.4.2), $S_2(v_i)$ still has $\Omega(d^2)$ vertices. Recall that each vertex in $S_2(v_i)$ sends $\Omega(d)$ edges to $S_3(v_i)$, then by a well-known result of Bondy and Simonovits [13], we have that there are at least $\Omega(d^{3-3/(k+1)})$ vertices available in $S_3(v_i)$ after cleaning $S_1(v_i)$ and $S_2(v_i)$. We further clean $S_3(v_i)$ by deleting at most $\ell^2 \cdot diam$ vertices. For $k \geq 4$, $d^{3-3/(k+1)}\varepsilon(d^{3-3/(k+1)}) \gg \ell^2 \cdot diam + d^2$, thus the robust diameter property is guaranteed for all connections.

# Chapter 3

# Triangle factors

The classical Corrádi-Hajnal theorem claims that every $n$-vertex graph $G$ with $\delta(G) \geq 2n/3$ contains a triangle factor, when $3|n$. In this paper we asymptotically determine the minimum degree condition necessary to guarantee a triangle factor in graphs with sublinear independence number. In particular, we show that if $G$ is an $n$-vertex graph with $\alpha(G) = o(n)$ and $\delta(G) \geq (1/2 + o(1))n$, then $G$ has a triangle factor and this is asymptotically best possible. We also propose many related open problems whose solutions could show a relationship with Ramsey-Turán theory.

Additionally, we also consider a fractional variant of the Corrádi-Hajnal Theorem, settling a conjecture of Balogh-Kemkes-Lee-Young. Let $t \in (0, 1)$ and $w : E(K_n) \to [0, 1]$. We call a triangle in $K_n$ heavy if the sum of the weights on its edges is more than $3t$. We prove that if $3|n$ and $w$ is such that for every vertex $v$ the sum of $w(e)$ over edges $e$ incident to $v$ is at least $\left(\frac{1+2t}{3} + o(1)\right) n$, then there are $n/3$ vertex disjoint heavy triangles in $G$.

## 3.1 Introduction

Given an $n$-vertex graph $G$ and an $h$-vertex graph $H$, an *$H$-tiling* is a collection of vertex disjoint copies of $H$ in $G$. A *perfect $H$-tiling* or an *$H$-factor* is an $H$-tiling that covers all of the vertices of $G$. One obvious necessary condition for an $H$-factor in $G$ is $h|n$. Throughout the rest of the paper we will assume that this divisibility condition holds whenever necessary. We also always assume that $n$ is sufficiently large.

For a given graph $H$, a fundamental problem in graph theory is to find sufficient conditions for a graph $G$ to have an $H$-factor. A classical result of Tutte gives necessary and sufficient conditions for the case $H = K_2$. Another celebrated result of this type is the Hajnal-Szemerédi Theorem [29] which states that every $n$-vertex graph $G$ with $\delta(G) \geq (1 - 1/r)n$ has a $K_r$-factor. The case $r = 3$ was proved earlier by Corrádi and Hajnal [14]. The almost balanced complete $r$-partite graph on $n$ vertices shows that the minimum degree condition in the Hajnal-Szemerédi theorem is sharp. This extremal example, which is very similar to the Turán graph, has chromatic number $r$, has an independent set of size greater than $n/r$, it is almost regular and very far from random-like.

Although the Hajnal-Szemerédi Theorem was proved many years ago, there has been significant recent activity on related theorems. For example, Alon-Yuster [4], Komlós-Sárközy-Szemerédi [37] and Kühn-Osthus [45] have all proved theorems similar to the Hajnal-Szemerédi Theorem where complete graphs factors are replaced with $H$-factors where $H$ is an arbitrary graph; Kierstead-Kostochka proved the Hajnal-Szemerédi Theorem with an Ore-type degree condition [34]; Fischer [26], Martin-Szemerédi [50], and Keevash-Mycroft [32] have proved multipartite variants; and Wang [62], Keevash-Sudakov [33], Czygrinow-Kierstead-Molla [16], Czygrinow-DeBiasio-Kierstead-Molla [15], Treglown [60] and Balogh-Lo-Molla [9] have all proved analogues of the Hajnal-Szemerédi Theorem in directed and oriented graphs.

In 1970, Erdős and Sós [24] began studying a variation on Turán's theorem that excludes graphs with high independence number such as Turán graph. They investigated the maximum number of edges in an $n$-vertex, $K_r$-free graph with independence number $o(n)$. These types of problems became known as Ramsey-Turán problems, and have been studied extensively over the past 40 years, see for example [7] [21] [22] [53] [56]. The following question is a Ramsey-Turán-type of variant of the Hajnal-Szemerédi theorem.

**Question 3.1.1.** Let $G$ be an $n$ vertex graph with $\alpha(G) = o(n)$. What is the minimum degree condition on $G$ that guarantees a $K_k$-factor in $G$ for $k \geq 3$?

As we mentioned earlier, the main motivation for Question 3.1.1 is the fact that the extremal example for the Hajnal-Szemerédi theorem is a very structured graph. In 2004, Krivelevich-Sudakov-Szabó [41] considered the pseudo-random version of the Corrádi-Hajnal theorem. In particular, they proved that every $n$-vertex graph $G$ satisfying some pseudo-random conditions has a triangle-factor. The pseudo-random condition they require implies $\alpha(G) = o(n)$. In fact, their condition implies that the graph has uniform edge distribution, a much stronger condition, in Question 3.1.1, we impose a much weaker hypothesis, though for this price we need a higher minimum degree condition. Our first main result is to answer Question 3.1.1 for $k = 3$.

**Theorem 3.1.2.** *For every $\varepsilon > 0$, there exists $\gamma > 0$ and $n_0$ such that the following holds. For every $n$-vertex graph $G$ with $n > n_0$, if $\delta(G) \geq (1/2 + \varepsilon)n$ and $\alpha(G) \leq \gamma n$, then $G$ has a triangle factor.*

The following examples show that the minimum degree condition in the statement of Theorem 3.1.2 is asymptotically best possible. For $n = 2k$, consider the graph $G = K_{k-1} \cup K_{k+1}$. This graph does not have a triangle factor and $\delta(G) = n/2 - 2$. Another example for $n = 2k$

is the following. Consider the graph consisting of $K_{k+2}$ and $K_{k-1}$ sharing one vertex. Since $3|2k$, we have that both $k+2 \equiv 2 \pmod 3$ and $k-1 \equiv 2 \pmod 3$. Hence, this graph has no triangle factor and $\delta(G) = n/2 - 2$. For $n = 2k+1$ consider the graph consisting of two copies of $K_{k+1}$ sharing one vertex. Since $3|2k+1$, we have $k+1 \equiv 2 \pmod 3$. Hence, this graph has no triangle factor and $\delta(G) = (n-1)/2$.

We also prove the triangle case of the conjecture proposed by Balogh-Kemkes-Lee-Young ([6], Conjecture 1). Let $t \in (0,1)$ and $w : E(K_n) \to [0,1]$. We call $x, y, z \in V(K_n)$ a *heavy triangle* if $w(xy) + w(xz) + w(yz) > 3t$, for any $v \in V(K_n)$, we let $d_w(v)$ be the sum of the weights on the edges incident to $v$ and let $\delta_w(K_n) = \min_{v \in V(K_n)} d_w(v)$.

**Theorem 3.1.3.** *For any $t \in (0,1)$ and $\varepsilon > 0$ there exists $n_0$ such that for $3k = n \geq n_0$, if $w : V(K_n) \to [0,1]$ is such that $\delta_w(K_n) \geq \left(\frac{1+2t}{3} + \varepsilon\right) n$ then there are $k$ vertex-disjoint heavy triangles in $G$.*

This theorem is asymptotically best possible for any $t \in (0,1)$ by the following example from [6]. Let $n$ be divisible by 3, let $U \subseteq V(K_n)$ such that $|U| = 2n/3 + 1$ and, for all $e \in E(G)$, set $w(e) = t$ if $e \in E(G[U])$, and otherwise set $w(e) = 1$. Since every heavy triangle intersects $U$ in at most two vertices, there are no $n/3$ vertex disjoint heavy triangles in $G$. Furthermore, we have that $\delta_w(G) = |V(G) \setminus U| + t(|U| - 1) = (1+2t)n/3 - 1$.

As was pointed out in [6], when $t = 2/3$ and $w(e) \in \{0,1\}$ for every $e \in E$, the Corrádi-Hajnal Theorem implies that $G$ has a heavy triangle factor when $\delta_w(G) \geq 2n/3$. It is interesting to note that when $w(e)$ is allowed to take any value in $(0,1)$ we can show that we must force $\delta_w(G)$ to be greater than $7n/9 - 1$ to guarantee a heavy triangle factor by replacing $t$ with $2/3$ in the example above.

**Notation.** Most of the notation that we use is standard. For a collection of subsets $\mathcal{U}$ of $G$ we let $V(\mathcal{U}) := \bigcup_{U \in \mathcal{U}} U$. Similarly, for a collection of subgraphs $\mathcal{U}$ we let $V(\mathcal{U}) := \bigcup_{U \in \mathcal{U}} V(U)$. For any $v \in V$ and $U \subseteq V$, we let $d_U(v) = d(v, U)$ be the number of edges incident to $v$ and a vertex in $U$. For $U, W \subseteq V$, we let $e_G(U, W) := \sum_{u \in U} d(u, W)$.

We use the notion of a multiset in several places, and when $U$ is a multiset, we write $\nu_U(u)$ to represent the multiplicity of the element $u \in U$.

The notation $a \ll b$ means that there exists an increasing function $f$ such that when $a$ and $b$ are constants and $a \leq f(b)$ the argument holds. The function $f$ is not always explicitly specified, but could be computed.

**Outline of the chapter.** We first introduce and prove all the tools for the absorbing method in Section 3.2. In Section 3.3 we prove Theorem 3.1.2. In Section 3.3.1, we state the two main lemmas and show how they imply Theorem 3.1.2. Then, in Sections 3.3.2 and 3.3.3, we prove the two main lemmas of Theorem 3.1.2. In Section 3.4, we prove Theorem 3.1.3.

## 3.2 Tools for the absorbing method

The absorbing method of Rödl, Ruciński and Szemerédi [54] is used in the proofs of both Theorem 3.1.2 and Theorem 3.1.3, and the results of this section are used in both of the proofs. When reading this section in the context of Theorem 3.1.3, all references to triangles should be interpreted as references to heavy triangles.

The proof of the absorbing lemma for Theorem 3.1.3 (Lemma 3.4.1), while non-trivial, is standard within the context of the absorbing method. However, the absorbing lemma for Theorem 3.1.2 (Lemma 3.3.1) is more involved. The framework for the proof of Lemma 3.3.1 is established in this section. This framework will also be used in the proof of Lemma 3.4.1, but most of it is not necessary for Theorem 3.1.3.

The main problem we had in applying a standard argument to create an absorbing lemma for Theorem 3.1.2 is that there does not necessarily exists $k \in \mathbb{N}$ such that for every set of 3-vertices $W$ there exist $\Omega(n^{3k})$ sets $U$ of size $3k$ such that both $G[U]$ and $G[W \cup U]$ have perfect triangle factors. Below, we construct a graph to demonstrate this property.

**Example 3.2.1.** Fix $k \in \mathbb{N}$ and $0 < \varepsilon < 1/6$. Let $V_1, V_2, \ldots, V_{2m+1}$ be disjoint sets that partition $[n]$ where $|V_1| = \lfloor (1/2 - \varepsilon)n \rfloor$ and $|V_2|, \ldots, |V_{2m+1}| \geq \lceil 2\varepsilon n \rceil$. Note that $m$ can be as large as $\left\lfloor \frac{\lceil n/2+\varepsilon n \rceil}{2\lceil 2\varepsilon n \rceil} \right\rfloor \geq \varepsilon^{-1}/8$. Let $G'$ be the graph on $[n]$, where for every $i \in [m]$ we add all possible edges between $V_1$, $V_{2i}$, $V_{2i+1}$, i.e. $G'[V_1 \cup V_{2i} \cup V_{2i+1}]$ is the complete 3-partite graph with parts $V_1$, $V_{2i}$ and $V_{2i+1}$ for every $i \in [m]$. Note that $\delta(G') \geq (1/2 + \varepsilon)n$, and every triangle in $G'$ has exactly one vertex in $V_1$. We obtain $G$ by adding edges inside $V_i$ for every $i \in [2m+1]$ so that $d_G(v, V_i) = o(n)$ for every $v \in V_i$ and $\alpha(G[V_i]) = o(n)$. It is well-known that, with high probability, if every possible edge in $G[V_i]$ is selected with probability $\frac{\log n}{n} = p = o(1)$, then $G$ will have the desired properties. Let $G'' := G - G'$.

**Claim 3.2.2.** For every fixed $k$, there exists a 3-set $W \subseteq V$ such that there are only $o(n^{3k})$ sets $U \subseteq V$ of size $3k$ such that both $G[U]$ and $G[U \cup W]$ have a triangle factor.

*Proof.* Let $\{w_1, w_2, w_3\} := W \subseteq V \setminus V_1$ such that $W$ is an independent set and $|W \cap (V_{2i} \cup V_{2i+1})| \neq 3$ for any $i \in [m]$. Let $U \subseteq V$ such that $G[U]$ has a triangle factor $\mathcal{T}_1$ and $G[U \cup W]$ has a triangle factor $\mathcal{T}_2$. If $E(G''[U \cup W]) = \emptyset$, then every $T \in \mathcal{T}_1 \cup \mathcal{T}_2$ has exactly one vertex in $V_1$, so $|U \cap V_1| = k$ and $|(W \cup U) \cap V_1| = k + 1$, but this contradicts the fact that $W \cap V_1 = \emptyset$. Therefore, $E(G''[U \cup W]) \neq \emptyset$, but there are only $o(n^{3k})$ sets $U \subseteq V$ of size $3k$ such that $G''[U \cup W]$ contains an edge. □

**Definition 3.2.3.** Let $G(V, E)$ be an $n$-vertex graph. Distinct vertices $x, y \in V$ are $(c, k)$-*linked* if there are at least $(cn)^{3k-1}$ multisets $U \subseteq V$ of size $3k-1$ such that the following holds. Let $U'$ be the set of elements of $U$, without repetition. Then, both $G[U' \cup \{x\}]$ and $G[U' \cup \{y\}]$ have triangle factors in the following sense: if a vertex in $U$ has multiplicity $i$ then it should be in exactly $i$ triangles. We also call $U$ a *$k$-linking set* for $\{x, y\}$.

For a vertex $v \in V$, denote by $L_{c,k}(v)$ the set of vertices that are $(c, k)$-linked with $v$. A set $V' \subseteq V$ is $(c, k)$-*linked* if every pair of vertices in $V'$ are $(c, k)$-linked.

**Definition 3.2.4.** For $k \in N$ and $0 < \phi < \psi \leq 1$, call a partition $\mathcal{M} = \{V_1, \ldots, V_d\}$ of $V$ $(\psi, \phi, k)$-*linked* if $|V_i| \geq \psi n$ and $V_i$ is $(\phi, k)$-linked for every $i \in [d]$. Note that $d \leq 1/\psi$.

In Example 3.2.1, for every $i \in [2m+1]$, $V_i$ is $(\varepsilon, 1)$-linked, in particular, $\{V_1, \ldots, V_{2m+1}\}$ is a $(2\varepsilon, \varepsilon, 1)$-linked partition of $G$.

**Claim 3.2.5.** Consider the graph from Example 3.2.1. For any $k \in \mathbb{N}$ and $\phi > 0$, if $v_i \in V_i$ and $v_j \in V_j$ where $i \neq j$, then $v_i$ and $v_j$ are not $(\phi, k)$-linked.

*Proof.* We show that there are $o(n^{3k-1})$ sets $U$ that are $k$-linking multiset for $\{v_i, v_j\}$. Let $U$ be such a multiset. Since there are only $o(n^{3k-1})$ multisets of order $3k-1$ such that an element of $U$ has multiplicity greater than 1, we can assume that $U$ is actually a set. Furthermore, we can assume that both $G''[U + v_i]$ and $G''[U + v_j]$ are independent sets, since there are only $o(n^{3k-1})$ sets of order $3k-1$ that do not have this property. This implies that $U + v_i$ and $U + v_j$ both have exactly $k$ vertices in $V_1$, so, since $i \neq j$, neither $i$ nor $j$ is 1. Therefore, we can assume without loss of generality that $i$ is even. Hence, $|(U + v_i) \cap V_i| = |(U + v_i) \cap V_{i+1}|$ and $|(U + v_j) \cap V_i| = |(U + v_j) \cap V_{i+1}|$, which is impossible since $i \neq j$. □

Now we study properties of a linked partition of any graph.

**Proposition 3.2.6.** *For a graph $G = (V, E)$, let $x_1, x_2 \in V$, $k_1, k_2 \in \mathbb{N}$, $c, c_1, c_2 > 0$, $k := k_1 + k_2$ and $c' := \min\{c, c_1, c_2\}$. If*

$$|L_{c_1,k_1}(x_1) \cap L_{c_2,k_2}(x_2)| \geq cn,$$

*then $x_1$ and $x_2$ are $(\frac{1}{3}c', k)$-linked.*

*Proof.* Assume $k_1 \leq k_2$. Let $(x, U_1, U_2)$ be an ordered triple such that $x \in L_{c_1,k_1}(x_1) \cap L_{c_2,k_2}(x_2)$ and $U_i$ is a $k_i$-linking set for $\{x_i, x\}$ and $i \in [2]$. There are at least

$$cn \cdot (c_1 n)^{3k_1 - 1} \cdot (c_2 n)^{3k_2 - 1} \geq (c' n)^{3k_1 + 3k_2 - 1}$$

such ordered triples and if $U := \{x\} \cup U_1 \cup U_2$ then $U$ is a $k_1 + k_2$ linking set for $\{x_1, x_2\}$. Let $(x', U'_1, U'_2)$ be another such triple such that $U = \{x'\} \cup U'_1 \cup U'_2$. By first picking $x'$ and then $U'_1$ from the multiset $U$ (and using the fact that $x + 1 \le 3 \cdot (3/2)^x$ for all values of $x > 0$), we have that there at most

$$(3k_1 + 3k_2 - 1) \cdot \binom{3k_1 + 3k_2 - 2}{3k_1 - 1} \le \left(3 \cdot \left(\frac{3}{2}\right)^{3k_1+3k_2-2}\right) \cdot 2^{3k_1+3k_2-2} = 3^{3k_1+3k_2-1}$$

such triples $(x', U'_1, U'_2)$ and the conclusion follows. □

**Definition 3.2.7.** Given a partition $\mathcal{M} = \{V_1, \ldots, V_d\}$ of $V$, $0 < \phi < 1$ and any multiset $I$ of $[d]$ of order 3, let $t(\mathcal{M}, I)$ be the number of triangles $T$ such that $|V(T) \cap V_i| = \nu_I(i)$ for every $1 \le i \le d$ and let

$$f_\phi(\mathcal{M}, I) = \begin{cases} 1 & \text{if } t(\mathcal{M}, I) \ge \phi n^3, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Also, let $F_\phi(\mathcal{M}) := \{I : f_\phi(\mathcal{M}, I) = 1\}$, and, for $i \in [d]$, let $t_\phi(\mathcal{M}, i)$ be the number of times the index $i$ appears in a multiset of $F_\phi(\mathcal{M})$ with multiplicity, i.e. $3 \cdot |F_\phi(\mathcal{M})| = \sum_{i=1}^{d} t_\phi(\mathcal{M}, i)$. When the partition $\mathcal{M}$ is clear from context, we often use $F_\phi$ and $t_\phi(i)$ to refer to $F_\phi(\mathcal{M})$ and $t_\phi(\mathcal{M}, i)$, respectively. For convenience, we let $k : F_\phi(\mathcal{M}) \times [3] \to [d]$ be the map defined by $\{k(I,1), k(I,2), k(I,3)\} = I$ and $k(I,1) \le k(I,2) \le k(I,3)$ for every $I \in F_\phi(\mathcal{M})$.

For the graph from Example 3.2.1, $F_{\varepsilon^2}(\{V_1, \ldots, V_{2m+1}\}) = \{\{1, 2i, 2i+1\} : i \in [m]\}$, $k(\{1, 2i, 2i+1\}, 1) = 1$, $k(\{1, 2i, 2i+1\}, 2) = 2i$, and $k(\{1, 2i, 2i+1\}, 3) = 2i+1$ for every $i \in [m]$.

**Definition 3.2.8.** Given constants $0 < \eta < \phi < \psi \le 1$ and a partition $\mathcal{M} = \{V_1, \ldots, V_d\}$ of $V$ and $A \subseteq V$, a collection $\mathcal{N}$ is called $(\mathcal{M}, \phi, \eta)$-*absorbable (with respect to $A$)* if it consists of $3 \cdot |F_\phi(\mathcal{M})|$ vertex disjoint subsets of $V \setminus A$ and if there exists a bijective map $X : F_\phi(\mathcal{M}) \times [3] \to \mathcal{N}$ such that

- $X(I, j) \subseteq V_{k(I,j)}$ for every $j \in [3]$ and
- $|X(I,1)| = |X(I,2)| = |X(I,3)| \le \eta n$.

For every $(\mathcal{M}, \phi, \eta)$-absorbable collection $\mathcal{N}$ we will always implicitly assume that a fixed function $X$ exists. Call $A$ an $(\mathcal{M}, \phi, \eta)$-*absorber* if for any collection $\mathcal{N}$ of disjoint sets that is $(\mathcal{M}, \phi, \eta)$-absorbable with respect to $A$, $G[A \cup V(\mathcal{N})]$ has a triangle factor.

When $d = 1$, Lemma 3.2.9 is very similar to lemmas that appear in other results which use the absorbing method and the proof is nearly identical, for example see Lemma 1.1 in [47] for a general result used for hypergraph matching.

**Lemma 3.2.9.** *For any $k$ and $0 < \eta \ll \sigma \ll \phi \ll \psi \leq 1$, the following holds. If $G = (V, E)$ is a graph and $\mathcal{M} = \{V_1, \ldots, V_d\}$ is a $(\psi, \phi, k)$-linked partition of $V$, then there exists an $(\mathcal{M}, \phi, \eta)$-absorber $A \subseteq V$ such that $|A| \leq \sigma n$.*

*Proof.* Let $\ell := 9 \cdot k$ and $\eta \ll \xi \ll \sigma$. For any 3-set $W = \{w_1, w_2, w_3\} \subseteq V$ denote by $\mathcal{L}_W$ the set of ordered $\ell$-tuples $(u_1, \ldots, u_\ell) \in V^\ell$ such that $u_{3k}u_{6k}u_{9k}$ is a triangle and, for $j \in [3]$, the multiset $\{u_{3k\cdot(j-1)+1}, \ldots, u_{3k\cdot j-1}\}$ is a $k$-linking multiset for $\{w_j, u_{3k\cdot j}\}$. Note that if the vertices $u_1, \ldots, u_\ell$ are distinct and $U := \{u_1, \ldots, u_\ell\}$, then $G[U]$ and $G[U \cup W]$ both have triangle factors. We say that the 3-set $W$ is *acceptable* if $|\mathcal{L}_w| \geq 4(\phi n)^\ell$.

Form a random subset of $\ell$-tuples $\mathcal{A}' \subseteq V^\ell$ where each $\ell$-tuple is picked independently at random with probability $p := \xi n^{1-\ell}$. We have the following:

$$\mathbb{E}|\mathcal{A}'| = p \cdot |V^\ell| = \xi n, \tag{3.2}$$
$$\mathbb{E}|\mathcal{A}' \cap \mathcal{L}_W| \geq p \cdot 4(\phi n)^\ell = 4\xi\phi^\ell n \text{ for every acceptable 3-set } W. \tag{3.3}$$

We call a pair of $\ell$-tuples $(u_1, \ldots, u_\ell)$ and $(u'_1, \ldots, u'_\ell)$ a *bad pair* if a vertex appears more than once in the list $u_1, \ldots, u_\ell, u'_1, \ldots, u'_\ell$. The number of bad pairs is at most $(2\ell)^2 \cdot n^{2\ell-1}$. Hence,

$$\mathbb{E}|\{\text{bad pairs in } \mathcal{A}'\}| \leq p^2(2\ell)^2 \cdot n^{2\ell-1} = 4\xi^2(\ell)^2 n. \tag{3.4}$$

Therefore, by Markov's inequality 1.2.1, with probability at most 1/2,

(a) $\mathcal{A}'$ has at most $8\xi^2(\ell)^2 n$ bad-pairs.

Furthermore, since there are at most $\binom{n}{3}$ acceptable sets $W$, the Chernoff bound and the union bound with (3.2) and (3.3) imply that w.h.p. $\mathcal{A}'$ is such that

(b) $|\mathcal{A}'| \leq 2\xi n$ and

(c) $|\mathcal{A}' \cap \mathcal{L}_W| \geq 2\xi\phi^\ell n$ for all acceptable 3-sets $W$.

Therefore there exists $\mathcal{A}'$ that satisfies properties (a), (b) and (c). We now remove both elements from every bad pair in $\mathcal{A}'$. We also remove any tuples in $\mathcal{A}'$ that are not in $\mathcal{L}_W$ for any acceptable 3-set $W$. We call $\mathcal{A}$ the remaining part of $\mathcal{A}'$. Note that for every $(u_1, \ldots, u_\ell) \in \mathcal{A}$, there is a triangle factor in $G[\{u_1, \ldots, u_\ell\}]$. Since $\phi^\ell \geq 16\ell^2\xi$,

$$|\mathcal{A} \cap \mathcal{L}_W| \geq \xi\phi^\ell n \text{ for every acceptable 3-set } W. \tag{3.5}$$

Let $A$ be the union of the vertices in the $\ell$-tuples of $\mathcal{A}$. We have that $|A| \leq 2\ell\xi n \leq \sigma n$. Let $\mathcal{N}$ be a collection of disjoint subsets of $V \setminus A$ that is $(\mathcal{M}, \phi, \eta)$-absorbable. For every $I \in F_\phi$, $|X(I,1)| = |X(I,2)| = |X(I,3)| \leq \eta n$, so there exists a partition $\mathcal{W}$ of $V(\mathcal{N})$ into parts of size 3 such that for every $W \in \mathcal{W}$ there exists $I \in F_\phi$ such that $W$ has one vertex in each of $X(I,1), X(I,2)$ and $X(I,3)$. Note that

$$|\mathcal{W}| = |V(\mathcal{N})|/3 \leq 3\eta n \cdot |F_\phi|/3 \leq \eta n d^3 \leq \xi \phi^\ell n. \tag{3.6}$$

We claim that every $W \in \mathcal{W}$ is acceptable. By construction, there exists an $I \in F_\phi$ such that $W$ has one vertex in each of $X(I,1), X(I,2)$ and $X(I,3)$. We can label $W$ as $\{w_1, w_2, w_3\}$ so that $w_j \in X(I,j) \subseteq V_{K(I,j)}$ for each $j \in [3]$. Since $f_\phi(I) = 1$, there are $\phi n^3$ triangles $u_{3k}u_{6k}u_{9k}$ such that $u_{3k\cdot j} \in V_{k(I,j)}$ for $j \in [3]$. Furthermore, for any $j \in [3]$, since $V_{K(I,j)}$ is $(\phi, k)$ linked, there are at least $(\phi n)^{3k-1}$ $k$-linking multisets for $\{w_j, u_{3k\cdot j}\}$, for each $j \in [3]$. Therefore,

$$|\mathcal{L}_W| \geq (\phi n)^{3(3k-1)} \phi n^3 \geq 4(\phi n)^\ell,$$

so $W$ is acceptable.

Hence, by (3.5) and (3.6), we can match every $W \in \mathcal{W}$ to a different $\ell$-tuple in $\mathcal{A} \cap \mathcal{L}_W$ to construct a triangle factor of $G[V(\mathcal{N}) \cup A]$. □

## 3.3 Proof of Theorem 3.1.2

### 3.3.1 Overview

Following the absorbing method, the heart of the proof is the following two lemmas, which we show implies the theorem.

**Lemma 3.3.1** (Absorbing Lemma for Theorem 3.1.2)**.** *For $0 < \gamma \ll \zeta \ll \sigma \ll \varepsilon < 1/6$ the following holds. If $G = (V, E)$ is a graph such that $\delta(G) \geq (1/2 + \varepsilon)n$ and $\alpha(G) \leq \gamma n$, then there exists $U \subseteq V$ such that $|U| \leq 2\sigma n$ and for every $W \subseteq V \setminus U$ such that $|W|$ is at most $\zeta n$ and divisible by 3, $G[U \cup W]$ has a triangle factor.*

**Lemma 3.3.2** (Triangle Covering Lemma for Theorem 3.1.2)**.** *For any $\varepsilon > 0$, there exists $\gamma > 0$ and $n_0$ such that the following holds. For every $n$-vertex graph $G$ with $n > n_0$, $\delta(G) \geq (1/2 + \varepsilon)n$ and $\alpha(G) \leq \gamma n$, there is a triangle tiling of all but at most $16/\varepsilon + 1$ vertices.*

*Proof of Theorem 3.1.2.* Let $0 < \gamma \ll \zeta \ll \sigma \ll \varepsilon < 1/6$ be as in Lemma 3.3.1 and such that $\gamma$ is small enough so that Lemma 3.3.2 holds when $\varepsilon$ and $\gamma$ are replaced with $\varepsilon' := \varepsilon - 2\sigma$ and $\gamma' := \gamma/(1-2\sigma)$, respectively. Let $U \subseteq V$ be a set of size at most $\sigma n$ that is guaranteed by Lemma 3.3.1 and let $V' := V \setminus U$, $n' := |V'|$ and $G' := G[V']$. Note that $\delta(G') \geq (1/2+\varepsilon')n'$ and $\alpha(G') \leq \gamma n \leq \gamma' n'$, so Lemma 3.3.2 implies that there exists a triangle tiling $\mathcal{T}_1$ such that if $W := V' \setminus V(\mathcal{T}_1)$, then $|W| \leq 16/\varepsilon' + 1$. Since $n$ is divisible by 3, $|W|$ is divisible by 3 and Lemma 3.3.1 implies that there exists a triangle factor $\mathcal{T}_2$ of $G[W \cup U]$, and $\mathcal{T}_1 \cup \mathcal{T}_2$ is a triangle factor of $G$. □

### 3.3.2 Proof of Lemma 3.3.1

First we prove a series of lemmas and claims as preparation for the proof of Lemma 3.3.1. The first lemma is similar to the Dependent Random Choice Lemma.

**Lemma 3.3.3.** *Let $F$ be a bipartite graph with classes $(A, B)$ and $0 < \varepsilon \leq 1$ be such that $d_F(a) \geq \varepsilon|B|$ for every $a \in A$ and $d_F(b) \geq \varepsilon|A|$ for every $b \in B$. If $B$ is sufficiently large, then for every $0 < \psi < \varepsilon^4/64$ there exists a collection of disjoint subsets $\{S_1, \ldots, S_d\}$ of $B$ such that*

1. *for every $i \in [d]$, $|S_i| \geq \psi|B|$,*
2. *$\left|\bigcup_{i=1}^{d} S_i\right| \geq (1-\psi)|B|$, and*
3. *for every $i \in [d]$, there are at most $\psi^3|B|$ pairs in $b, b' \in S_i$ such that $|N_F(b) \cap N_F(b')| < \psi^4|A|$.*

*Proof.* Since $0 < \varepsilon \leq 1$ and $0 < \psi < \varepsilon^4/64$, we have the following:

$$-\log(\psi/2)/\varepsilon < 4\psi^{-1/2}/\varepsilon - 1 = 8\psi^{1/2}/(2\psi \cdot \varepsilon) - 1 < \varepsilon/(2\psi) - 1.$$

Hence, we can pick a positive integer $d$ so that

$$-\log(\psi/2)/\varepsilon < d < \varepsilon/(2\psi). \tag{3.7}$$

Call a pair $(b, b') \in B^2$ *bad* if $|N_F(b) \cap N_F(b')| < \psi^4|A|$ and let $Z \subseteq B^2$ be the set of bad pairs. Let $U = \{a_1, \ldots, a_d\} \subseteq A$ be a set of $d$ vertices selected uniformly at random and independently with repetition for $A$, and define $f_i$ to be the random variable counting

$|N(a_i)^2 \cap Z|$ for every $i \in [d]$. By (3.7),

$$\mathbb{E} f_i = \sum_{(b,b')\in Z} \mathbb{P}(a_i \in N_F(b) \cap N_F(b')) = \sum_{(b,b')\in Z} \frac{|N_F(b) \cap N_F(b')|}{|A|} < \sum_{(b,b')\in Z} \psi^4 < \frac{\psi^3}{2d}|B|^2. \tag{3.8}$$

Let $Y := \{b \in B : b \notin \bigcup_{i=1}^d N_F(a_i)\}$, therefore, using (3.7),

$$\mathbb{E}|Y| = \sum_{b\in B} \mathbb{P}(N_F(b) \cap U = \emptyset) = \sum_{b\in B} \left(1 - \frac{|N_F(b)|}{|A|}\right)^d \leq (1-\varepsilon)^d |B| \leq e^{-\varepsilon d}|B| < \frac{\psi|B|}{2}.$$

Markov's inequality 1.2.1 and the union bound implies that there exist a choice of $\{a_1, \dots, a_d\} \subseteq A$ such that $|N_F(a_i)^2 \cap Z| \leq \psi^3|B|^2$ for every $i \in [d]$, and $|V \setminus \bigcup_{i=1}^d N(a_i)| \leq \psi|B|$. Fix such an $\{a_1, \dots, a_d\}$ and let $S_i' := N(a_i)$ for $i \in [d]$.

To make the sets $S_i'$ disjoint, we use the following probabilistic argument. For every vertex $v \in \bigcup_{i=1}^d S_i'$ we select uniformly at random and independently of other vertices an index $j$ from the set $\{j \in [d] : v \in S_j'\}$, and then assign $v$ to the set $S_j$. At the end of this process, the sets $\{S_1, \dots, S_d\}$ are disjoint, and using (3.7) we have,

$$\mathbb{E}|S_i| = \sum_{v\in S_i'} |\{j \in [d] : v \in S_j'\}|^{-1} \geq \frac{d_F(a_i)}{d} \geq \frac{\varepsilon|B|}{d} \geq 2\psi|B| \qquad \text{for all } i \in [d].$$

Because each $S_i$ is the sum of independent random indicator variables, the Chernoff bound implies that

$$\mathbb{P}(|S_i| \leq \psi|B|) \leq 2\exp(-((1/2)^2 \cdot 2\psi|B|)/3) < 1/d \qquad \text{for all } i \in [d],$$

and, with the union bound, there is an assignment such that $|S_i| \geq \psi|B|$ for every $i \in [d]$. □

**Proposition 3.3.4.** *For any $0 < \varepsilon < 1/6$, if $G = (V, E)$ is a graph on $n$ vertices such that $\delta(G) \geq (1/2 + \varepsilon)n$, then for every vertex $v \in V$, $|L_{\varepsilon^2,1}(v)| \geq \frac{3}{2}\varepsilon^2 n$, for $n$ sufficiently large.*

*Proof.* For a vertex $v \in V$ define

$$F(v) := \{(u, e) \in (V - v) \times E : ve \text{ and } ue \text{ are triangles}\}.$$

Since $\delta(G) \geq (1/2 + \varepsilon)n$, we have $e(G[N(v)]) \geq ((1/2 + \varepsilon)n \cdot 2\varepsilon n)/2$ for $n$ sufficiently large, furthermore for every edge $uu' \in E(G[N(v)])$, $|N(u) \cap N(u') - v| \geq 2\varepsilon n - 1$. Hence,

$$|F(v)| \geq \left(\frac{1}{2} + \varepsilon\right) n \cdot \varepsilon n \cdot (2\varepsilon n - 1) \geq \varepsilon^2 n^3. \tag{3.9}$$

On the other hand,

$$|F(v)| \leq (n - |L_{\varepsilon^2,1}(v)|) \cdot (\varepsilon^2 n)^2 + |L_{\varepsilon^2,1}(v)||E| \leq \varepsilon^4 n^3 + |L_{\varepsilon^2,1}(v)|\frac{n^2}{2}. \tag{3.10}$$

Since $\varepsilon < 1/6$, (3.9) and (3.10) imply that $|L_{\varepsilon^2,1}(v)| \geq \frac{3}{2}\varepsilon^2 n$. □

For reference, we now list the relationship between the constants used in the rest of this section:

$$0 < \gamma \ll \zeta \ll \beta \ll \eta \ll \sigma \ll \phi \ll \psi \ll \varepsilon < 1/6. \tag{3.11}$$

We will also have that $d$ is a positive integer such that

$$d \leq 1/\psi \tag{3.12}$$

**Lemma 3.3.5.** *Assuming* (3.11)*, if $G = (V, E)$ is a graph on $n$ vertices where $\delta(G) \geq (1/2 + \varepsilon)n$, then there exists a $(\phi, \psi, 6)$-linked partition $\mathcal{M} = \{V_1, \ldots, V_d\}$ of $V$ for some $d \leq 1/\psi$.*

*Proof.* Let $F$ be the bipartite graph with parts $E$ and $V$ such that $ev \in E(F)$ if $ev$ is a triangle in $G$. For every $v \in V$,

$$d_F(v) = \frac{1}{2} \cdot \sum_{v' \in N_G(v)} |N_G(v) \cap N_G(v')| \geq \frac{1}{2} \cdot \delta(G) \cdot (2\delta(G) - n) \geq \varepsilon |E|,$$

and, for every $vv' \in E(G)$,

$$d_F(vv') = |N_G(v) \cap N_G(v')| \geq 2 \cdot \delta(G) - n \geq 2\varepsilon |V|.$$

Therefore, by Lemma 3.3.3, there exists a disjoint collection of vertex sets $\{V_1', \ldots, V_d'\}$ such that if $R' := V \setminus \left(\bigcup_{i=1}^d V_i'\right)$, then $|R'| \leq 2\psi n$, and, for every $i \in [d]$, $|V_i'| \geq 2\psi n$ and, for all $i \in [d]$, all but at most $(2\psi)^3 n^2$ pairs $v, v' \in V_i'$ are such that

$$|N_F(v) \cap N_F(v')| \geq (2\psi)^4 n^2. \tag{3.13}$$

In the remainder of the proof, we will potentially remove some vertices from the each of the sets $V_1', \ldots, V_d'$ and the distribute these removed vertices and the vertices in $R'$ into the sets to create the desired partition. To help achieve this, we build an auxiliary graph $H$ with $V(H) = V(G)$ and in which two vertices $v, v' \in V(H)$ are adjacent if and only if $v$ and

$v'$ satisfy (3.13). Also, define $H_i := H[V_i']$ for $i \in [d]$. For any $i \in [d]$, note that

$$N_{H_i}(v) \subseteq L_{4\psi^2,1}(v) \text{ for any } v \in V_i' \tag{3.14}$$

Let $J_i := \{v \in V(H_i) : d_{\overline{H_i}}(v) \geq 8\psi^2 n\}$ and $V_i'' := V_i' \setminus J_i$. Since $e(\overline{H_i}) \leq (2\psi)^3 n^2$, we have that

$$|J_i| \leq \frac{8\psi^3 n^2}{8\psi^2 n} = \psi n \text{ and } |V_i''| \geq \psi n \text{ for every } i \in [d].$$

Let $v, v' \in V_i''$. Since $v, v' \notin J_i$,

$$|N_{H_i}(v) \cap N_{H_i}(v')| \geq 2 \cdot (|V_i'| - 8\psi^2 n) - |V_i'| \geq 27\phi n. \tag{3.15}$$

By (3.14) and (3.15), Proposition 3.2.6 with $k_1 = 1$, $k_2 = 1$, $c = 27\phi$ and $c_1 = c_2 = 4\psi^2$, implies that $v$ and $v'$ are $(9\phi, 2)$-linked. Therefore, $V_i''$ is $(9\phi, 2)$-linked. Similarly, Proposition 3.2.6 also implies that $V_i''$ is $(3\phi, 3)$-linked and $(\phi, 6)$-linked.

Let $v \in J_1 \cup \cdots \cup J_d \cup R$. By Proposition 3.3.4, there exists $i \in [d]$ such that

$$\left|\{u \in V_i'' \colon u \text{ and } v \text{ are } (\varepsilon^2, 1)\text{-linked}\}\right| \geq \frac{\frac{3}{2}\varepsilon^2 n - |R|}{d} - |J_i| \geq 9\phi n. \tag{3.16}$$

Therefore, we can construct a partition (that may contain empty parts) $\{W_1, \ldots, W_d\}$ of $J_1 \cup \cdots \cup J_d \cup R$ such that for every $i \in [d]$ and every $w \in W_i$, $|L_{\varepsilon^2,1}(w) \cap V_i''| \geq 9\phi n$.

Since $V_i''$ is $(9\phi, 2)$-linked, (3.16) and Proposition 3.2.6 imply that, for every $w \in W_i$, $V_i'' + w$ is $(3\phi, 3)$-linked and also $(\phi, 6)$-linked. Therefore, for every two distinct vertices $w_1, w_2 \in W_i$, since $|V_i''| \geq 3\phi n$, Proposition 3.2.6 implies that $w_1$ and $w_2$ are $(\phi, 6)$-linked. Hence, if $V_i := V_i'' \cup W_i$ for every $i \in [d]$, then $\mathcal{M} := \{V_1, \ldots, V_d\}$ is a $(\psi, \phi, 6)$-linked partition of $V$. □

**Definition 3.3.6.** Let $G$ be an $n$-vertex graph. Define $\mathcal{S} := \{S_1, \ldots, S_m\}$ be a family of subsets of $V(G)$. For $a > 0$, let $C(\mathcal{S}, a)$ be the graph with vertex set $\mathcal{S}$ where the following holds

$$\begin{aligned} S_i S_j \in E(C) \iff \quad & |\{v \in S_i : |N(v) \cap S_j| \geq an\}| \geq an \text{ and} \\ & |\{v \in S_j : |N(v) \cap S_i| \geq an\}| \geq an. \end{aligned} \tag{3.17}$$

**Proposition 3.3.7.** *Let $0 < \gamma < a < 1$ and $d \in \mathbb{N}$. Let $G = (V, E)$ be a graph on $n$ vertices such that $\alpha(G) \leq \gamma n$, $\mathcal{S} = \{S_1, \ldots, S_d\}$ a collection of disjoint subsets of $V$, and $W \subseteq V$ such that $|W| < (a - \gamma)n$ the following holds. If $P$ is a $(S, S')$-path in $C(\mathcal{S}, a)$, then there exists a set of vertex disjoint triangles $\mathcal{Y}$ in $G[V(\mathcal{S}) \setminus W]$ such that:*

- $|\mathcal{Y}| = |E(P)|$, $|V(\mathcal{Y}) \cap S| = 1$, $|V(\mathcal{Y}) \cap S'| = 2$ *and*
- $|V(\mathcal{Y}) \cap S''| \in \{0, 3\}$ *for every* $S'' \in \mathcal{S} - S - S'$.

*Proof.* Let $S = S_1, \ldots, S_\ell = S'$ be $P$. We will iteratively construct vertex disjoint triangles $v_1e_1, \ldots, v_{\ell-1}e_{\ell-1}$, so that $v_i \in S_i \setminus W$ and $e_i \in E(G[S_{i+1} \setminus W])$. We always select $v_i$ so that $d(v_i, S_{i+1}) \geq an$, which is possible by the definition of $C(\mathcal{S}, a)$. Selecting $e_i$ is then possible because $\alpha(G) \leq \gamma n < an - |W|$. □

The following lemma relies heavily on Definitions 3.2.3, 3.2.7, 3.2.8 and 3.3.6.

**Lemma 3.3.8.** *For any $k$ and assuming* (3.11)*, then if $G = (V, E)$ is a graph on $n$ vertices such that $\delta(G) \geq (1/2 + \varepsilon)n$, $\mathcal{M} = \{V_1, \ldots, V_d\}$ is a $(\psi, \phi, k)$-linked partition of $V$ and $A$ is an $(\mathcal{M}, \phi, \eta)$-absorber such that $|A| \leq \sigma n$, then there exists $\mathcal{N}$ an $(\mathcal{M}, \phi, \eta)$-absorbable collection with respect to $A$ such that:*

*(a) for every $I \in F_\phi(\mathcal{M})$ and $j \in [3]$, $|X(I, j)| = \lfloor \eta n \rfloor$,*

*(b) the graph $C(\mathcal{N}, \beta)$ is connected,*

*(c) for every $v \in V$, there exists $I \in F_\phi(\mathcal{M})$ and $j \in [3]$ such that $d(v, X(I, j)) \geq \beta n$, and*

*(d) for every $I \in F_\phi(\mathcal{M})$, $X(I, 1)X(I, 2)X(I, 3)$ is a triangle in $C(\mathcal{N}, \beta)$.*

*Proof.* Chose $\tau$ so that $\sigma \ll \tau \ll \phi$, and define $V_i' := V_i \setminus A$ for every $i \in [d]$. For every $i, i' \in [d]$, let

$$U_{(i,i')} := \{v \in V_i' : d(v, V_{i'}) \geq \tau n\}.$$

Note that if $V_i$ and $V_i'$ are adjacent in $C(\mathcal{M}, \tau)$, then, by the definition of $C(\mathcal{M}, \tau)$ and the fact that $|A| \leq \sigma n$, we have that $|U_{(i,i')}|, |U_{(i',i)}| \geq \tau n - |A| \geq \tau n/2$.

We first establish the following three simple claims.

**Claim 3.3.9.** For every $i \in [d]$, $t_\phi(\mathcal{M}, i) \geq 1$.

*Proof.* Assume that $t_\phi(\mathcal{M}, i) = 0$. Then the number of triangles containing vertices of $V_i$ is less than $d^2\phi n^3$, but there are at least $\left(\sum_{v \in V_i} e(G[N(v)])\right)/3 \geq \psi n \cdot \varepsilon n^2 \cdot 1/3$ such triangles, a contradiction. □

**Claim 3.3.10.** If $I \in F_\phi(\mathcal{M})$ where $\{i, i', i''\} = I$, then $|U_{(i,i')}| \geq \tau n/2$.

*Proof.* Note that if $\nu_I(i) \geq 2$, then it could be that $i = i'$. Since $I \in F_\phi(\mathcal{M})$, there are at least $\phi n^2$ edges with one end in $V_i$ and the other end in $V_{i'}$, therefore, since $d_G(v, V_{i'}) \leq |V_{i'}| \leq n$, for every $v \in U_{(i,i')}$,

$$\begin{aligned}|U_{(i,i')}| &\geq e_G(U_{(i,i')}, V_{i'})/n \\ &= \left(e_G(V_i, V_{i'}) - e(V_i \setminus U_{(i,i')}, V_{i'})\right)/n \geq \left(\phi n^2 - \tau n \cdot (|V_i| - |U_{(i,i')}|)\right)/n \geq \tau n/2.\end{aligned}$$ □

**Claim 3.3.11.** The graph $C(\mathcal{M}, \tau)$ is connected.

*Proof.* We can assume $d \geq 2$, so let $\mathcal{C}_1, \mathcal{C}_2$ be an arbitrary partition of $\mathcal{M}$ and let $U_i := \bigcup \mathcal{C}_i$ for $i \in [2]$. Without loss of generality we can assume that $|U_1| \leq |U_2|$, so $|U_1| \leq n/2$. We will show that there is an edge in $C(\mathcal{M}, \tau)$ between the sets $\mathcal{C}_1, \mathcal{C}_2$, which will prove the claim. We can assume that $V_1 \in \mathcal{C}_1$. For every $v \in V_1$, we have $|N_G(v) \cap (V \setminus U_1)| \geq \delta(G) - |U_1| \geq \varepsilon n$, so

$$e_G(V_1, U_2) \geq |V_1| \cdot \varepsilon n.$$

Hence, there exists some $V_i \notin \mathcal{C}_2$, say $V_2$, such that $e_G(V_1, V_2) \geq |V_1| \cdot \varepsilon n/d$. For $i \in [2]$, let $x_i$ be the number of vertices in $v \in V_i$ such that $|N_G(v) \cap V_{3-i}| \geq \tau n$. We have the following inequality,

$$x_i \cdot |V_{3-i}| + (|V_i| - x_i) \cdot \tau n \geq |V_1| \cdot \varepsilon n/d.$$

Since $\psi n \leq |V_1|, |V_2| \leq n$ and $\psi\varepsilon/d \geq \psi^2\varepsilon \geq 2\tau$, we have

$$x_i \geq \frac{|V_1| \cdot \varepsilon n/d - |V_i| \cdot \tau n}{|V_{3-i}| - \tau n} \geq \frac{(\psi\varepsilon/d - \tau)n^2}{n} \geq \tau n,$$

which means that $V_1$ and $V_2$ are adjacent in $C(\mathcal{M}, \tau)$. □

Now we proceed to prove Lemma 3.3.8. For every $i \in [d]$, let the collection $\mathcal{U}_i$ contain the sets $N(v) \cap V'_i$ for every $v \in V(G)$ and $U_{(i,i')}$ for every $i' \in [d]$. Note that $|\mathcal{U}_i| = n + d$ and that every set $U \in \mathcal{U}_i$ is is a subset of $V'_i$.

We will use the following probabilitic argument to contruct the desired $(\mathcal{M}, \phi, \eta)$-absorbable collection $\mathcal{N}$. Let $m := \lfloor \eta n \rfloor$ and select a set $Z_i \subseteq V'_i$ of size $t_\phi(\mathcal{M}, i) \cdot m$ uniformly at random. Then uniformly at random select a partition of $Z_i$ into $t_\phi(\mathcal{M}, i)$ parts each of size $m$ over all such partitions. Note that any such partition corresponds to an $(\mathcal{M}, \phi, \eta)$-absorbable collection, since for every $I \in F_\phi(\mathcal{M})$ and $j \in [3]$, we can uniquely assign $X(I, j)$ to one of the $t_\phi(\mathcal{M}, k(I, j))$ parts of $Z_{k(I,j)}$. Assume there exists such a fixed assigned for every such collection. For any $I \in F_\phi(\mathcal{M})$, $j \in [3]$ and $U \in \mathcal{U}_{k(I,j)}$, the random variable $|U \cap X(I, j)|$ is hypergeometrically distributed [1] and

$$\mathbb{E}|U \cap X(I,j)| = \frac{m}{|V'_{k(I,j)}|} \cdot |U| \geq 0.9 \cdot \eta \cdot |U|.$$

For any $I \in F_\phi$, $j \in [3]$, and any $U \in \mathcal{U}_{k(I,j)}$, when $|U| < \beta n$ the following probabilty estimate is trivially true and when $|U| \geq \beta n$ it is implied by the Chernoff bound for the hypergeometric distribution:

$$\mathbb{P}(|U \cap X(I,j)| < \mathbb{E}|U \cap X(I,j)| - \beta n) \leq \exp(-\beta^2/3 \cdot \mathbb{E}|U \cap X(I,j)|) \leq \exp(-\beta^3 n/3).$$

Hence, by the union bound, w.h.p.

$$|U \cap X(I,j)| \geq \mathbb{E}|U \cap X(I,j)| - \beta n$$

for each of the $n+d$ sets $U \in \mathcal{U}_{k(I,j)}$ simultaneously. Finally, this with the union bound again imply that there exists an $(\mathcal{M}, \phi, \eta)$-absorbable collection $\mathcal{N}$ such that, for every $I \in F_\phi$ and $j \in [3]$,

$$|U \cap X(I,j)| \geq 0.9 \cdot \eta \cdot |U| - \beta n \text{ for every } U \in \mathcal{U}_{k(I,j)}.$$

Rewriting this, we have that, for every $i' \in [d]$, $I \in F_\phi$ and $j \in [3]$,

$$d(v, X(I,j)) \geq 0.9 \cdot \eta \cdot d(v, V'_{k(I,j)}) - \beta n \text{ for every } v \in V, \tag{3.18}$$

and

$$|U_{(k(I,j),i')} \cap X(I,j)| \geq 0.9 \cdot \eta \cdot |U_{(k(I,j),i')}| - \beta n. \tag{3.19}$$

For any $I, I' \in F_\phi(\mathcal{M})$ and $j, j' \in [3]$, (3.18) and (3.19) imply that

$$\begin{aligned}&\text{if } k(I,j) \neq k(I',j') \text{ and } V_{k(I,j)}V_{k(I',j')} \in E(C(\mathcal{M},\tau)), \text{ then}\\ &X(I,j)X(I',j') \in E(C(\mathcal{N},\beta)).\end{aligned} \tag{3.20}$$

Also note that Claim 3.3.9 implies that,

$$\text{for every } i \in [d], \text{ there exists } I \in F_\phi(\mathcal{M}) \text{ and } j \in [3] \text{ such that } X(I,j) \subseteq V_i. \tag{3.21}$$

Combining (3.20) and (3.21), we have that for any $I, I' \in F_\phi(\mathcal{M})$ and $j, j' \in [3]$, if $k(I,j) \neq k(I',j')$ and there is a path from $V_{k(I,j)}$ to $V_{k(I',j')}$ in $C(\mathcal{M}, \tau)$, then there is a path from

[1] That is, if we have a bin with $|V_{k(I,j)}|$ balls and exactly $|U|$ of them are red, then the probability that there are exactly $t$ red balls after drawing $m$ balls without replacement from the bin is $\mathbb{P}(|U \cap X(I,j)| = t)$.

$X(I, j)$ to $X(I', j')$ in $C(\mathcal{N}, \beta)$. This, (3.21) and Claim 3.3.11 imply that when $d \geq 2$, the graph $C(\mathcal{N}, \beta)$ is connected. Also, for all $d \geq 1$, (3.20) and Claim 3.3.10, imply that $X(I, 1)X(I, 2)X(I, 3)$ is a triangle in $C(\mathcal{N}, \beta)$ for every $I \in F_\phi(\mathcal{M})$. Therefore, (d) holds. This and Claim 3.3.9, imply that $C(\mathcal{N}, \beta)$ is isomorphic to $K_3$ when $d = 1$, so (b) holds for all $d \geq 1$. Since (a) is true by construction, only (c) remains to be proved. To see that (c) holds, note that for every $v \in V$, there exists $i \in [d]$ such that $d(v, V_i') \geq ((1/2+\varepsilon)n - |A|)/d \geq \phi n$. Since (3.21) implies that there exist $I \in F_\phi(\mathcal{M})$ and $j \in [3]$ such that $X(I, j) \subseteq V_i$, (3.18) implies that $d(v, X(I, j)) \geq \beta n$.

□

*Proof of Lemma 3.3.1.* Assume (3.11) holds. Lemma 4.7.1 implies that there exists a $(\psi, \phi, 6)$-linked partition $\mathcal{M}$ of $V$. Lemma 3.2.9 implies that there exists $A \subseteq V$ such that $|A| \leq \sigma n$ and $A$ is an $(\mathcal{M}, \phi, \eta)$-absorber. Lemma 3.3.8 then implies that there exists a collection $\mathcal{N}$ of disjoint subsets of $V \setminus A$ such that $\mathcal{N}$ is $(\mathcal{M}, \phi, \eta)$-absorbable with respect to $A$ and that properties (a), (b), (c) and (d) of Lemma 3.3.8 hold. Let $N := V(\mathcal{N})$, i.e. $N := \bigcup\{X(I, j) : I \in F_\phi, j \in [3]\}$. Let $U := A \cup N$ and $W \subseteq V \setminus U$ such that $|W| \leq \zeta n$ and $|W|$ is divisible by 3. We have that $|U| \leq \sigma n + 3d^3\eta n \leq 2\sigma n$ and we will show that there is a triangle factor of $G[W \cup U]$ which will complete the proof.

For every $w \in W$, by Lemma 3.3.8(c), there exists some $I \in F_\phi$ and $j \in [3]$, such that $d(w, X(I, j)) \geq \beta n > \gamma n + 2|W|$. Therefore, since $\alpha(G) \leq \gamma n$, for every $w \in W$, we can assign some edge $e_w \in E(G[N(w) \cap X(I, j)])$ to $w$ so that $\mathcal{W} := \{we_w : w \in W\}$ is a collection of vertex disjoint triangles.

The idea of the remainder of the proof is the following. We iteratively construct another small collection $\mathcal{Y}$ of vertex disjoint triangles in $G[N \setminus V(\mathcal{W})]$. For convenience, we will use $\mathcal{Y}$ to represent the triangles that have been constructed so far in this iterative process. In particular, at the beginning of this process $\mathcal{Y} = \emptyset$. For every $I \in F$ and $j \in [3]$, we define $X'(I, j) := X(I, j) \setminus V(\mathcal{W} \cup \mathcal{Y})$. We also define $\mathcal{N}' := \{X'(I, j) : I \in F_\phi, j \in [3]\}$ and $N' := V(\mathcal{N}') = \bigcup \mathcal{N}'$. After this process is completed and we have finished constructed $\mathcal{Y}$, we will have that, for every $I \in F_\phi$, $|X'(I, 1)| = |X'(I, 2)| = |X'(I, 3)|$. Note that then because $A$ is an $(\mathcal{M}, \phi, \eta)$-absorber, and Lemma 3.3.8(a) implies that $|X'(I, 1)| = |X'(I, 2)| = |X'(I, 3)| \leq \eta n$, the collection $\mathcal{N}'$ is $(\mathcal{M}, \phi, \eta)$-absorbable with respect to $A$, so there exists a triangle factor $\mathcal{Z}$ of $G[A \cup N']$. Therefore, $\mathcal{W} \cup \mathcal{Y} \cup \mathcal{Z}$ is a triangle factor of $G[W \cup A \cup N] = G[W \cup U]$, which completes the proof.

We will now describe the two stage process for constructing $\mathcal{Y}$. Our goal in the first stage is for the following to hold for every $I \in F_\phi$:

$$|X'(I,1) \cup X'(I,2) \cup X'(I,3)| \equiv 0 \pmod 3. \tag{3.22}$$

At any step of the first stage of the algorithm, we call a triple $I \in F_\phi$, *bad* if it does not satisfy (3.22). Pick a bad $I \in F_\phi$ such that $|X'(I,1) \cup X'(I,2) \cup X'(I,3)| \equiv 1 \pmod 3$ if possible. Note that $|N'|$ is always divisible by three, because and $|N'| = |N|-2|W|-3|\mathcal{Y}|$ and $|W|$ and $|N|$ are both divisible by 3. Therefore, there exists another bad triple $I' \in F_\phi - I$. By Lemma 3.3.8(b) there exists a path $P$ from $X(I,1)$ to $X(I',1)$ in the graph $C(\mathcal{N},\beta)$. Hence, by Proposition 3.3.7, we can add a collection of at most $|P|-1$ vertex disjoint triangles to $\mathcal{Y}$, so that after this step, at least one of $I$ or $I'$ is no longer bad and every triple in $F_\phi$ that was good before this step remains good after this step is completed. Note that we finish the first phase in at most $|\mathcal{N}|$ steps, so $|\mathcal{Y}| \le |\mathcal{N}|(|\mathcal{N}|-1) \le (3\cdot d^3)^2$ after the first phase.

In each step of the second and final stage of the algorithm, we pick some $I \in F_\phi$ such that $|X'(I,1)| = |X'(I,2)| = |X'(I,3)|$ does not hold and add triangles contained in $G[X(I,1) \cup X(I,2) \cup X(I,3)]$ to $\mathcal{Y}$ until $|X'(I,1)| = |X'(I,2)| = |X'(I,3)|$ holds. We continue in this manner until we have the desired collection $\mathcal{Y}$. We will now describe this process for a fixed $I \in F_\phi$. Before each triangle is constructed, we relabel $\{j_1,j_2,j_3\} = [3]$ so that $|X'(I,j_1)| \le |X'(I,j_2)| \le |X'(I,j_3)|$ and let

$$c(I) := (|X'(I,j_2)| - |X'(I,j_1)|) + (|X'(I,j_3)| - |X'(I,j_1)|)\,.$$

We also fix $\Phi := c(I)$ before any triangle are constructed. Because $|\mathcal{Y}| \le 9\cdot d^6$ at the start of the second stage of the algorithm, $|\mathcal{W}| = |W|$, and every triangle in $\mathcal{Y} \cup \mathcal{W}$ has at most 2 vertices in $X(I,j)$ for any $j \in [3]$, we have that

$$\Phi \le 2\cdot 2(9\cdot d^6 + |W|) < 2\zeta n.$$

Note that because $I$ satisfies (3.22), we can conclude that $\Phi \equiv c(I) \equiv 0 \pmod 3$ throughout this process.

We now add a triangle to $\mathcal{Y}$ with one vertex in $X(I,j_2)$ and two vertices in $X(I,j_3)$ until $c(I) = 0$, which implies $|X'(I,1)| = |X'(I,2)| = |X'(I,3)|$ (recall that we relabel $\{j_1,j_2,j_3\} = [3]$ before each triangle is constructed). By Lemma 3.3.8(d), $X(I,1)X(I,2)X(I,3)$ is a triangle in $C(\mathcal{N},\beta)$. Therefore, there exists $v \in X'(I,j_2)$ such that $d(v, X'(I,j_3)) > \gamma n =$

$\alpha(G)$ and, hence, a triangle with one vertex in $X'(I, j_2)$ and two vertices in $X'(I, j_3)$, provided

$$|V(\mathcal{Y} \cup \mathcal{W}) \cap X(I, j_2)|, |V(\mathcal{Y} \cup \mathcal{W}) \cap X(I, j_3)| < (\beta - \gamma)n. \tag{3.23}$$

Assuming (3.23) always holds, this process will terminate after constructing at most $2 \cdot \Phi/3$ triangles, because $c(I)$ decreases by 3 after each triangle is added to $\mathcal{Y}$ unless $|X'(I, j_1)| = |X'(I, j_2)|$, and when $|X'(I, j_1)| = |X'(I, j_2)|$, $c(I)$ does not change, but $c(I)$ decreases by 3 when the following triangle is added to $\mathcal{Y}$. Therefore, $V(\mathcal{Y} \cup \mathcal{W})$ intersects any set in $\mathcal{N}$ in at most $2(2 \cdot \Phi/3 + 9 \cdot d^6 + |W|) < (\beta - \gamma)n$ vertices. Hence, (3.23) always holds and we can find the required triangles between $X'(I, j_2)$ and $X'(I, j_3)$. □

### 3.3.3 Proof of Lemma 3.3.2

*Proof of Lemma 3.3.2.* Set $\gamma < \varepsilon/36$. Let $\mathcal{T}$ be a maximum family of disjoint triangles in $G$, and $\mathcal{M}$ be a maximum matching in $G[V \setminus V(\mathcal{T})]$. Denote $\mathcal{V}$ the set of remaining vertices and let $v = |\mathcal{V}|$, i.e. $v = |G \setminus V(\mathcal{T} \cup \mathcal{M})|$. Denote $t := |\mathcal{T}|$ and $m := |\mathcal{M}|$, then we have $n = 3t + 2m + v$, $v \le \alpha(G) \le \gamma n$ and $t \ge (\delta(G) - \alpha(G))/3 \ge n/6$ by greedy construction.

**Claim 3.3.3.** $m < 8/\varepsilon$.

*Proof.* For a contradiction, assume $\varepsilon m \ge 8$. Note that for every vertex $u \in V(\mathcal{M})$, its degree in $G[V \setminus V(\mathcal{T})]$ is at most $v + m$, otherwise $u$ is adjacent to both ends of a matching edge in $\mathcal{M}$, contradicting the maximality of $\mathcal{T}$. Thus

$$\begin{aligned} d(u, V(\mathcal{T})) &\ge \left(\frac{1}{2} + \varepsilon\right) n - v - m = \left(\frac{1}{2} + \varepsilon\right)(3t + 2m + v) - v - m \\ &\ge \left(\frac{3}{2} + 3\varepsilon\right) t + \varepsilon m - \frac{v}{2} \ge \left(\frac{3}{2} + \varepsilon\right) t, \end{aligned} \tag{3.24}$$

where the last inequality follows from the fact that $v \le \gamma n$ and $t \ge n/6$. Thus $e(V(\mathcal{M}), V(\mathcal{T})) \ge (\frac{3}{2} + \varepsilon)t \cdot 2m = (3 + 2\varepsilon)tm$.

Let $\mathcal{T}'$ be the collection of triangles in $\mathcal{T}$, each sending at least $3m + 9$ edges to $\mathcal{M}$ and write $t' = |\mathcal{T}'|$. Note that each triangle $T \in \mathcal{T}$ can send at most $6m$ edges to $\mathcal{M}$, thus

$$e(V(\mathcal{M}), V(\mathcal{T})) \le t' \cdot 6m + (t - t')(3m + 8) = (3m + 8)t + (3m - 8)t'.$$

Together with (3.24) we have that

$$t' \ge \frac{2\varepsilon m - 8}{3m - 8} \cdot t \ge \frac{\varepsilon m}{3m - 8} \cdot t \ge \frac{\varepsilon}{3} \cdot t \ge \frac{\varepsilon n}{18}. \tag{3.25}$$

Note that for every $T \in \mathcal{T}'$, there is at least one vertex $s_T \in V(T)$ that sends at least $(3m+9)/3 = m+3$ edges to $\mathcal{M}$. Hence, $s_T$ forms a triangle with at least 3 edges in $\mathcal{M}$. Let $S := \{s_T : T \in \mathcal{T}'\}$ and $R := V(\mathcal{T}') \setminus S$. By the definition of $\mathcal{T}'$, we have $e(V(\mathcal{M}), V(\mathcal{T}')) \geq (3m+9)t'$. Thus there exists $u \in V(\mathcal{M})$ such that

$$d(u, V(\mathcal{T}')) \geq \frac{e(V(\mathcal{M}), V(\mathcal{T}'))}{2m} \geq \frac{(3m+9)t'}{2m} \geq \frac{3t'}{2}.$$

With (3.25) we have that $d(u, V(\mathcal{R})) \geq d(u, V(\mathcal{T}')) - |S| \geq t'/2 \geq (\varepsilon n)/36 > \gamma n$.

Since $\alpha(G) \leq \gamma n$, there is at least one edge $y_1 y_2 \in N_R(u)$. Let $T$ be the triangle $uy_1y_2$ and let $T_1, T_2 \in \mathcal{T}$ such that $y_i \in T_i$ for $i \in [2]$. Since, for $i \in [2]$, $s_{T_i}$ forms a triangle with at least three edges in $\mathcal{M}$, we can pick distinct edges in $e_1, e_2 \in \mathcal{M}$ such that neither contains $u$ and $s_{T_i} e_i$ is a triangle for $i \in [2]$. If $T_1 \neq T_2$, then $\mathcal{T} - T_1 - T_2 + T + s_{T_1} e_1 + s_{T_2} e_2$ contradicts the maximality of $\mathcal{T}$, and if $T_1 = T_2$, then $\mathcal{T} - T_1 + T + s_{T_1} e_1$ contradicts the maximality of $\mathcal{T}$. □

**Claim 3.3.4.** $v \leq 1$.

*Proof.* Suppose to the contrary that there exists two vertices $x, y \in V(\mathcal{V})$. $V(\mathcal{V})$ is an independent set, hence $v \leq \gamma n$, and, by Claim 3.3.3, $m < 8/\varepsilon$, therefore

$$e(\{x, y\}, V(\mathcal{T})) \geq 2(\delta(G) - m) \geq (1+\varepsilon)n > 3t + \varepsilon n.$$

Denote $\mathcal{T}'' := \{T \in \mathcal{T} : e(\{x, y\}, T) \geq 4\}$. It follows that $t'' := |\mathcal{T}''| \geq \varepsilon n/3 > \gamma n$. Fix a triangle $T = abc \in \mathcal{T}''$. If $d(x, V(T)) = 3$ and $d(y, V(T)) = 1$, say $ya \in E(G)$, then we get a triangle $xbc$ and an edge $ya$, contradicting to the maximality of $\mathcal{M}$. Thus we may assume that $d(x, V(T)) = d(y, V(T)) = 2$. Note that if $x$ is adjacent to $\{a, b\}$ and $y$ is adjacent to $\{a, c\}$, then we get the triangle $xab$ and an edge $yc$, contradicting to the maximality of $\mathcal{M}$. Hence, both $x$ and $y$ are adjacent to the same two vertices in $T$. Let $S := N(x) \cap N(y) \cap V(\mathcal{T}'')$, and $R := V(\mathcal{T}'') \setminus S$. Since $|R| = t'' > \gamma n$, there exist two triangles $abc, a'b'c' \in \mathcal{T}''$ such that $cc' \in E(G[R])$. Now we can take $xab, ya'b'$ and $cc'$, again contradicting to the maximality of $\mathcal{M}$. □

The number of vertices not covered in $\mathcal{T}$ is then $2m + v < 16/\varepsilon + 1$.

□

## 3.4 Proof of Theorem 3.1.3

We prove Theorem 3.1.3 in roughly the same way as we proved Theorem 3.1.2. That is, we prove an absorbing lemma (Lemma 3.4.1) and an almost tiling lemma (Lemma 3.4.4) and then we use them both to obtain the desired result. We omit the details of proving Theorem 3.1.3, given Lemma 3.4.1 and Lemma 3.4.4, since they are identical to the analogous proof of Theorem 3.1.2.

**Notation.** For disjoint vertices $x, y, z$, we will let $x$, $xy$ and $xyz$ represent the sets $\{x\}$, $\{x, y\}$ and $\{x, y, z\}$ respectively. It should be clear from context whether we mean for $x$ to represent the vertex $x$ or the singleton set $\{x\}$. For any $U \subseteq V$, we will let $\overline{U} = V \setminus U$, $\mathrm{ind}U := \sum_{e \in \binom{U}{2}} w(e) \cdot 3$ and for $W \subseteq \overline{U}$ we will let $\|U, W\| := \sum_{e \in E(U,W)} w(e) \cdot 3$. For disjoint vertices $x, y$ and $z$ we call $xyz$ a *heavy triangle* if $\mathrm{ind}xyz > 9t$. We multiply by three here purely for notational convenience.

To prove the absorbing lemma, we will consider the very simply partition $\mathcal{M} := \{V_1\}$ of $V$, i.e. $V_1 := V$. We show that there are at least $\phi n^3$ heavy triangles in $G$, i.e. $t_\phi(\mathcal{M}, \{1, 1, 1\}) = 1$ and that the entire vertex set $V_1 := V$ is $(\phi, 1)$-linked. Applying Lemma 3.2.9 will essentially complete the proof of the absorbing lemma.

**Lemma 3.4.1** (Absorbing Lemma)**.** *For any $t \in (0, 1)$ let $0 < \zeta \ll \sigma \ll \varepsilon < 1$ and $n_0$ such that the following holds. For any $n \geq n_0$ that is divisible by 3, graph $(V, E) = G = K_n$ and $w : E \to [0, 1]$ such that $\delta_w(G) \geq \left(\frac{1+2t}{3} + \varepsilon\right) n$, there exists $U \subset V$ such that $|U| \leq \sigma n$ and for any $W \subseteq V \setminus U$ such that $|W|$ is at most $\zeta n$ and divisible by 3, there exists a perfect tiling of $G[U \cup W]$ with heavy triangles.*

*Proof.* Let $\sigma \ll \phi \ll \varepsilon$. The following two claims make up the bulk of the proof.

**Claim 3.4.2.** There are at least $\frac{1}{4}n^3$ ordered triples $(x, y, z) \in V^3$ such that $xyz$ is a heavy triangle.

*Proof.* Pick any $x \in V$. For any $y \in V - x$, let $V' = V - x - y$ and

$$Z_y := \{z \in V' : xyz \text{ is heavy triangle}\}.$$

By $\delta_w(G) \geq \left(\frac{1+2t}{3} + \varepsilon\right) n$,

$$(2{+}4t)|V'| < \|xy, V'\| \leq 6 \cdot |Z_y| + (9t - \mathrm{ind}xy)|V' \setminus Z_y| = (6 - 9t + \mathrm{ind}xy)\,|Z_y| + (9t - \mathrm{ind}xy)|V'|,$$

so, since $|Z_y| \geq 0$ and $\text{ind}xy \leq 3$,

$$|Z_y| > \frac{\text{ind}xy - 5t + 2}{6 - 9t + \text{ind}xy}|V'| \geq \frac{\text{ind}xy - 5t + 2}{9(1-t)}|V'|.$$

Therefore, there are at least

$$\sum_{y \in V - x} |Z_y| > \sum_{y \in V - x} \frac{\text{ind}xy - (5t-2)}{9(1-t)} \cdot |V'| > \frac{1 + 2t - (5t-2)}{9(1-t)} \cdot (n-2)^2 = \frac{1}{3}(n-2)^2$$

pairs $(y, z)$ such that $xyz$ is a heavy triangle, and this completes the proof. □

**Claim 3.4.3.** For every pair of distinct vertices $x$ and $y$ there are at least $2\phi^2 n$ ordered pairs $(z, w) \in (V - x - y)^2$ such that $xyz$ and $xyw$ are both heavy triangles.

*Proof.* Assume the contrary. For $0 \leq c \leq 6$, let

$$Z_c := \{z \in V - x - y : \|z, xy\| > c\}.$$

For any $z \in V$ we will say $w \in V - xyz$ *works* with $z$ if both $xzw$ and $yzw$ are heavy triangles.

First note that if $z \in V - x - y$ is such that $\|x, z\|, \|y, z\| > 3t$, then any vertex $w \in V - x - y - z$ such that $\|w, xyz\| \geq 3 + 6t$ works with $z$.

If $z \in Z_{3+3t}$, then, because $\|xyz, V \setminus xyz\| > (3 + 6t + 9\varepsilon)n - 2\text{ind}xyz$, there are at least $2\phi n$ vertices $w$ such that $\|w, xyz\| > 3 + 6t$. By the previous observation, every such $w$ works with $z$. Therefore, we can now assume that $|Z_{3+3t}| < \phi n$.

Since $\|xy, V \setminus xy\| \geq (2 + 4t + 6\varepsilon)n - 2\text{ind}xy$, we have $|Z_{2+4t}| \geq 2\phi n$. Therefore, if for every vertex in $z \in Z_{2+4t}$ there are $\phi n$ vertices that work with $z$, then we are done. Assume that this is not the case, and let $z \in Z_{2+4t}$ such that there are fewer than $\phi n$ vertices work with $z$. Let $G^*$ be the graph obtained from $G$ by removing the vertices in $Z_{3+3t}$ and the vertices that work with $z$ from $G$. Note that we removed at most $2\phi n$ vertices, so $G^*$ has the following properties:

$$\text{(a) } Z_{3+3t} = \emptyset, \text{ (b) no vertices work with } z, \text{ and (c) } \delta_w(G^*) \geq \left(\frac{1+2t}{3} + \frac{\varepsilon}{2}\right)n. \tag{3.26}$$

Assume without loss of generality, that $\text{ind}xz \geq \text{ind}yz$.

Let $V' := V(G^*) \setminus xyz$, $Y := \{w \in V' : yzw \text{ is a heavy triangle}\}$ and $X := V' \setminus Y$.

By (3.26)(c), there exists $w \in V'$ such that $\|w, xyz\| \geq 3 + 6t$. If $\mathrm{ind}xz \geq \mathrm{ind}yz > 3t$, then, because $\mathrm{ind}xw, \mathrm{ind}yw \leq 3$, both $\mathrm{ind}xzw$ and $\mathrm{ind}yzw$ are heavy triangles, contradicting (3.26)(b). Hence $\mathrm{ind}yz \leq 3t$, which implies that if we let $\bar{t} := 1 - t$ and $c := (\mathrm{ind}yz - 3t) + \bar{t} = \mathrm{ind}yz - (3 - 4\bar{t})$, then $c \leq \bar{t}$. Because $z \in Z_{2+4t} = Z_{6-4\bar{t}}$, it must be that

$$\mathrm{ind}xz > 3 - c, \tag{3.27}$$

so $c > 0$. Combining the upper and lower bounds on $c$ gives

$$\bar{t} \geq c > 0. \tag{3.28}$$

Note that

$$w \in Y \text{ if and only if } \|w, yz\| \geq 6t + \bar{t} - c = 5t + 1 - c. \tag{3.29}$$

Therefore, we have that

$$\|yz, V'\| < 3|Y| + \|z, Y\| + (5t+1-c)|X| = \|z, Y\| + (5\bar{t} - 3 + c)|Y| + (5t+1-c)|V'|, \tag{3.30}$$

and, by (3.26)(c),

$$\|yz, V'\| \geq (2 + 4t)|V'| = \left(\bar{t} + c\right)|V'| + (5t + 1 - c)|V'|. \tag{3.31}$$

If $w \in X$, then (3.29) implies $\|w, xyz\| < 5t + 4 - c$. If $w \in Y$ and $\|w, xyz\| \geq 9t + c$, then (3.27) implies that

$$\mathrm{ind}xzw \geq \|w, xyz\| - \|wy\| + \|xz\| > (9t + c) - 3 + (3 - c) = 9t,$$

which contradicts (3.26)(b). Combining this with (3.26)(c), implies

$$(3 + 6t)|V'| < \|xyz, V'\| < (5t + 4 - c)|X| + (9t + c)|Y| = (5t + 4 - c)|V'| - (4\bar{t} - 2c)|Y|.$$

Then combining this with the obvious bound $\|z, Y\| \leq 3|Y|$, (3.30) and (3.31), implies

$$\frac{\bar{t} - c}{4\bar{t} - 2c} > \frac{|Y|}{|V'|} > \frac{\bar{t} + c}{5\bar{t} + c} \text{ which implies } c^2 - 6\bar{t}c + \bar{t}^2 > 0.$$

With (3.28), we have that

$$0 \leq c < \bar{t}(3 - 2\sqrt{2}) < \bar{t}/2. \tag{3.32}$$

Again using the fact that $\|xyz, w\| < 9t + c$ for every vertex $w \in Y$, but this time also using (3.26)(a), we have that

$$(3+3t)|X| + \|z, X\| + (9t+c)|Y| \geq \|xyz, V'\| > (2+4t)|V'| + \|z, X\| + \|z, Y\|$$

so

$$0 > \|z, Y\| - \bar{t}|X| + (5\bar{t} - 3 - c)|Y| = \|z, Y\| - \bar{t}|V'| + (6\bar{t} - 3 - c)|Y|.$$

By (3.30) and (3.31), $\|z, Y\| - (\bar{t} + c)|V'| + (5\bar{t} - 3 + c)|Y| > 0$, so $c|V'| + (\bar{t} - 2c)|Y| < 0$. This contradicts (3.32). □

Now we can quickly prove Lemma 3.4.1. Recall definitions 3.2.4, 3.2.7 and 3.2.8. Claim 3.4.2 implies that $V$ is $(1, \phi, 1)$-linked, so if we let $V_1 := V$ and $\mathcal{M} = \{V_1\}$, then $\mathcal{M}$ is a $(1, \phi, 1)$-linked partition of $V$. Claim 3.4.3 implies that $t_\phi(\mathcal{M}, \{1, 1, 1\}) = 1$ and $F_\phi(\mathcal{M}) = \{\{1, 1, 1\}\}$. Now we can apply Lemma 3.2.9 to $\mathcal{M}$. Let $U$ and $\zeta$ be $A$ and $\eta$ from Lemma 3.2.9, respectively. The set $U$ is the desired set, since when $W \subseteq V \setminus U$ is such that $|W|$ is at most $\zeta n$ and divisible by 3, any parition of $W$ into three parts each of size $|W|/3$ is $(\mathcal{M}, \phi, \eta)$-absorbable with respect to $A$. □

**Lemma 3.4.4** (Triangle Covering Lemma)**.** *For any $\varepsilon > 0$ there exists $n_0$ such that for any $n \geq n_0$, if $(V, E) = G = K_n$ and $w : E \to [0, 1]$ such that $\delta_w(G) \geq \left(\frac{1+2t}{3} + \varepsilon\right) n$ then there is a heavy triangle tiling on all but at most 6 vertices.*

*Proof.* Let $\mathcal{R}$ be a collection of vertex disjoint heavy triangles in $G$, let $U := V(\mathcal{R})$, $W := V \setminus U$, and $\rho := \sum_{T \in \mathcal{R}} \mathrm{ind} T$. Let $M \subseteq E(G[U])$ be a matching such that for every $e \in M$, $\mathrm{ind} e > 3t$, and let $I := W \setminus (\bigcup M)$. Assume that $\mathcal{R}$ and $M$ are picked to maximize the triple $(|\mathcal{R}|, |M|, \rho)$ lexicographically.

Clearly $|W| = 2|M| + |I|$, so the following two claims complete the proof.

**Claim 3.4.5.** $|M| \leq 2$.

*Proof.* Suppose there exist three distinct edges $e_1, e_2, e_3 \in M$. By the maximality of $|\mathcal{R}|$, for $i \in \{1, 2, 3\}$ and any $x \in W - e_i$, $\|e_i, x\| < 6t$. Therefore, $\|e_1, e_2, e_3, W\| \leq 6t|W|$, so

$$\|e_1 \cup e_2 \cup e_3, U\| > 6 \cdot 3\delta_w(G) - 6t|W| > 6 \cdot (1+2t)|U| = (18+36t)|\mathcal{R}|,$$

so there exist $T \in \mathcal{R}$ such that $\|e_1 \cup e_2 \cup e_3, T\| > 18 + 36t$. Without loss of generality assume that $\|e_1, T\| \geq \|e_2, T\| \geq \|e_3, T\|$.

Since $18 \geq \|e_1, T\| > 6 + 12t$, $\|e_2, T\| > 18t$. Now, label $\{t_1, t_2, t_3\} := V(T)$ so that $\|e_1, t_1\| \geq \|e_1, t_2\| \geq \|e_1, t_3\|$. Since $6 \geq \|e_1, t_1\| > 2 + 4t$, we have that $\|e_1, t_2\| > 6t$, and both $e_1 t_1$ and $e_1 t_2$ are heavy triangles. Because $\|e_2, T\| > 18t$, there exists $i \in \{1, 2, 3\}$ such that $\|e_2, t_i\| > 6t$ which implies $e_2t_i$ is a heavy triangle. Let $j \in \{1,2\} - i$. Since $e_1t_j$ and $e_2t_i$ are disjoint heavy triangle, we have violated the maximality of $|\mathcal{R}|$. ☐

**Claim 3.4.6.** $|I| \leq 2$.

*Proof.* Suppose there are disjoint vertices $x_1, x_2, x_3 \in I$. By the maximality of $|\mathcal{R}|$, $\|x_i, e\| < 6t$ for every $e \in M$ and $i \in [3]$. Furthermore, by the maximality of $|M|$, $\|x_i, y\| \leq 3t$ for every $y \in I - x_i$. Therefore, $\|x_1x_2x_2, W\| \leq 3t|W|$ and

$$\|x_1x_2x_3, U\| > 3 \cdot 3\delta_w(G) - 3t|W| > 3 \cdot (1+2t)|U| = (9+18t)|\mathcal{R}|,$$

so there exists $T \in \mathcal{R}$ such that $\|x_1x_2x_3, T\| > 9 + 18t$. Without loss of generality assume that $\|x_1, T\| \geq \|x_2, T\| \geq \|x_3, T\|$.

Note that $9 \geq \|x_1, T\| > 3 + 6t$ which implies $\|x_2, T\| > 9t$ and $\|x_2, t_1\| > 3t$ for some $t_1 \in T$. Therefore, by the maximality of $|M|$, to complete the proof we only need to show that $x_1t_2t_3$ is a heavy triangle where $\{t_2, t_3\} = V(T) - t_1$. For the rest of the proof we will focus on $x_1$ so, for notation simplicity, let us define $x := x_1$.

Now suppose $xt_2t_3$ is not a heavy triangle, i.e.

$$\text{ind}xt_2 + \text{ind}xt_3 + \text{ind}t_2t_3 \leq 9t. \tag{3.33}$$

Note that for any labeling $\{i,j,k\} = \{1,2,3\}$ since $\text{ind}xt_k \leq 3$, we have $\|x, t_it_j\| > 6t$, so $xt_it_j$ is a heavy triangle when $\text{ind}t_it_j \geq 3t$. Therefore, $\text{ind}t_2t_3 < 3t$, and, furthermore, because $t_1t_2t_3$ is a heavy triangle, we have that $\text{ind}t_1t_2 + \text{ind}t_1t_3 > 6t$. Assume without loss of generality, that $\text{ind}t_1t_2 \geq \text{ind}t_1t_3$, so $\text{ind}t_1t_2 > 3t$. This implies that $xt_1t_2$ is a heavy triangle, and, by the maximality of $\rho$,

$$\text{ind}xt_1 + \text{ind}xt_2 \leq \text{ind}t_1t_3 + \text{ind}t_2t_3. \tag{3.34}$$

Furthermore, since $\text{ind}xt_1 + \text{ind}xt_2 > 6t$ and $\text{ind}t_2t_3 < 3t$, this implies $\text{ind}t_1t_3 > 3t$. Therefore, $xt_1t_3$ is a heavy triangle, and, again by the maximality of $\rho$,

$$\text{ind}xt_1 + \text{ind}xt_3 \leq \text{ind}t_1t_2 + \text{ind}t_2t_3. \tag{3.35}$$

By (3.33), $\text{ind}t_2t_3 \leq 9t - (\text{ind}xt_2 + \text{ind}xt_3)$. Combining this with (3.34) and (3.35), we get that

$$2\text{ind}xt_1 + \text{ind}xt_2 + \text{ind}xt_3 \leq \text{ind}t_1t_2 + \text{ind}t_1t_3 + 18t - 2(\text{ind}xt_2 + \text{ind}xt_3).$$

Hence,

$$\text{ind}xt_2 + \text{ind}xt_3 + 2\|x, T\| \leq \text{ind}t_1t_2 + \text{ind}t_1t_3 + 18t.$$

This is a contradiction, because

$$\text{ind}xt_2 + \text{ind}xt_3 + 2\|x, T\| > 6t + 2(3 + 6t) = 6 + 18t \text{ and } \text{ind}t_1t_3 + \text{ind}t_1t_2 + 18t \leq 6 + 18t.$$

□

## 3.5 Concluding Remarks

In this paper we answered Question 3.1.1 for $k = 3$, but this question remains open for $k \geq 4$. We now give constructions which show that the minimum degree necessary for Questions 3.1.1 is at least $(\frac{k-2}{k} + o(1))n$ for every $k \geq 4$. In the following constructions, we call an $n$ vertex triangle-free graph with independence number $o(n)$ and minimum degree $o(n)$ an *Erdős graph.*

For the case $k = 2\ell + 1$, consider the complete $(\ell + 1)$-partite graph with one part $V_0$ of size $n/k - 1$, another part $V_1$ of size $2n/k + 1$ and the remaining parts $V_2, \ldots, V_\ell$ each of size $2n/k$. To complete the construction, for $i = 0, \ldots, \ell$, put a copy of an Erdős graph on the set $V_i$. This graph does not have a $K_k$-tiling, because each $K_k$ has at most 2 vertices in $V_1$ and a $K_k$-tiling can have at most $n/k$ copies of $K_k$. The minimum degree of this graph is $(\frac{k-2}{k} + o(1))n$ and it has sublinear independence number. Note that this construction has the additional property of being $K_{k+2}$-free. For the case $k = 2\ell$, start with the complete $\ell$-partite graph with parts $V_1, \ldots, V_\ell$ where $V_1$ has size $2n/k + 1$, $V_2$ has size $2n/k - 1$ and the remaining parts each have size $2n/k$, and place an Erdős graph on each of the parts $V_1, \ldots, V_\ell$. This again gives a graph with no $K_k$-factor, sublinear independence number and minimum degree $(\frac{k-2}{k} + o(1))n$. Note that, in this case, the graph is $K_{k+1}$-free.

Another question, motivated by the fact that all of our examples which show that the minimum degree condition in Theorem 3.1.2 is asymptotically sharp contain very large cliques, is the following.

**Question 3.5.1.** Let $G$ be an $n$-vertex $K_r$-free graph with $\alpha(G) = o(n)$ for some constant $r \geq 4$. What is the minimum degree condition on $G$ that guarantees a triangle tiling in $G$?

For the case $r = 4$, we use a modified version of the Bollobás-Erdős graph [10] to construct a lower bound. For every large even $n$, the Bollobás-Erdős graph is an $n$-vertex $K_4$-free graph with independence number $o(n)$. The vertex set is the disjoint union of two sets $V_1$ and $V_2$ of the same order such that the graphs $G[V_1]$ and $G[V_2]$ are triangle-free and

$d(v_i, V_{3-i}) \geq (1/4 - o(1))n$ for every $v_i \in V_i$. To construct our example, we start with the Bollobás–Erdős graph on $4/3n+2$ vertices, and then remove a random subset of size $n/3+2$ from one of the two parts. Note that the two parts now have sizes $n/3-1$ and $2n/3+1$. With high-probability, this gives a $K_4$-free with minimum degree $(1/6-o(1))n$ that does not have a triangle factor. We call this construction the modified Bollobás-Erdős graph on $n$ vertices.

For the case $r=5$, we can use the example above with $k=3$, i.e. just the parts $V_0$ and $V_1$, to show that we need $\delta(G) \geq (1/3+o(1))n$. It might be true that instead of $K_5$, forbidding any larger clique does not affect the bound on the minimum degree.

**Question 3.5.2.** Let $G$ be an $n$-vertex $K_r$-free graph with $\alpha(G) = o(n)$ for some constant $r \geq 5$. Is $\delta(G) \geq (1/3+o(1))n$ sufficient for the existence of a triangle tiling?

Noga Alon commented that if one is only looking for $n/3-1$ vertex disjoint triangles, instead of a triangle factor, then maybe the minimum degree condition $(1/3+o(1))n$ is sufficient (with no condition on the clique number).

One can also consider a more general question.

**Question 3.5.3.** Let $r, k$ be such that $r > k$, let $G$ be an $n$-vertex $K_r$-free graph with $\alpha(G) = o(n)$. What is the minimum degree condition on $G$ that guarantees a triangle tiling in $G$?

When $k$ is even and $r = k+1$, the example above shows that the minimum degree must be at least $(\frac{k-2}{k}+o(1))n$. Note that this minimum degree condition agrees with the minimum degree condition in Question 3.5.2. When $k = 2\ell+1$ and $r = k+1$, we can modify the construction above by replacing the parts $V_0$ and $V_1$ with the modified Bollbás Erdős graph on $3n/k$ vertices. The minimum degree of this graph is $\left(\frac{k-2}{k} - \frac{1}{2k} - o(1)\right)n = \left(\frac{2k-5}{2k} - o(1)\right)n$.

It should also be noted that when $\alpha(G)$ is at most a constant, the fact that $G$ has a $K_k$ tiling on all but at most a constant number of vertices is a direct consequence of Ramsey's Theorem. Furthermore, when we add the condition $\delta(G) \geq (1/2+\varepsilon)n$, a counting argument and Ramsey's Theorem show that there are $\Omega(n^{k-1})$ copies of $K_{k-1}$ in the intersection of the neighborhoods of any two distinct vertices, so the absorbing method gives a $K_k$-factor.

# Chapter 4

# Ramsey-Turán-type of extremal problems

Many classical results in extremal graph theory provide sufficient conditions for the appearance of a certain structure. For example, the fundamental theorem of Ramsey states that one can find a monochromatic clique of a given size in any edge-coloring of a sufficiently large graph. Another example is Turán theorem which determines the maximum size of a graph without a fixed size clique. The extremal example for this theorem is Turán graph, an $n$-vertex complete $k$-partite graph, denoted by $T_k(n)$, where all the partite sets have size $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$. Motivated by the fact that the Turán graph has linear-sized independent sets, Erdős and Sós initiated the so-called Ramsey-Turán theory, where they studied the maximum size of an $H$-free graph $G$ with the additional condition that $\alpha(G) = o(|G|)$. Here, we will study the Ramsey-Turán variation of some classical results, whose extremal graphs are close to the Turán graph.

## 4.1 Introduction

One of the central topics in extremal combinatorics is Turán-type problems: Given an integer $n$ and graphs $F$ and $H$, determine $\mathrm{ex}(n, F, H)$, i.e. the maximum number of copies of $F$ in an $n$-vertex $H$-free graph. Mantel theorem [48] and Turán theorem [61] are the first results of this type. Since then, this function has been studied for many different pairs of graphs $F$ and $H$ (see [55] for a survey, where they considered the case $F = K_2$). The following theorem of Erdős [17] determines $\mathrm{ex}(n, K_\ell, K_r)$ for all $\ell < r$. Note that for $\ell = 2$ it is Turán theorem [61].

**Theorem 4.1.1.** *Let $r > \ell \geq 2$ and $n$ be positive integers. Among all $n$-vertex $K_r$-free graphs, the $T_{r-1}(n)$ has the maximum number of $K_\ell$'s.*

In 1970, motivated by Ramsey and Turán theorem, Erdős and Sós [24] introduced the function $\mathrm{RT}(n, H, f(n))$, which is the maximum number of edges of an $n$-vertex $H$-free graph $G$ with $\alpha(G) \leq f(n)$. The problem of determining $\mathrm{RT}(n, H, f(n))$ is called a *Ramsey-Turán-type problem*, and in the last forty years there has been a significant amount of research on this topic (see [56] for a survey). The function $\mathrm{RT}(n, H, f(n))$ can be viewed as a variation of $\mathrm{ex}(n, F, H)$ for $F = K_2$. Therefore one natural question is to determine this variation for other graphs $F$. Denote $\mathrm{RT}(F, H, f(n))$ the *generalized Ramsey-Turán function*, which is the maximum number of copies of $F$ in an $H$-free $n$-vertex graph $G$ with $\alpha(G) \leq f(n)$.

**Definition 4.1.2.** For graphs $F$, $H$, and a function $f(n)$, let

$$\begin{aligned}\mathrm{RT}(n, H, o(f(n))) &= n^2 \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathrm{RT}(n, H, \varepsilon f(n))}{n^2}, \quad \text{and} \\ \mathrm{RT}(F, H, o(f(n))) &= n^{|F|} \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathrm{RT}(F, H, \varepsilon f(n))}{n^{|F|}}.\end{aligned}$$

The following problem [8] is a Ramsey-Turán variation of Theorem 4.1.1.

**Problem 4.1.3.** Determine $\mathrm{RT}(K_s, K_t, o(n))$ for all $s < t$.

We answer Problem 4.1.3 for various values of $s$ and $t$.

**Theorem 4.1.4.** $\mathrm{RT}(K_s, K_{s+2}, o(n)) = \left(2^{-\binom{s}{2}} + o(1)\right) \left(\frac{n}{s}\right)^s$.

The extremal example for Theorem 4.1.4 is based on the Bollobás-Erdős graph (see Section 1.1 for more details). In [10], Bollobás and Erdős constructed a dense, $K_4$-free graph with independence number $o(n)$ (see Section 1.1 for more details). Here, we consider an $s$-partite graph $G$ with vertex set $V_1, \cdots, V_s$, such that $G[V_i \cup V_j]$ is a copy of the Bollobás-Erdős graph for all $1 \leq i, j \leq s$. Before stating our next result, we will define

$$a_t = \begin{cases} 0 & t = 2, 3, 4, \\ \left(\frac{1}{2}\right)^3 \left(\frac{1}{3}\right)^3 & t = 5, \\ \max\limits_{0 \leq x \leq 1} \binom{\ell-2}{3} \left(\frac{1-x}{\ell-2}\right)^3 + x \binom{\ell-2}{2} \left(\frac{1-x}{\ell-2}\right)^2 + \frac{1}{2} \cdot \left(\frac{x}{2}\right)^2 (\ell - 2) \left(\frac{1-x}{\ell-2}\right) & t = 2\ell, t \geq 6, \\ \left(\frac{1}{\ell}\right)^3 \binom{\ell}{3} & t = 2\ell + 1, t \geq 7. \end{cases} \tag{4.1}$$

The following construction explains the meaning of $a_t$.

**Construction 4.1.5.** For any $t \geq 4$, $a_t n^3$ counts the number of triangles in the following graphs. For $t = 2\ell + 1$, let $G$ be a complete balanced $\ell$-partite graph with partite sets $V_1$, $V_2$, ..., $V_\ell$ and put a copy of Kim's graph (see Section 1.1 for more details) inside each partite set. For $t = 2\ell$, let $G$ be a balanced $\ell$-partite graph with partite sets $V_1$, $V_2$, ..., $V_\ell$. Now, let $G[V_1 \cup V_2]$ be a copy of the Bollobás-Erdős graph and for all other pairs of $1 \le i < j \le \ell$, $G[V_i \cup V_j]$ is the complete bipartite graph. Next, for all $3 \le i \le \ell$, put a copy of Kim's graph inside $V_i$. In this case, we solve an optimization problem to find the size of $V_i$'s that maximizes the number of triangles, which is how $a_\ell$ was defined in (4.1).

The following theorem answers Question 4.1.3 for $s = 3$ and all integers $t \ge 4$.

**Theorem 4.1.6.** *Let $t \ge 4$ be an integer, then as $n$ tends to infinity*

$$\mathrm{RT}(K_3, K_t, o(n)) = a_t n^3(1 + o(1)),$$

*where $a_t$ is defined in* (4.1).

We need the following definition to state our next result.

**Definition 4.1.7.** For an integer $s \ge 2$, a graph $H$ and two functions $f(n) \le g(n)$, we say that the Ramsey-Turán function for $H$ exhibits a *jump* or has a *phase transition* from $g(n)$ to $f(n)$ if

$$\limsup_{n\to\infty} \frac{\mathrm{RT}(n, H, f(n))}{n^2} < \liminf_{n\to\infty} \frac{\mathrm{RT}(n, H, g(n))}{n^2}.$$

Otherwise, if

$$\mathrm{RT}(n, H, f(n)) = (1 + o(1))\mathrm{RT}(n, H, g(n)),$$

we say that the Ramsey-Turán function for $H$ is *stable* from $g(n)$ to $f(n)$.

For a function $w(n)$ define $g_r(n) = n2^{-\omega(n)\log^{1-1/r} n}$. For the rest of this chapter, let $\omega(n)$ be a function of $n$ such that $\omega(n) \to \infty$ arbitrary slowly. Balogh, Hu and Simonovits [5] showed that the Ramsey-Turán function for the even clique $K_{2r}$ exhibits a jump from $o(n)$ to $g_r(n)$:

**Theorem 4.1.8.** *For every integer $r \ge 2$,*

$$\limsup_{n\to\infty} \frac{\mathrm{RT}(n, K_{2r}, g_r(n))}{n^2} < \liminf_{n\to\infty} \frac{\mathrm{RT}(n, K_{2r}, o(n))}{n^2}.$$

Our next result shows a similar phenomenon for the generalized Ramsey-Turán function.

**Theorem 4.1.9.** *(i)* $\mathrm{RT}(K_3, K_5, g_3(n)) = o(n^3)$ *and* $\mathrm{RT}(K_3, K_6, g_3(n)) = o(n^3)$.

*(ii) Odd cliques larger than 5 are stable: for every* $\ell \geq 3$,

$$\mathrm{RT}(K_3, K_{2\ell+1}, g_{\ell+1}(n)) = (1+o(1))\mathrm{RT}(K_3, K_{2\ell+1}, o(n)).$$

*(iii) Even cliques always exhibit a jump: for every* $\ell \geq 3$,

$$\begin{aligned}\mathrm{RT}(K_3, K_{2\ell+2}, g_{\ell+1}(n)) &= (1+o(1))\mathrm{RT}(K_3, K_{2\ell+1}, o(n)) = (1+o(1))a_{2\ell+1}n^3,\\ \mathrm{RT}(K_3, K_{2\ell+2}, o(n)) &= (1+o(1))a_{2\ell+2}n^3,\end{aligned}$$

*where* $a_{2\ell+1}$ *and* $a_{2\ell+2}$ *is defined in* (4.1)*. Note that by taking* $x = 1/\ell$ *in* (4.1)*, we have* $a_{2\ell+2} > a_{2\ell+1}$.

First, Bollobás and Győri [11] and more recently, Alon and Shikhelman [3] studied the maximum number of triangles that a $C_5$-free graph can have, i.e. $\mathrm{ex}(n, K_3, C_5)$. Their main result was that $\mathrm{ex}(n, K_3, C_5) = \Theta(n^{3/2})$. It was noted in [3] that it would be of some interest to prove sparse versions of their results. In particular, what can one say about the expected number of triangles in a $C_5$-free subgraph of the Erdős-Rényi random graph $G(n,p)$. Here, we made some steps toward this direction, below we summarize our observations. We think that everywhere the lower bounds are close to be the best possible. Note that for $p = n^{-1/3+o(1)}$, the lower and the upper bounds are both $n^{4/3+o(1)}$. Also, it is surprising that there are many different phases.

**Theorem 4.1.10.** *There exist constants* $c$ *and* $C$*, such that w.h.p. the followings hold:*

*(i) If* $n^{-1} \ll p \ll n^{-7/12}$ *then* $\mathrm{ex}(G(n,p), K_3, C_5) = (\frac{1}{6} + o(1))p^3n^3$;

*(ii) If* $n^{-7/12} \ll p \ll n^{-5/9}$ *then* $n^{5/4-o(1)} \leq \mathrm{ex}(G(n,p), K_3, C_5) \leq (\frac{1}{6}+o(1))p^3n^3$;

*(iii) If* $n^{-5/9} \ll p \ll n^{-3/8}$ *then* $n^{5/4-o(1)} \leq \mathrm{ex}(G(n,p), K_3, C_5) \leq Cn^{4/3}(\log n)^4$;

*(iv) If* $n^{-3/8} \ll p \ll n^{-1/3}$ *then* $\frac{p^2n^2}{8}(1+o(1)) \leq \mathrm{ex}(G(n,p), K_3, C_5) \leq Cn^{\frac{4}{3}}(\log n)^4$;

*(v) If* $n^{-1/3}(\log n)^4 \leq p \ll n^{-1/6}$ *then* $n^{4/3-o(1)} \leq \mathrm{ex}(G(n,p), K_3, C_5) \leq Cp^{1/2}n^{3/2}$;

*(vi) If* $n^{-1/6}(\log n)^2 \leq p \ll 1$ *then* $cpn^{3/2} \leq \mathrm{ex}(G(n,p), K_3, C_5) \leq Cp^{1/2}n^{3/2}$.

Denote $F(n,r,k)$ the maximum number of $r$-edge colorings that an $n$-vertex graph can have without a monochromatic copy of $K_k$. More than thirty years ago, Erdős and Rothschild [19] conjectured that, for sufficiently large $n$, $F(n,2,k) = 2^{\mathrm{ex}(n,K_2,K_k)}$. This conjecture was proved for $k=3$ by Yuster [63]. In 2004, Alon, Balogh, Keevash and Sudakov [1] proved this conjecture for all $k \geq 3$.

For positive integers $r$ and $k$, and a function $f(n)$, we define $F(r,k,f(n))$ to be the Ramsey-Turán variation of $F(n,r,k)$, i.e., the maximum number of $r$-edge colorings that an $n$-vertex graph with independence number at most $f(n)$ can have without a monochromatic copy of $K_k$. One natural guess is that $F(r,k,f(n))$ will have similar behavior as $F(n,r,k)$, i.e. $F(r,k,f(n)) = r^{RT(n,K_k,f(n))(1+o(1))}$. The following example shows that this is not true even for $r = 2$, $k = 4$ and $f(n) = o(n)$, more precisely $F(2,4,o(n)) \geq 2^{n^2/4}$, note that $\mathrm{RT}(n,K_4,o(n)) = \frac{n^2}{8}(1+o(1))$.

Let $G$ be an $n$-vertex complete balanced bipartite graph and put a copy of Kim's graph (defined in Section 1.1) inside each partite set. Consider the following set of 2-edge-colorings of $G$. All the edges inside one part are red and the other part are blue. For the rest of the edges, we can color them either red or blue. Note that since Kim's graph is triangle-free, these colorings do not contain a monochromatic copy of $K_4$. Hence the total number of 2-edge colorings of $G$ with no monochromatic copy of $K_4$ is at least $2^{n^2/4}$. Our next theorem shows that this bound is asymptotically sharp.

**Theorem 4.1.11.** *Let $G$ be an $n$-vertex graph with $\alpha(G) = o(n)$. Then the number of 2-edge-colorings of $G$ without a monochromatic copy of $K_4$ is at most $2^{(1/4+o(1))n^2}$.*

**Organization.** We first introduce some tools in Section 4.2; we will prove Theorem 4.1.4 in Section 4.3. In Section 4.4, we state and prove a main lemma and show how it implies Theorem 4.1.6. In Section 4.5, the proof of Theorem 4.1.9 will be given. In Section 4.6, we will prove Theorem 4.1.10. We will prove Theorem 4.1.11 in Section 4.7.

**Notation.** Let $G = (V,E)$ be an $n$-vertex graph. Denote $k_3(G)$ the number of triangles in $G$. For every $A, B \subseteq V(G)$, let $G[A]$ be the subgraph of $G$ induced by vertex set $A$ and $G[A \cup B]$ be the bipartite subgraph of $G$ induced on partite sets $A$ and $B$. Also, for an $r$-coloring of $E(G)$ with colors $\{c_1,\dots,c_r\}$, let $G_{c_i}[A]$ be the $c_i$-colored subgraph of $G$ induced by the vertex set $A$. We will write $G_{c_i}$ instead of $G_{c_i}[V(G)]$.

## 4.2 Preliminaries

We will use the following theorem of Morris and Saxton [52].

**Theorem 4.2.1.** *There exists a constant $C$ such that*

$$ex(G(n,p),C_4) \leq \begin{cases} Cn^{4/3}(\log n)^2 & \text{if } p \leq n^{-1/3}(\log n)^4, \\ Cp^{1/2}n^{3/2} & \text{otherwise.} \end{cases}$$

*The second bound is sharp up to the constant factor $C$.*

We also need the following result of Erdős and Gallai [20].

**Theorem 4.2.2.** *Let $G$ be an $n$-vertex graph. If $G$ does not contain a path with more than $\ell$ edges then $e(G) \le \frac{1}{2}n\ell$.*

We will use the following definitions.

**Definition 4.2.3.** A *weighted graph* $G$ is an ordered triple $(V, E, w)$ where $E$ is the set of all unordered pairs of distinct vertices, and $w : E \to \{0, 1/2, 1\}$. Define $G_{1/2} = (V, E_{1/2})$ where $E_{1/2} = \{e \in E : w(e) \ge 1/2\}$ and $G_1 = (V, E_1)$ where $E_1 = \{e \in E : w(e) = 1\}$.

A set of three distinct vertices $\{x, y, z\}$ is a triangle if $w(xy) \ne 0$, $w(xz) \ne 0$, and $w(yz) \ne 0$. For a triangle $T = xyz$, define $w(T) = w(xy)w(xz)w(yz)$. Denote $T(G) = \sum_{T\in\mathcal{T}(G)} w(T)$, where $\mathcal{T}(G)$ is the set of all triangles in $G$. For two sets $U, U' \subseteq V$ and a vertices $u, v \in V$, define

$$e(G[U]) = \sum_{e\in E(G[U])} w(e), \qquad T(G[U]) = \sum_{T\in\mathcal{T}(G[U])} w(T),$$
$$T_v(G[U]) = \sum_{T\in\mathcal{T}(G[U]),v\in T} w(T), \qquad T_v(G) = \sum_{T\in\mathcal{T}(G),v\in T} w(T),$$
$$T_{uv}(G[U]) = \sum_{\substack{T\in\mathcal{T}(G[U]),\\ u,v\in T}} w(T), \qquad T_{uv}(G) = \sum_{\substack{T\in\mathcal{T}(G),\\ u,v\in T}} w(T).$$

For a given weighted graph $G = (V, E, w)$ and $X \subseteq Y \subseteq V$, we call $(X, Y)$ a *weighted clique* or *weighted complete subgraph* of size $\ell$ if $X^2 \subseteq E_1$ and $Y^2 \subseteq E_{1/2}$ and $|X| + |Y| = \ell$. Also, let the *weighted clique number* of $G$ be the size of the largest weighted complete subgraph of $G$.

**Definition 4.2.4.** For every $\varepsilon > 0$, positive integer $t$, and an $n$-vertex graph $G = (V, E)$, let $C = \{C_1, \dots, C_m\}$ be an $\varepsilon$-regular partition of $V(G)$ given by the Szemerédi Regularity Lemma 1.3.2 with $m \ge t$. Denote $R$ the *cluster graph* with respect to $C$ and minimum density $10\varepsilon$. We now define the *weighted cluster graph*, $R = (C, w)$, on the vertex set $C$ as follows. For an $\varepsilon$-regular pair $(C_i, C_j)$, we will define $w(C_i, C_j)$ to be

$$\begin{cases} 0 & \text{if } d(C_i, C_j) \le 10\varepsilon \text{ or } (C_i, C_j) \text{ is an irregular pair}, \\ \frac{1}{2} & \text{if } 10\varepsilon < d(C_i, C_j) \le 1/2 + 10\varepsilon, \\ 1 & \text{if } d(C_i, C_j) > 1/2 + 10\varepsilon. \end{cases} \tag{4.2}$$

We need the following lemma which has been proved in the proof of Theorem 2 in [21].

**Lemma 4.2.5.** *Let $G$ be an $n$-vertex graph with $\alpha(G) = o(n)$. If the weighted cluster graph of $G$, $R(C, w)$, contains a weighted complete subgraph of size $\ell$, then $G$ contains a copy of $K_\ell$.*

Next lemma can be proved with a small modification in the proof of Lemma 4.2.5.

**Lemma 4.2.6.** *Let $G$ be an $n$-vertex graph. If the weighted cluster graph of $G$, $R(C, w)$, contains a weighted clique $(X, Y)$ of size $\ell$ such that $\alpha(G[X]) = o(n)$, then $G$ contains a copy of $K_\ell$.*

We will use the following multicolored version of the Szemerédi regularity lemma (for example, see [36]).

**Theorem 4.2.7.** *For every $\varepsilon > 0$ and integer $r$, there exists an $M$ such that for every $n > M$ and any $r$-coloring of the edges of an $n$-vertex graph $G$ with colors $\{c_1, \ldots, c_r\}$, there exists a partition of $V(G)$ into sets $V_1, \ldots, V_m$, for some $1/\varepsilon < m < M$, which is $\varepsilon$-regular with respect to $G_{c_i}$ for every $1 \leq i \leq r$.*

## 4.3 Proof of Theorem 4.1.4

Note that if there are no restrictions on the independence number, by Theorem 4.1.1, an $n$-vertex $K_5$-free graph can have up to $4\left(\frac{n}{4}\right)^3$ triangles. The following theorem shows that if we bound the independence number of the graph by $o(n)$, then the number of triangles will decrease by a factor of $2/27$. For Theorem 4.1.4, we will provide a proof only for $s = 3$, a similar method works for general $s$, we omit the technical details.

**Theorem 4.3.1.** $RT(K_3, K_5, o(n)) = \left(\frac{1}{8} + o(1)\right)\left(\frac{n}{3}\right)^3$.

*Proof.* (Lower bound) Let $G$ be an $n$-vertex graph with a balanced vertex partition $V_1$, $V_2$, and $V_3$. The vertex sets $V_1$, $V_2$ and $V_3$ are the same set of uniformly distributed points from the high dimensional sphere, as in the Bollobás-Erdős graph. The edge set of $G[V_i \cup V_j]$ is the same copy of BE$(2n/3)$, for all $1 \leq i < j \leq 3$. Note that all $V_i$'s are triangle-free and also $G[V_i \cup V_j]$ is $K_4$-free, hence $G$ is $K_5$-free. We will count the number of triangles with exactly one vertex from each $V_i$. Fix a vertex $v_1 \in V_1$, w.l.o.g. we can assume $v_1 = (1, 0, \ldots, 0)$ is labeled as a point of the sphere. Now, we need to pick a vertex $v_2 \in N_{V_2}(v_1)$. Note that $|d_{V_2}(v_1)| = (1/2 - o(1))n/3$, also we can assume $v_2 = (x_1, x_2, 0, \ldots, 0)$. Since the distance between $v_1$ and $v_2$ is less than $\sqrt{2}$, we have $x_1 \geq 0$. Additionally, w.l.o.g. we can assume that $x_2 \geq 0$. Now, for choosing the third vertex $v_3 \in N_{V_3}(v_1) \cap N_{V_3}(v_2)$, w.l.o.g. we can assume $v_3 = (x_1', x_2', x_3', 0, \ldots, 0)$. If $x_1' \geq 0$ and $x_2' \geq 0$ then the distance between $v_3$ and both $v_1$

and $v_2$ will be less than $\sqrt{2}$. Therefore, number of choices for $v_3$ is at least $(1/4 - o(1))n/3$. Hence, we have

$$k_3(G) \geq \frac{n}{3} \cdot \left(\frac{1}{2} - o(1)\right) \frac{n}{3} \cdot \left(\frac{1}{4} - o(1)\right) \frac{n}{3} = \left(\frac{1}{8} - o(1)\right) \left(\frac{n}{3}\right)^3.$$

Upper bound: Let $G$ be an $n$-vertex $K_5$-free graph with $\alpha(G) = o(n)$. Let $R$ be the cluster graph of $G$. Call an edge $XY$ in $R$ *heavy* if the density $d(X, Y)$ is at least $1/2 + 10\varepsilon$. Then by Lemma 4.2.5, we have that $R$ is $K_4$-free and does not have a triangle with a heavy edge. By Theorem 4.1.1, the number of triangles in the $K_4$-free graph $R$ is at most $(|R|/3)^3$. Not having a triangle with a heavy edge implies that each triangle in $R$ has weight at most $1/8 + 100\varepsilon$, i.e. at most

$$\left(\frac{1}{8} + 100\varepsilon\right) \left(\frac{n}{|R|}\right)^3 \left(\frac{|R|}{3}\right)^3 = \left(\frac{1}{8} + 100\varepsilon\right) \left(\frac{n}{3}\right)^3.$$

Standard argument gives that the number of other types of triangles is $o(n^3)$, we omit the details. □

## 4.4 Proof of Theorem 4.1.6

Before we start the proof of Theorem 4.1.6, we need the following lemma.

**Lemma 4.4.1.** *For every integer $t \geq 4$ and $n$-vertex weighted graph $G = (V, E, w)$ (as in Definition 4.2.3) with no weighted complete subgraph of size $t$, we have*

$$T(G) \leq a_t n^3 (1 + o(1)), \tag{4.3}$$

*where $a_t$ is as in* (4.1).

*Proof.* Let $G = (V, E, w)$ be an $n$-vertex weighted graph that satisfies the hypothesis and has the maximum number of triangles. First, we will apply two rounds of the so-called symmetrization method to the graph $G$. Let $V(G) = \{v_1, \ldots, v_n\}$ such that $T_{v_1}(G) \geq \ldots \geq T_{v_n}(G)$. Define $S_1$ to be the following operation: For every $1 \leq i \leq n$ and $i < j \leq n$, if $v_i v_j \notin G_{1/2}$ then we replace $v_j$ with a copy of $v_i$. Operation $S_1$ will not increase the weighted clique number and will not decrease the number of triangles. Consequently, for every pair of vertices $v_i v_j \notin G_{1/2}$ and for every vertex $v_k \neq v_i, v_j$, we have, after $S_1$, $w(v_i v_k) = w(v_j v_k)$. Therefore in the resulting graph $v_i v_j \notin G_{1/2}$ is an equivalence relation. Denote $\mathcal{A} = \{A_1, \ldots, A_m\}$ the equivalence classes of this relation, i.e. two vertices $u$ and $v$ are in the same class if and only if $uv \notin G_{1/2}$. Therefore, for fixed $1 \le i, j, k \le m$, all the edges between $A_i$ and $A_j$ have equal weights, which we denote by $w(A_iA_j)$. Also, for all vertices $x, x' \in A_i$, $y, y' \in A_j$, and $z, z' \in A_k$, we have

$$T_x(G) = T_{x'}(G) \quad \text{and} \quad T_{xy}(G) = T_{x'y'}(G).$$

Therefore, we define $T_{A_i}(G) = T_x(G)$ and $T_{A_iA_j}(G) = T_{xy}(G)$. Note that if $(X, Y)$ is one of the largest weighted complete subgraphs of $G$, then $|Y| = m$.

We summarize the structure of $G$ as follows: Let $H$ be a weighted graph on vertex set $\{a_1, \ldots, a_m\}$ with all its edges having weight either 1 or 1/2, and $w(a_ia_j) = w(A_iA_j)$. The graph $G$ is a blow-up of $H$ where we replace each $a_i$ with a set of $|A_i|$ vertices, and inside each $A_i$ the weight of all edges is zero.

Our next goal is to show that a second round of symmetrization can be carried out in $G$, in other words, in $H$, $w(a_ia_j) = 1/2$ is an equivalence relation. Without loss of generality we can assume $T_{A_1}(G) \ge \ldots \ge T_{A_m}(G)$. For every $1 \le i, j \le m$, define $S_2(i, j)$ to be the following operation: Change $w(A_jA_k)$ to $w(A_iA_k)$ for all $k \ne i, j$, and denote $G_{A_i}$ the resulting graph. Define $G_{A_j}$ analogously as the graph obtained from applying $S_2(j, i)$ to $G$. The following claim states that when the edges between two classes $A_i, A_j$ have weight 1/2, we can replace vertices in $A_i$ with copies of vertices in $A_j$, or the other way around, without decreasing the number of triangles.

**Claim 4.4.2.** For every pair of integers $1 \le i < j \le m$ with $w(A_iA_j) = 1/2$,

(i) $T_{A_i}(G) = T_{A_j}(G)$;

(ii) $k_3(G_{A_i}) = k_3(G_{A_j}) = k_3(G)$.

*Proof.* Define

$$T^o_{A_i}(G) = T_{A_i}(G) - T_{A_iA_j}(G) \quad \text{and} \quad G' = G \setminus \{A_i \cup A_j\}.$$

Since $T_{A_i}(G) \ge T_{A_j}(G)$, we have

$$T^o_{A_i}(G) + T_{A_iA_j}(G) \ge T^o_{A_i}(G) + T_{A_iA_j}(G) \quad \Leftrightarrow \quad T^o_{A_i}(G) \ge T^o_{A_j}(G). \tag{4.4}$$

For (i), it suffices to show

$$T^o_{A_i}(G) = T^o_{A_j}(G). \tag{4.5}$$

Note that

$$k_3(G) = k_3(G') + |A_i| \cdot T^o_{A_i}(G) + |A_j| \cdot T^o_{A_j}(G) + |A_i| \cdot |A_j| \cdot T_{A_iA_j}(G), \tag{4.6}$$

$$k_3(G_{A_i}) = k_3(G') + (|A_i| + |A_j|) \cdot T^o_{A_i}(G) + |A_i| \cdot |A_j| \cdot T_{A_iA_j}(G_{A_i}). \tag{4.7}$$

Then, since $G$ is maximal,

$$0 \geq k_3(G_{A_i}) - k_3(G) = |A_j| \cdot (T^o_{A_i}(G) - T^o_{A_j}(G)) + |A_i| \cdot |A_j| \cdot (T_{A_iA_j}(G_{A_i}) - T_{A_iA_j}(G)).$$

Therefore, by (4.4), we only need to show $T_{A_iA_j}(G_{A_i}) \geq T_{A_iA_j}(G)$. Let

$$\begin{aligned} V_{1,1/2} &= \left\{A_\ell : w(A_iA_\ell) = 1 \text{ and } w(A_jA_\ell) = \frac{1}{2}\right\}, \\ V_{1/2,1} &= \left\{A_\ell : w(A_iA_\ell) = \frac{1}{2} \text{ and } w(A_jA_\ell) = 1\right\}, \\ V_{1/2,1/2} &= \left\{A_\ell : w(A_iA_\ell) = \frac{1}{2} \text{ and } w(A_jA_\ell) = \frac{1}{2}\right\}, \\ V_{1,1} &= \{A_\ell : w(A_iA_\ell) = 1 \text{ and } w(A_jA_\ell) = 1\}. \end{aligned}$$

Denote $|V_{p,q}| = \sum_{A_\ell \in V_{p,q}} |A_\ell|$ for $p, q \in \{1/2, 1\}$. We have

$$T_{A_iA_j}(G_{A_i}) - T_{A_iA_j}(G) = \left(\frac{1}{2} - \frac{1}{4}\right)|V_{1,1/2}| - \left(\frac{1}{4} - \frac{1}{8}\right)|V_{1/2,1}| = \frac{1}{4}|V_{1,1/2}| - \frac{1}{8}|V_{1/2,1}|,$$

Therefore, it suffices to show $2|V_{1,1/2}| \geq |V_{1/2,1}|$. For the sake of contradiction, assume

$$|V_{1/2,1}| > 2|V_{1,1/2}|. \tag{4.8}$$

We will show that (4.8) contradicts the maximality of $G$. Note that

$$k_3(G_{A_j}) = k_3(G') + (|A_i| + |A_j|) \cdot T^o_{A_j}(G) + |A_i| \cdot |A_j| \cdot T_{A_iA_j}(G_{A_j}). \tag{4.9}$$

By (4.6),(4.7),(4.9), and the maximality of $G$ we have

$$k_3(G_{A_i}) \leq k_3(G) \Leftrightarrow \left(\frac{1}{4}|V_{1,1/2}| - \frac{1}{8}|V_{1/2,1}|\right)|A_i| \cdot |A_j| + |A_j| \cdot T^o_{A_i}(G) \leq |A_j| \cdot T^o_{A_j}(G), \tag{4.10}$$

$$k_3(G_{A_j}) \leq k_3(G) \Leftrightarrow \left(\frac{1}{4}|V_{1/2,1}| - \frac{1}{8}|V_{1,1/2}|\right)|A_i| \cdot |A_j| + |A_i| \cdot T^o_{A_j}(G) \leq |A_i| \cdot T^o_{A_i}(G). \tag{4.11}$$

Then (4.10) and (4.11) imply

$$\left(\frac{1}{4}|V_{1,1/2}| - \frac{1}{8}|V_{1/2,1}|\right)|A_i| + T^o_{A_i}(G) \leq T^o_{A_j}(G),$$
$$\left(\frac{1}{4}|V_{1/2,1}| - \frac{1}{8}|V_{1,1/2}|\right)|A_j| + T^o_{A_j}(G) \leq T^o_{A_i}(G).$$

Therefore,

$$\begin{aligned}
&\left(\frac{1}{4}|V_{1/2,1}| - \frac{1}{8}|V_{1,1/2}|\right)|A_j| \leq T^o_{A_i}(G) - T^o_{A_j}(G) \leq \left(\frac{1}{8}|V_{1/2,1}| - \frac{1}{4}|V_{1,1/2}|\right)|A_i| \\
\Rightarrow\ &\frac{1}{8}|V_{1/2,1}|(2|A_j| - |A_i|) \leq \frac{1}{8}|V_{1,1/2}|(|A_j| - 2|A_i|) \overset{(4.8)}{<} \frac{1}{16}|V_{1/2,1}|(|A_j| - 2|A_i|) \\
\Rightarrow\ &\frac{1}{4}|A_j| - \frac{1}{8}|A_i| < \frac{1}{16}|A_j| - \frac{1}{8}|A_i| \quad \Rightarrow \quad \frac{1}{4}|A_j| < \frac{1}{16}|A_j|,
\end{aligned}$$

a contradiction.

For (ii), by the maximality of $G$, it suffices to show that $k_3(G_{A_i}) + k_3(G_{A_j}) \geq 2k_3(G)$. By (i) and (4.5),(4.6),(4.7),(4.9), we have

$$k_3(G_{A_i}) + k_3(G_{A_j}) - 2k_3(G) = |A_i||A_j| \cdot (T_{A_iA_j}(G_{A_i}) + T_{A_iA_j}(G_{A_j}) - 2T_{A_iA_j}(G)).$$

It is left to show $T_{A_iA_j}(G_{A_i}) + T_{A_iA_j}(G_{A_j}) - 2T_{A_iA_j}(G) \geq 0$. Indeed,

$$\begin{aligned}
&T_{A_iA_j}(G_{A_i}) + T_{A_iA_j}(G_{A_j}) - 2T_{A_iA_j}(G) \\
&= \frac{1}{2}\sum_{\substack{1\leq k\leq m,\\ k\neq i,j}} w(A_iA_k)^2 + \frac{1}{2}\sum_{\substack{1\leq k\leq m,\\ k\neq i,j}} w(A_jA_k)^2 - 2\cdot\frac{1}{2}\sum_{\substack{1\leq k\leq m,\\ k\neq i,j}} w(A_iA_k)w(A_jA_k) \\
&= \frac{1}{2}\sum_{\substack{1\leq k\leq m,\\ k\neq i,j}} (w(A_iA_k) - w(A_jA_k))^2 \geq 0.
\end{aligned}$$

□

Define $S_2$ to be the following operation: For every $1 \leq i \leq n$ and $i < j \leq n$, if $w(A_iA_j) = 1/2$ then change $w(A_jA_k)$ to $w(A_iA_k)$ for all $k \neq i, j$.

**Claim 4.4.3.** The operation $S_2$ is not changing the weighted clique number of $G$.

*Proof.* Let $(X, Y)$ be one of the largest weighted complete subgraph of $G$ of size $\ell$. Note that $|Y|$ is still $m$. Also, since we only repeat this operation for vertices $x$ and $y$ with

$w(xy) = 1/2$, the operation is not changing $|X|$ either. Hence, after repeated applications of this operation, the weighted clique number of $G$ will not change. $\square$

**Claim 4.4.4.** After applying $S_2$, in the resulted weighted graph $xy \notin G_1$ is an equivalence relation.

*Proof.* The reflexivity and symmetry properties are obviously satisfied. For the transitivity property, we need to show that for every three distinct integers $1 \leq i, j, k \leq m$, $w(A_iA_j) \neq 1$ and $w(A_jA_k) \neq 1$ implies $w(A_iA_k) \neq 1$, i.e., if $w(A_iA_j) = 1/2$ and $w(A_jA_k) = 1/2$ then $w(A_iA_k) = 1/2$. Since $w(A_iA_j) = 1/2$, depending on whether $i > j$ or $j > i$, during the above process we changed $w(A_iA_k)$ to $w(A_jA_k)$ or $w(A_jA_k)$ to $w(A_iA_k)$. In both cases, $w(A_iA_k) = w(A_jA_k) = 1/2$. $\square$

Denote $\mathcal{B} = \{B_1, \ldots, B_{m'}\}$ the equivalence classes of this relation, i.e. two vertices $u$ and $v$ are in the same class if and only if $uv \notin G_1$. Notice that the $\mathcal{A}$-partition is a refinement of the $\mathcal{B}$-partition. More importantly, the size of the largest weighted complete subgraph is $m + m'$.

We now study the structure of these partitions.

**Claim 4.4.5.** Each $B_i$ contains at most two $A_j$'s.

*Proof.* Let us assume that $B_1$ contains $k$ $A_j$'s, $A_1, \ldots, A_k$, where $k \geq 3$. Denote $U$ the vertex set of $B_1$ and write $u = |U|$. Note that the edges between two $B_i$'s always have weight 1 and the edges inside an $A_i$ have weight 0 and all the other edges have weight 1/2. We will divide the proof into three cases depending on the value of $k$. In each case, we will modify $B_1$ by splitting it into multiple parts. This modification will only change the weight of the edges with both ends in $U$ and also the equivalence classes $\mathcal{A}$ and $\mathcal{B}$. Then we need to prove that the weighted clique number did not increase, and the number of triangles did not decrease. For the latter, since the weight of the edges with at least one end in $V \setminus U$ remain the same, we only need to show that the number of triangles with two or three vertices in $U$ did not decrease. Therefore, it suffices to show that both $e(U)$ and $T(U)$ did not decrease.

**Case 1:** Assume $k \geq 5$, which implies $u \geq 5$. We will split vertices in $U$ into three parts, $B_{11}$, $B_{12}$ and $B_{13}$, such that $|B_{11}| \leq |B_{12}| \leq |B_{13}| \leq |B_{11}| + 1$ for all $1 \leq i, j \leq 3$. Also, define $A_i = B_{1i}$ for all $1 \leq i \leq 3$. For every $u \in U$ and $v \in V \setminus U$, we will not change $w(uv)$. For all vertices $u, u' \in U$ if they belong to the same $B_{1i}$, let $w(uu') = 0$, otherwise let $w(uu') = 1$. The equivalence classes $\mathcal{A}$ and $\mathcal{B}$ will change to $\{A_1, A_2, A_3, A_{k+1}, \ldots, A_m\}$ and $\{B_{11}, B_{12}, B_{13}, B_2, \ldots, B_{m'}\}$. Since $k \geq 5$, the number of classes in the $\mathcal{A}$ partition decreased by at least two and the number of classes in the $\mathcal{B}$ partition increased by exactly 2, hence, the weighted clique number of $G$ will not increase. Now, we only need to show that the number of triangles in the graph $G$ does not decrease.

$$\text{before:}\quad e(U) \leq \binom{k}{2}\frac{u^2}{k^2}\cdot\frac{1}{2} \leq \frac{u^2}{4},$$

$$\text{after:}\quad e(U) = \begin{cases} 3\cdot\frac{u^2}{9} = \frac{u^2}{3} & \text{if } u \equiv 0 \pmod 3,\\ \frac{(u-1)^2}{9} + 2\cdot\frac{(u-1)(u+2)}{9} = \frac{u^2-1}{3} & \text{if } u \equiv 1 \pmod 3,\\ \frac{(u+1)^2}{9} + 2\cdot\frac{(u-2)(u+1)}{9} = \frac{u^2-1}{3} & \text{if } u \equiv 2 \pmod 3. \end{cases}$$

Therefore $e(U)$ does not decrease for $u \geq 2$. Now, for $T(U)$ we have

$$\text{before:}\quad T(U) \leq \binom{k}{3}\frac{u^3}{k^3}\cdot\frac{1}{8} \leq \frac{u^3}{48},$$

$$\text{after:}\quad T(U) = \begin{cases} \frac{u^3}{27} & \text{if } u \equiv 0 \pmod 3,\\ \frac{(u-1)(u-1)(u+2)}{27} & \text{if } u \equiv 1 \pmod 3,\\ \frac{(u-2)(u+1)(u+1)}{27} & \text{if } u \equiv 2 \pmod 3, \end{cases}$$

which means that $T(U)$ does not decrease if $u \geq 3$.

**Case 2:** Assume $k = 4$ which implies $u \geq 4$. Let us split vertices in $U$ into three parts $A_1$, $A_2$ and $A_3$, such that $|A_1| \leq |A_2| \leq |A_3| \leq |A_1| + 1$. Also let $B_{11} = A_1 \cup A_2$ and $B_{12} = A_3$. For all vertices $u, u' \in U$ if they are in different $B_{1i}$'s then $w(uu') = 1$. If they are both in $B_{11}$ but in different $A_i$'s then $w(uu') = 1/2$, and $w(uu') = 0$ if they are in the same $A_i$. The equivalence classes $\mathcal{A}$ and $\mathcal{B}$ will change to $\{A_1, A_2, A_3, A_5, \ldots, A_m\}$ and $\{B_{11}, B_{12}, B_2, \ldots, B_{m'}\}$. Notice that the number of classes in the $\mathcal{A}$ partition decreased by one and the number of classes in the $\mathcal{B}$ partition increased by one. Hence, the weighted clique number of $G$ will not change.

$$\text{before:}\quad e(U) \leq \binom{4}{2}\frac{u^2}{16}\cdot\frac{1}{2} = \frac{3u^2}{16},$$

$$\text{after:}\quad e(U) = \begin{cases} \frac{1}{2}\cdot\frac{u^2}{9} + 2\cdot\frac{u^2}{9} & \text{if } u \equiv 0 \pmod 3,\\ \frac{1}{2}\cdot\frac{(u-1)(u-1)}{9} + 2\cdot\frac{(u-1)(u+2)}{9} & \text{if } u \equiv 1 \pmod 3,\\ \frac{1}{2}\cdot\frac{(u-2)(u+1)}{9} + \frac{(u-2)(u+1)}{9} + \frac{(u+1)(u+1)}{9} & \text{if } u \equiv 2 \pmod 3. \end{cases}$$

For $u \geq 2$, $e(U)$ does not decrease. We also need to show that $T(U)$ does not decrease.

$$\text{before:} \quad T(U) \leq 4 \cdot \frac{u^3}{4^3} \cdot \frac{1}{8} = \frac{u^3}{4^3} \cdot \frac{1}{2},$$
$$\text{after:} \quad T(U) = \begin{cases} \frac{1}{2} \cdot \frac{u^3}{27} & \text{if } u \equiv 0 \pmod 3, \\ \frac{1}{2} \cdot \frac{(u-1)(u-1)(u+2)}{27} & \text{if } u \equiv 1 \pmod 3, \\ \frac{1}{2} \cdot \frac{(u-2)(u+1)(u+1)}{27} & \text{if } u \equiv 2 \pmod 3. \end{cases}$$

Therefore $T(U)$ will increase for $u \geq 3$.

**Case 3:** Assume $k = 3$, which implies $u \geq 3$. First, suppose that $u \leq 12n/13$, then split vertices in $U$ into two equal parts $B_{11}$ and $B_{12}$. Also define $A_1 = B_{11}$ and $A_2 = B_{12}$. For all vertices $u, u' \in U$ if they are in different $B_{1i}$'s then $w(uu') = 1$, and $w(uu') = 0$ otherwise. The equivalence classes $\mathcal{A}$ and $\mathcal{B}$ will change to $\{A_1, A_2, A_4, \ldots, A_m\}$ and $\{B_{11}, B_{12}, B_2, \ldots, B_{m'}\}$. Notice that the number of classes in the $\mathcal{A}$ partition decreased by one and the number of classes in the $\mathcal{B}$ partition increased by one, hence, the weighted clique number of $G$ will not change. We also have

$$\text{before:} \quad T(U) + e(U)(n-u) \leq \frac{u^3}{3^3} \cdot \frac{1}{2^3} + \frac{1}{2} \cdot \binom{3}{2} \frac{u^2}{9}(n-u),$$
$$\text{after:} \quad T(U) + e(U)(n-u) = \begin{cases} 0 + \frac{u^2}{4} \cdot (n-u) & \text{if } u \text{ is even}, \\ 0 + \frac{(u-1)(u+1)}{4} \cdot (n-u) \geq \frac{u^2}{4.5}(n-u) & \text{if } u \text{ is odd}. \end{cases}$$

Since $u \leq 12n/13$, we have

$$\frac{u^3}{3^3} \cdot \frac{1}{2^3} + \frac{3}{2} \cdot (n-u) \cdot \frac{u^2}{3^2} \leq (n-u) \cdot \frac{u^2}{4.5} \quad \Leftrightarrow \quad \frac{u^3}{6^3} \leq (n-u) \cdot \frac{u^2}{18} \quad \Leftrightarrow$$
$$\frac{u}{12} \leq (n-u) \quad \Leftrightarrow \quad u \leq \frac{12}{13}n.$$

We may now assume that $u > 12n/13$. Let $U'$ be the vertex set of $B_2$ and $u' = |B_2|$. Since $u \geq 12n/13$, $u' < n/13$, and therefore $B_2$ contains at most two $A_i$'s. Note that $u' \leq u/12$. We split $U \cup U'$ into three classes of the same size, $B_0$, $B_1$ and $B_2$. Define $A_0 = B_0$, $A_1 = B_1$, and $A_2 = B_2$. For two vertices $u, u' \in U \cup U'$, if they belong to the same $B_i$ then $w(uu') = 0$, otherwise $w(uu') = 1$. The equivalence classes $\mathcal{A}$ and $\mathcal{B}$ will change to $\{A_0, A_1, A_2, A_5, \ldots, A_m\}$ and $\{B_0, B_1, B_2, B_3, \ldots, B_{m'}\}$. Notice that the number of classes in the $\mathcal{A}$ partition decreased by one and the number of classes in the $\mathcal{B}$ partition increased by one, which implies that the weighted clique number of $G$ will not change. We are left to show that this operation will not decrease $e(U \cup U')$ and $T(U \cup U')$.

$$\text{before:} \quad e(U \cup U') \le \frac{u^2}{3^2} \cdot \frac{3}{2} + uu' + \frac{u'^2}{8} \le \frac{u^2}{6} + \frac{u^2}{12} + \frac{u'^2}{8} = \frac{3u^2}{12} + \frac{u'^2}{8},$$

$$\text{after:} \quad e(U \cup U') = \begin{cases} 3 \cdot \frac{(u+u')^2}{9} & \text{if } u + u' \equiv 0 \pmod 3, \\ \frac{(u+u'-1)(u+u'-1)}{9} + 2 \cdot \frac{(u+u'-1)(u+u'+2)}{9} & \text{if } u + u' \equiv 1 \pmod 3, \\ 2 \cdot \frac{(u+u'-2)(u+u'+1)}{9} + \frac{(u+u'+1)^2}{9} & \text{if } u + u' \equiv 2 \pmod 3. \end{cases}$$

Hence $e(U \cup U')$ is increasing for $u \ge 3$. We also have

$$\text{before:} \quad T(U \cup U') \le \left(\frac{u}{3}\right)^3 \cdot \frac{1}{8} + \frac{3}{2} \cdot \frac{u^2}{3^2} \cdot u' + u \cdot \frac{u'^2}{8} \le \frac{u^3}{6^3} + \frac{u^3}{6 \cdot 12} + \frac{u^3}{8 \cdot 12^2} \le \frac{u^3}{51.5},$$

$$\text{after:} \quad T(U \cup U') = \begin{cases} \frac{(u+u')^3}{27} & \text{if } u + u' \equiv 0 \pmod 3, \\ \frac{(u+u'-1)(u+u'-1)(u+u'+2)}{27} & \text{if } u + u' \equiv 1 \pmod 3, \\ \frac{(u+u'-2)(u+u'+1)(u+u'+1)}{27} & \text{if } u + u' \equiv 2 \pmod 3. \end{cases}$$

Therefore $T(U \cup U')$ increases for $u + u' \ge 3$. □

**Claim 4.4.6.** There is at most one $B_i$ that contains two $A_j$'s.

*Proof.* Now, we know that no $B_i$ contains three or more $A_i$'s. Let us assume that $B_1 = A_1 \cup A_2$ and $B_2 = A_3 \cup A_4$. Denote $U$ the vertex set of $B_1 \cup B_2$, and write $u = |U| \ge 4$. We will split the vertices in $U$ into three equal pieces, $B_{11}$, $B_{12}$ and $B_{13}$, and define $A_1 = B_{11}$, $A_2 = B_{12}$, and $A_3 = B_{13}$. For two vertices $u, u' \in U$ if they are in two different $B_{1i}$'s then $w(uu') = 1$, otherwise $w(uu') = 0$. This operation will change the $\mathcal{A}$ and $\mathcal{B}$ partition to $\{A_1, A_2, A_3, A_5, \ldots, A_m\}$ and $\{B_{11}, B_{12}, B_{13}, B_3, \ldots, B_{m'}\}$, therefore the weighted clique number does not change. We only need to show that $e(U)$ and $T(U)$ do not decrease.

$$\text{before:} \quad e(U) \le \frac{u^2}{4^2} + \frac{u^2}{4} = \frac{5u^2}{16},$$

$$\text{after:} \quad e(U) = \begin{cases} 3 \cdot \frac{u^2}{9} & \text{if } u \equiv 0 \pmod 3, \\ \frac{(u-1)^2}{9} + 2 \cdot \frac{(u-1)(u+2)}{9} & \text{if } u \equiv 1 \pmod 3, \\ \frac{(u+1)^2}{9} + 2 \cdot \frac{(u-2)(u+1)}{9} & \text{if } u \equiv 2 \pmod 3, \end{cases}$$

$$\text{before:} \quad T(U) \le \frac{u^2}{16} \cdot \frac{u}{2} = \frac{u^3}{32},$$

$$\text{after:} \quad T(U) = \begin{cases} \frac{u^3}{27} & \text{if } u \equiv 0 \pmod 3, \\ \frac{(u-1)^2(u+2)}{27} & \text{if } u \equiv 1 \pmod 3, \\ \frac{(u+1)^2(u-2)}{27} & \text{if } u \equiv 2 \pmod 3. \end{cases}$$

It can be easily checked that for $u \ge 4$, both $e(U)$ and $T(U)$ are not decreasing. □

Now, we will use the Claims 4.4.5 and 4.4.6 to complete the proof of Lemma 4.4.1. Let us assume that the extremal graph has $\mathcal{A} = \{A_1, \ldots, A_m\}$ and $\mathcal{B} = \{B_1, \ldots, B_{m'}\}$ partitions. Also, since $\mathcal{A}$ is a refinement of $\mathcal{B}$ and also by Claims 4.4.5 and 4.4.6, we have $m' \leq m \leq m' + 1$. For the case $t = 2\ell + 1$, the graph does not contain a weighted clique of size $2\ell + 1$, which implies $m + m' \leq 2\ell$. Therefore $m' = m = \ell$ will maximize the number of triangles. In particular, the extremal graph is an $\ell$-partite graph with partite sets $B_1 \cup \ldots \cup B_\ell$, where $||B_i| - |B_j|| \leq 1$ for all $1 \leq i < j \leq \ell$. Define $A_i = B_i$ for all $1 \leq i \leq \ell$, and for two vertices $u$ and $v$ if they belong to two different $B_i$'s then $w(uv) = 1$, otherwise $w(uv) = 0$. For the case $t = 2\ell$, the graph does not contain a weighted clique of size $2\ell$ which implies $m + m' \leq 2\ell - 1$. Therefore, in the extremal example, $m' = \ell - 1$ and $m = \ell$. Hence, the extremal example is an $(\ell - 1)$-partite graph, with partite sets $B_1 \cup \ldots \cup B_{\ell-1}$, where $|B_1| = x$ and $|B_i| = (n - x)/(\ell - 2)$ for all $2 \leq i \leq \ell - 1$. Also let $A_1 \cup A_2 = B_1$ such that $||A_1| - |A_2|| \leq 1$, and $A_{i+1} = B_i$ for all $2 \leq i \leq \ell - 1$. For two vertices $u$ and $v$ if they belong to two different $B_i$'s then $w(uv) = 1$. Otherwise, if they both belong to $B_1$ but to different $A_i$'s then $w(uu') = 1/2$, and $w(uu') = 0$ in all other cases. Now, we only need to maximize the number of triangles with respect to $x$, which completes the proof of Lemma 4.4.1. □

*Proof of Theorem 4.1.6.* Construction 4.1.5 shows the lower bound.

For the upper bound, let $G$ be an $n$-vertex $K_t$-free graph with $\alpha(G) = o(n)$. Let $R(C, w)$ be the weighted cluster graph of $G$. By Lemma 4.2.5, we have that $R(C, w)$ does not contain a weighted clique of size $\ell$. Then Lemma 4.4.1 implies the upper bound. □

## 4.5 Proof of Theorem 4.1.9

We will need a lemma by Balogh-Hu-Simonovits (Claim 6.1 in [5]).

**Lemma 4.5.1.** *Let $G$ be an $n$-vertex graph with $\alpha(G) = g_q(n)$, where $g_q(n) = n2^{-w(n)\log^{1-1/q} n}$ and $\omega(n) \to \infty$ arbitrary slowly. If there exists a $K_q$ in the cluster graph of $G$, then $K_{2q} \subseteq G$.*

*Proof of Theorem 4.1.9.* For (i), note that $\mathrm{RT}(K_3, K_5, g_3(n)) \leq \mathrm{RT}(K_3, K_6, g_3(n))$, therefore, it is sufficient to prove that $\mathrm{RT}(K_3, K_6, g_3(n)) = o(n^3)$. By Lemma 4.5.1, if $G$ is an $n$-vertex $K_6$-free graph with $\alpha(G) \leq g_3(n)$ then the cluster graph of $G$ is $K_3$-free, which means that $k_3(G) = o(n^3)$.

For (ii), note that $\mathrm{RT}(K_3, K_{2\ell+1}, g_{\ell+1}(n)) \leq \mathrm{RT}(K_3, K_{2\ell+1}, o(n))$, hence, by Theorem 4.1.6, it is sufficient to prove $\mathrm{RT}(K_3, K_{2\ell+1}, g_{\ell+1}(n)) \geq (1+o(1))\binom{\ell}{3}\left(\frac{n}{\ell}\right)^3$. Construction 4.1.5 shows that this inequality holds.

For (iii), note that $\mathrm{RT}(K_3, K_{2\ell+2}, g_{\ell+1}(n)) \geq \mathrm{RT}(K_3, K_{2\ell+1}, g_{\ell+1}(n))$, hence, using (ii), we only need to show that $\mathrm{RT}(K_3, K_{2\ell+2}, g_{\ell+1}(n)) \leq \mathrm{RT}(K_3, K_{2\ell+1}, o(n))$. By Lemma 4.5.1, if $G$ is an $n$-vertex $K_{2\ell+2}$-free graph with $\alpha(G) \leq g_{\ell+1}(n)$ then the cluster graph of $G$ is $K_{\ell+1}$-free. Then, by Theorem 4.1.1, the $\ell$-partite Turán graph has the maximum number of triangles among all $K_{\ell+1}$-free graphs. Hence, we have

$$\mathrm{RT}(K_3, K_{2\ell+2}, g_{\ell+1}(n)) \leq (1+o(1))\binom{\ell}{3}\left(\frac{n}{\ell}\right)^3 = RT(K_3, K_{2\ell+1}, o(n)),$$

where the last equality is by Theorem 4.1.6. □

## 4.6 Proof of Theorem 4.1.10

In this section we study $\mathrm{ex}(G(n,p), K_3, C_5)$, i.e. the maximum number of triangles in a $C_5$-free subgraph of $G(n,p)$. Theorem 4.1.10 exhibits multiple stages of phase transitions for $\mathrm{ex}(G(n,p), K_3, C_5)$. Note that for $p \ll n^{-1}$, with high probability there are no triangles in $G(n,p)$.

*Proof of Theorem 4.1.10.* First, we will prove that for $p \gg n^{-1}$, w.h.p. the number of triangles in $G(n,p)$ is $(1/6 + o(1)n^3p^3$. Let $T$ be the random variable denoting the number of triangles in $G(n,p)$, then

$$\mathbb{E}(T) = \binom{n}{3}p^3 \quad \text{and} \quad \mathrm{Var}(T) \leq \binom{n}{3}p^3 + 12\binom{n}{4}p^5.$$

By the second moment method, for any constant $\varepsilon > 0$,

$$\mathbb{P}(|T - \mathbb{E}(T)| \geq \varepsilon \cdot \mathbb{E}(T)) \leq \frac{\mathrm{Var}(T)}{\varepsilon^2 \cdot \mathbb{E}(T)^2} = o(1),$$

where the last equality holds for all $p \gg n^{-1}$. Therefore, with high probability

$$T = \left(\frac{1}{6} + o(1)\right) n^3p^3. \tag{4.12}$$

**Lower bounds.** For (i), we say that a copy of $C_5$ in $G(n,p)$ is *good* if it has at least one edge which is not contained in any triangle, otherwise it is *bad*. Notice that we can delete all good $C_5$'s by removing a non-triangle edge in them, without deleting any triangle. Now, we need to show that we can remove $o(p^3n^3)$ triangles to delete all bad $C_5$'s. Therefore, it suffices to show that the number of bad $C_5$'s is $o(p^3n^3)$. A bad $C_5$ has each of its edges in a triangle. Let $\mathcal{H}$ be the family of all such configurations, i.e. a $C_5$ with all of its edges in a triangle (see Figure 4.1). For every $H \in \mathcal{H}$, with $v := v(H)$ and $e := e(H)$, let $X$ be the number of copies of $H$ in $G(n,p)$. We have

$$\mathbb{E}(X) = \binom{n}{v} p^e < n^v p^e.$$

We claim that $n^v p^e \ll p^3 n^3$. Since $p \ll n^{-7/12}$,

$$n^{v-3} p^{e-3} \ll n^{v-3-\frac{7}{12}(e-3)} = n^{v-\frac{7}{12}e-\frac{5}{4}}.$$

Hence, we only need to show that $12v - 7e \leq 15$. It is not hard to see that among all such configurations $H \in \mathcal{H}$, the one that maximizes the value $12v - 7e$ is a $C_5$ with one triangle on each edge. For this configuration we have $v = 10$ and $e = 15$. Hence $12v - 7e = 15$, as desired. Therefore, for any constant $\varepsilon > 0$, by the first moment method

$$\mathbb{P}(X \geq \varepsilon \cdot p^3 n^3) \leq \frac{n^v p^e}{\varepsilon \cdot p^3 n^3} = o(1).$$

We then delete one triangle in each copy of $H$ to make $G(n,p)$ $H$-free. Notice that we have deleted $o(p^3 n^3)$ edges so far.

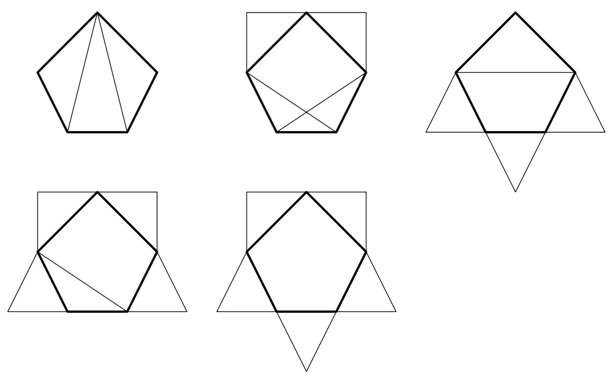

Figure 4.1: The family of all bad $C_5$'s, though vertices not in the $C_5$ can be the same.

One final thing to show is that every edge is in constant many triangles. We claim that each edge is in at most 8 triangles. Define a *book* of size $k$ to be $k$ triangles all sharing one edge. Let $X$ be the random variable for the number of books of size 9 in $G(n,p)$. By the first moment method we have:

$$\mathbb{P}(X \neq 0) \leq n^{11} p^{19} = o(1).$$

Hence, almost surely there is no book of size 9.

For (ii) and (iii), we sparsen the graph $G(n,p)$ to density $n^{-7/12}/\omega(n)$, then, we use the lower bound in (i). In particular, we will take a random subgraph of $G(n,p)$ where each edge appears with probability $p'$, such that $pp' = n^{-7/12}/\omega(n)$. Hence, we can use the bound in (i).

For (iv), we split the vertex set of $G(n,p)$ into two equal parts $X$ and $Y$. Then, we find a perfect matching $M$ on $X$. Note that $G(n,p)$ has a perfect matching with high probability for $p \gg \log n/n$ [23]. Now, we create an auxiliary bipartite graph $H$ with vertex set $M \cup Y$, in which an edge in $M$ is adjacent to a vertex in $Y$ if they form a triangle. Note that $H$ is a random bipartite graph in which each edge is drawn with probability $p^2 \ll n^{-2/3}$, independent of others. Let $C$ be the random variable denoting the number of $C_4$'s in $H$, then we have,

$$\begin{aligned}
\mathbb{E}(e(H)) &= p^2 \cdot |M| \cdot |Y| = p^2 \cdot \frac{n^2}{8}, \\
\mathbb{E}(C) &= (p^2)^4 \binom{|M|}{2}\binom{|Y|}{2} = p^8 \cdot \binom{\frac{n}{4}}{2} \cdot \binom{\frac{n}{2}}{2}, \\
\mathrm{Var}(C) &\leq n^4 p^8 + n^6 p^{14} + n^5 p^{12}.
\end{aligned}$$

Therefore, by Chernoff bound and the second moment method, for any $\varepsilon > 0$ we have

$$\begin{aligned}
\mathbb{P}\Big\{|e(H) - \mathbb{E}(e(H))| \geq \varepsilon \cdot \mathbb{E}(e(H))\Big\} &\leq 2e^{\frac{-\varepsilon^2 p^2 n^2}{24}} = o(1), \\
\mathbb{P}\Big\{|C - \mathbb{E}(C)| \geq \varepsilon \cdot \mathbb{E}(C)\Big\} &\leq \frac{\mathrm{Var}(C)}{\varepsilon^2 \cdot \mathbb{E}(C)^2} = o(1).
\end{aligned}$$

The last equality holds because for $p \gg n^{-1/2}$, we have $n^4p^8 = o(p^{16}n^8)$, $n^6p^{14} = o(p^{16}n^8)$, and $n^5p^{12} = o(p^{16}n^8)$. Now, we can delete one edge from each copy of $C_4$ to make $H$ $C_4$-free. Let $H'$ be the resulted graph. Hence, with high probability,

$$e(H') \geq \left(p^2 \cdot \frac{n^2}{8} - p^8 \cdot \frac{n^2}{16} \cdot \frac{n^2}{4}\right)(1+o(1)) \geq p^2 \cdot \frac{n^2}{8}(1+o(1)),$$

where the last inequality holds because $p \ll n^{-1/3}$. Note that every edge in $H'$ corresponds to a different triangle in $G(n,p)$, thus we obtain a lower bound of $\mathrm{ex}(G(n,p), K_3, C_5) \geq p^2n^2(1/8 + o(1))$.

For (v), again we sparsen the graph to density $n^{-1/3}/\omega(n)$. In particular, we will take a random subgraph of $G(n,p)$ where each edge appears with probability $p'$, such that $pp' = n^{-1/3}/\omega(n)$. Then, we use the lower bound in (iv).

For (vi), we use the same construction as in (iv) and we try to find a large $C_4$-free subgraph in a random auxiliary bipartite graph. The only difference is that we use the second lower bound on $\mathrm{ex}(G(n, p^2), C_4)$ in Theorem 4.2.1.

**Upper bounds.** For (i) and (ii), it follows from (4.12).

For (3)-(6), we use an argument by Alon and Shikhelman [3] and the bounds in Theorem 4.2.1.

Let $G$ be a $C_5$-free subgraph of $G(n, p)$. Note that for every $v \in V(G)$, $N(v)$ is $P_3$-free, therefore by Theorem 4.2.2, $e(N(v)) \leq d(v)$. Then

$$k_3(G) = \frac{\sum e(N(v))}{3} \leq \frac{\sum d(v)}{3} = \frac{2e(G)}{3}. \tag{4.13}$$

Now, we 2-color $V(G)$ independently and uniformly at random with red and blue. Denote $R$ the set of all the red and $B$ be the set of blue vertices, also, set $b = |B|$. Note that we can assume every edge of $G$ is in at least one triangle, otherwise we can delete that edge without decreasing the number of triangles. We assign a vertex $w$ to every $e = uv \in E(G)$ such that $uvw$ is a triangle. Let $G' = (B, E')$, where $E'$ is the set of all $e \in E(G)$ with both ends in $B$ and their assigned vertex in $R$. Notice that since $G$ is $C_5$-free, $E'$ is $C_4$-free. Hence, we have

$$|E'| \leq \mathrm{ex}(G(b, p), C_4).$$

Also, since $\mathbb{E}(b) = n/2$, Chernoff bound implies that with high probability $b = (1+o(1))n/2$, and

$$\mathbb{E}|E'| \leq \mathrm{ex}\left(G\left(\frac{n}{2} + o(n), p\right), C_4\right). \tag{4.14}$$

For every $e \in E(G)$,

$$\mathbb{P}(e \in E') = \frac{1}{8} \quad \Rightarrow \quad \mathbb{E}|E'| = \frac{e(G)}{8}, \tag{4.15}$$

where the last equality is by the linearity of expectation. Hence, by (4.13), (4.14), and (4.15) we have

$$k_3(G) \leq \frac{16}{3}\mathrm{ex}\left(G\left(\frac{n}{2} + o(n), p\right), C_4\right). \tag{4.16}$$

Combining (4.16) with the bounds in Theorem 4.2.1 will give us the upper bounds. □

## 4.7 Proof of Theorem 4.1.11

We will also use this lemma.

**Lemma 4.7.1.** *For every $c > 0$, there exists $a > 0$, such that for every $n$-vertex graph $G$ with $\alpha(G) \leq an$ and any 2-edge-coloring of $G$, $C : E(G) \to \{R, B\}$, the following holds. There exists a partition of $V(G)$ into sets $X$ and $Y$ such that $\alpha(G_R[X]) \leq cn$ and $\alpha(G_B[Y]) \leq cn$.*

*Proof.* We can assume $\alpha(G_R) > cn$ and $\alpha(G_B) > cn$. Let $X_0 = V(G)$, $Y_0 = \emptyset$ and $a \leq c^2$. We iterate the following operation. At step $i$, if $\alpha(G_R[X_{i-1}]) \leq cn$ then we will stop. Otherwise, let $I$ be a maximum independent set in $G_R[X_{i-1}]$. Since $\alpha(G) \leq an$, we have $\alpha(G_B[I]) \leq an$. Hence, we define $X_i := X_{i-1} \setminus I$ and $Y_i := Y_{i-1} \cup I$. Notice that $\alpha(G_B[Y_i]) \leq \alpha(G_B[Y_{i-1}]) + an$.

Let us assume that the iteration stops after $k$ steps, i.e. $\alpha(G_R[X_k]) \leq cn$. Note that $k \leq \frac{n}{cn} = 1/c$, which implies that $\alpha(G_B[Y_k]) \leq k \cdot an \leq cn$. □

*Proof of Theorem 4.1.11.* For an arbitrary small constant $\varepsilon > 0$, let $G$ be a graph with $\alpha(G) \leq \varepsilon n$. For any fixed 2-edge-coloring of $G$, $C : E(G) \to \{\text{red}, \text{blue}\}$, apply Lemma 4.7.1 with some $c > \sqrt{\varepsilon}$. Let $\{X, Y\}$ be the resulted partition. We also apply Theorem 4.2.7 to $G$ with the coloring $C$, and let $\mathcal{P} = \{P_1, \ldots, P_m\}$ be the resulted partition of $V(G)$, where we can require that $\mathcal{P}$ refines the $\{X, Y\}$-partition. Also, let $R_R$ and $R_B$ be the red and blue cluster graphs respectively. We have that

$$\begin{aligned} \text{the number of ways to fix an } \{X, Y\}\text{-partition of } V(G) &\leq 2^n, \\ \text{the number of ways to fix a } \mathcal{P}\text{-partition of } V(G) &\leq m^n. \end{aligned}$$

Now, we will count the number of colorings with a fixed $\{X, Y\}$-partition, $\mathcal{P}$-partition and cluster graphs $R_R$ and $R_B$. First, note that the total number of edges with both ends in one of the $P_i$'s, in between irregular pairs or sparse pairs is at most $o(n^2)$. Hence, total number of ways to color these edges is at most $2^{o(n^2)}$. We claim that $e(R_R \cap R_B) \leq (1/4 + o(1))m^2$. If not, the graph $R_R \cap R_B$ has at least one triangle, $P_i P_j P_k$, and we shall find a monochromatic $K_4$ in it. We can assume that $P_i \cup P_j \subseteq X$, i.e. $\alpha(G_R[P_i \cup P_j]) = o(n)$. Hence, Lemma 4.2.6 shows that the red subgraph of $G$ contains a copy of $K_4$, a contradiction. Therefore, for the total number of colorings, we have

$$F(2, 4, o(n)) \leq 2^n m^n 2^{o(n^2)} 2^{\left(\frac{1}{4} + o(1)\right) m^2 \cdot \left(\frac{n}{m}\right)^2} = 2^{\left(\frac{1}{4} + o(1)\right) n^2}.$$

□

# References


[1] N. Alon, J. Balogh, P. Keevash, and B. Sudakov, The number of edge colorings with no monochromatic cliques, *J. London Math. Soc. (2), 70*, (2004), 273–288.

[2] N. Alon, M. Krivelevich and B. Sudakov, Turán numbers of bipartite graphs and related Ramsey-type questions. *Combin. Probab. Comput.*, 12, (2003), 477–494.

[3] N. Alon and C. Shikhelman, Many $T$ copies in $H$-free graphs, *submitted*.

[4] N. Alon and R. Yuster, H-factors in dense graphs, *J. Combin. Theory B, 66*, (1996), 269–282.

[5] J. Balogh, P. Hu, and M. Simonovits, Phase transitions in Ramsey-Turán theory, *J. Combin. Theory Ser. B, 114*, (2015), 148–169.

[6] J. Balogh, G. Kemkes, C. Lee, and S. J. Young, Towards a weighted version of the Hajnal-Szemerédi theorem, *Combinatorics, Probability and Computing, 22*, (2013), 346–350.

[7] J. Balogh and J. Lenz, Some exact Ramsey-Turán numbers, *Bull. Lond. Math. Soc., 44*, (2012), 1251–1258.

[8] J. Balogh, H. Liu, and M. Sharifzadeh, Two extensions in Ramsey-Turán theory, *In preparation.*

[9] J. Balogh, A. Lo, and T. Molla, Transitive triangle tilings in oriented graphs, *submitted.*

[10] B. Bollobás and P. Erdős, On a Ramsey-Turán type problem, *J. Combin. Theory B, 21*, (1976), 166–168.

[11] B. Bollobás and E. Győri, Pentagons vs. triangles, *Discrete Mathematics, 308*, (2008), 4332–4336.

[12] B. Bollobás and A. Thomason, Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs, *European J. Combin., 19*, (1998), 883–887.

[13] J. Bondy and M. Simonovits, Cycles of even length in graphs, *J. Combin. Theory Ser. B, 16*, (1974), 97–105.

[14] K. Corrádi and A. Hajnal, On the maximal number of independent circuits in a graph, *Acta Math. Acad. Sci. Hungar, 14*, (1963), 423–439.

[15] A. Czygrinow, L. DeBiasio, H. A. Kierstead, and T. Molla, An extension of the Hajnal–Szemerédi theorem to directed graphs, *Combin. Probab. Comput., 24*, (2015), 754-773.

[16] A. Czygrinow, H. A. Kierstead, and T. Molla, On directed versions of the Corrádi–Hajnal Corollary, *European J. Combin., 42*, (2014), 1–14.

[17] P. Erdős, On the number of complete subgraphs contained in certain graphs, *Magyar Tud. Akad. Mat. Kut. Int. Kőzl, 7*, (1962), 459–474.

[18] P. Erdős, Problems and results in graph theory and combinatorial analysis, *Graph theory and related topics*, (Proc. Conf. Waterloo, 1977), Academic Press, New York (1979), 153–163.

[19] P. Erdős, Some new applications of probability methods to combinatorial analysis and graph theory, *Proceedings of the Fifth Southeastern Conference on Combinatorics, Graph Theory and Computing. Congressus Numerantium X*, (1974), 39–51.

[20] P. Erdős and T. Gallai, On Maximal Paths and Circuits of Graphs, *Acta Math. Acad. Sci. Hung., 10*, (1959), 337–356.

[21] P. Erdős, A. Hajnal, V. T. Sós, and E. Szemerédi, More results on Ramsey-Turán type problems, *Combinatorica, 3*, (1983), 69–81.

[22] P. Erdős, A. Hajnal, M. Simonovits, V. T. Sós, and E. Szemerédi, Turán-Ramsey theorems and $K_p$-independence numbers, *Combin. Probab. Comput., 3*, (1994), 297–325.

[23] P. Erdős and A. Rényi, On random matrices, *Magyar Tud. Akad. Mat. Kutat Int. Kzl, 8*, (1964), 455–461.

[24] P. Erdős and V. T. Sós, Some remarks on Ramsey's and Turán's theorem, *Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969)*, (1970), 395–404.

[25] P. Erdős and P. Turán, On some sequences of integers, *Journal of the London Mathematical Society, 11*, (1936), 261–164.

[26] E. Fischer, Variants of the Hajnal–Szemerédi theorem, *J. Graph Theory, 31* (1999), 275–282.

[27] J. Fox and B. Sudakov, Dependent random choice, *Random Structures & Algorithms, 38*, (2011), 68–99.

[28] E. Győri, $C_6$-free bipartite graphs and product representation of squares, *Discrete Math., 165/166*, (1997), 371–375.

[29] A. Hajnal and E. Szemerédi, Proof of a conjecture of P. Erdős, *Combinatorial theory and its applications, II*, (1970), 601–623.

[30] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, *Wiley, New York*, (2000).

[31] H. A. Jung, Eine Verallgemeinerung des $n$-fachen Zusammenhangs für Graphen, *Math. Ann., 187*, (1970), 95–103.

[32] P. Keevash and R. Mycroft, A multipartite Hajnal–Szemerédi theorem, *J. Combin. Theory B, 114*, (2015), 187–236.

[33] P. Keevash and B. Sudakov, Triangle packings and 1-factors in oriented graphs, *J. Combin. Theory B, 99*, (2009), 709–727.

[34] H. A. Kierstead and A. V. Kostochka , An Ore-type Theorem on Equitable Coloring, *J. Combin. Theory B, 98*, (2008), 226–234.

[35] J. H. Kim, The Ramsey number $\mathrm{R}(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures & Algorithms, 7*, (1995), 173–207.

[36] J. Komlós and M. Simonovits, 'Szemerédi regularity lemma and its application in graph theory', Paul Erdős is 80, *Proceedings of Colloquia of the Bolyai Mathematical Society, 2*, (Keszthely, 1993), 295–352.

[37] J. Komlós, G. Sárközy, and E. Szemerédi, Proof of the Alon-Yuster conjecture, *Discrete Mathematics, 235*, (2001), 255–269.

[38] J. Komlós and E. Szemerédi, Topological cliques in graphs, *Combin. Probab. Comput., 3*, (1994), 247–256.

[39] J. Komlós and E. Szemerédi, Topological cliques in graphs II, *Combin. Probab. Comput., 5*, (1996), 79–90.

[40] T. Kővári, V. T. Sós, and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math., 3*, (1954), 50–57.

[41] M. Krivelevich, B. Sudakov,and T. Szabó, Triangle factors in sparse pseudo-random graphs, *Combinatorica, 24*, (2004), 403–426.

[42] D. Kühn and D. Osthus, Four-cycles in graphs without a given even cycle, *J. Graph Theory, 48*, (2005), 147–156.

[43] D. Kühn and D. Osthus, Induced subdivisions in $K_{s,s}$-free graphs of large average degree, *Combinatorica, 24*, (2004), 287–304.

[44] D. Kühn and D. Osthus, Large topological cliques in graphs without a 4-cycle, *Combin. Probab. Comput., 13*, (2004), 93–102.

[45] D. Kühn and D. Osthus, The minimum degree threshold for perfect graph packings, *Combinatorica, 29*, (2009), 65–107.

[46] D. Kühn and D. Osthus, Topological minors in graphs of large girth, *J. Combin. Theory Ser. B, 86*, (2002), 364–380.

[47] A. Lo and K. Markström, $F$-factors in hypergraphs via absorption, *Graphs Combin., 31*, (2015), 679–712.

[48] W. Mantel, Problem 28, *Wiskundige Opgaven, 10*, (1907), 60–61.

[49] W. Mader, An extremal problem for subdivisions of $K_5^-$, *J. Graph Theory, 30*, (1999), 261–276.

[50] R. Martin and E. Szemerédi, Quadripartite version of the Hajnal-Szemerédi theorem, *Discrete Math., 308*, (2008), 4337–4360.

[51] C. McDiarmid, On the method of bounded differences, *Surveys in combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., 141*, (1989), 148–188.

[52] R. Morris and D. Saxton, The number of $C_{2k}$-free graphs, *submitted*.

[53] D. Mubayi and V. T. Sós, Explicit constructions of triple systems for Ramsey-Turán problems, *J. Graph Theory, 52*, (2006), 211–216.

[54] V. Rödl, A. Ruciński and E. Szemerédi, A Dirac-type theorem for 3-uniform hypergraphs, *Combin. Probab. Comput., 15* (2006), 229–251.

[55] M. Simonovits, Paul Erdős influence on extremal graph theory, *The mathematics of Paul Erdős II*, (Springer, Berlin, 1996) 148–192.

[56] M. Simonovits and V. T. Sós, Ramsey-Turán theory, *Discrete Math., 229*, (2001), 293–340.

[57] B. Sudakov and J. Verstraëte, Cycle lengths in sparse graphs, *Combinatorica, 28*, (2008), 357–372.

[58] E. Szemerédi, On graphs containing no complete subgraph with 4 vertices (Hungarian), *Mat. Lapok, 23*, (1972), 113–116.

[59] E. Szemerédi, On sets of integers containing no $k$ elements in arithmetic progression, *Polska Akademia Nauk. Instytut Matematyczny. Acta Arithmetica 27*, (1975), 199–245.

[60] A. Treglown, On directed versions of the Hajnal–Szemerédi theorem, *Combin. Probab. Comput., 24*, (2015), 873–928.

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

[62] H. Wang, Independent directed triangles in a directed graph, *Graphs Combin., 16*, (2000), 453–462.

[63] R. Yuster, The number of edge colorings with no monochromatic triangle, *J. Graph Theory, 21*, (1996), 441–452.