Combinatorics, Probability and Computing (2012) 21, 661–682. © Cambridge University Press 2012
doi:10.1017/S0963548312000181

Perfect Graphs of Fixed Density: Counting and Homogeneous Sets

JULIA BÖTTCHER[1†‡], ANUSCH TARAZ[2‡] and ANDREAS WÜRFL[2‡]

[1] Department of Mathematics, London School of Economics,
Houghton Street, London WC2A 2AE, UK
(e-mail: j.boettcher@lse.ac.uk)
[2] Zentrum Mathematik, Technische Universität München,
Boltzmannstraße 3, D-85747 Garching bei München, Germany
(e-mail: taraz@ma.tum.de, wuerfl@ma.tum.de)

Received 24 November 2010; revised 27 March 2012; first published online 14 May 2012

For $c \in (0,1)$ let $\mathcal{P}_n(c)$ denote the set of n-vertex perfect graphs with density c and let $\mathcal{C}_n(c)$ denote the set of n-vertex graphs without induced C_5 and with density c.

We show that

$$\lim_{n\to\infty} \log_2 |\mathcal{P}_n(c)|/\binom{n}{2} = \lim_{n\to\infty} \log_2 |\mathcal{C}_n(c)|/\binom{n}{2} = h(c)$$

with $h(c) = \frac{1}{2}$ if $\frac{1}{4} \leqslant c \leqslant \frac{3}{4}$ and $h(c) = \frac{1}{2}H(|2c-1|)$ otherwise, where H is the binary entropy function.

Further, we use this result to deduce that almost all graphs in $\mathcal{C}_n(c)$ have homogeneous sets of linear size. This answers a question raised by Loebl and co-workers.

AMS 2010 *Mathematics subject classification*: Primary 05C30; 05C69, 05C17

1. Introduction and results

In this paper we investigate classes of graphs that are defined by forbidding certain substructures. Let $\mathcal{H}$ be such a class. We focus on two related goals: to approximate the cardinality of $\mathcal{H}$ and to determine the structure of a typical graph in $\mathcal{H}$. In particular, we add the additional constraint that all graphs in $\mathcal{H}$ must have the same density c and would like to know how the answer to these questions depends on the parameter c.

The quantity $|\mathcal{H}_n|$, where $\mathcal{H}_n := \{G \in \mathcal{H} : V(G) = [n]\}$, is also called the *speed* of $\mathcal{H}$. Often exact formulas or good estimates for $|\mathcal{H}_n|$ are out of reach. In these cases, however, one might

† The first author was partially supported by FAPESP (Proc. 2009/17831-7), and is grateful to NUMEC/USP, Núcleo de Modelagem Estocástica e Complexidade of the University of São Paulo.
‡ The authors were partially supported by DFG grant TA 309/2-2.

still ask for the asymptotic behaviour of the speed of $\mathcal{H}$. One prominent result in this direction was obtained by Erdős, Frankl and Rödl [9] who considered properties $\mathcal{F}orb(F)$ defined by a single forbidden (weak) subgraph F. They proved that, for each graph F with $\chi(F) \geqslant 3$, the class $\mathcal{F}orb_n(F)$ of n-vertex graphs that do not contain F as a subgraph satisfies $|\mathcal{F}orb_n(F)| = 2^{\mathrm{ex}(F,n)+o(n^2)}$, where $\mathrm{ex}(F,n) := (\chi(F)-2)\binom{n}{2}/(\chi(F)-1)$. In other words, if $\chi(F) \geqslant 3$ then the speed of $\mathcal{F}orb(F)$ asymptotically only depends on the chromatic number of F.

In this paper we are interested in features of the picture at a more fine-grained scale. More precisely, we fix a density $0 < c < 1$ and are interested in the number $|\mathcal{H}_n(c)|$ of graphs on n vertices with property $\mathcal{H}$ and density c. Let $\mathcal{F}orb_n(F,c) = \mathcal{F}orb_n(F) \cap \mathcal{G}_n(c)$, where $\mathcal{G}_n(c)$ is the set of all graphs on vertex set $[n]$ with $\lfloor c\binom{n}{2}\rceil$ edges. Here $\lfloor x \rceil$ denotes the nearest integer to x. For the sake of readability, we will always assume in the following that $c\binom{n}{2}$ is an integer, since rounding issues would not affect our asymptotic considerations.

Straightforward modifications of the proof of the theorem of Erdős, Frankl and Rödl [9] yield the following bounds for $|\mathcal{F}orb_n(F,c)|$ (we will sketch this argument in Section 2.3). Let F be a graph with $\chi(F) = r$. For all $c \in (0, \frac{r-2}{r-1})$, we have

$$\lim_{n\to\infty} \frac{\log|\mathcal{F}orb_n(F,c)|}{\binom{n}{2}} = \tfrac{r-2}{r-1} H\big(\tfrac{r-1}{r-2}c\big), \tag{1.1}$$

where $H(x)$ is the *binary entropy* function, that is, for $x \in (0,1)$ we set $H(x) := -x\log x - (1-x)\log(1-x)$. Here we denote by log the logarithm to base 2. Notice that

$$\lim_{n\to\infty} \log|\mathcal{F}orb_n(F,c)|/\binom{n}{2} = 0$$

for $c > \frac{r-2}{r-1}$, by the theorem of Erdős and Stone [11].

The analogous problem for a graph class $\mathcal{F}orb^*(F)$, characterized by a forbidden *induced* subgraph F, is more challenging and was first considered by Prömel and Steger [20]. They specified a graph parameter, the so-called colouring number $\chi^*(F)$ of F, that serves as a suitable replacement of the chromatic number in the theorem of Erdős, Frankl and Rödl. More precisely, they showed that $|\mathcal{F}orb_n^*(F)| = 2^{\mathrm{ex}^*(F,n)+o(n^2)}$ with $\mathrm{ex}^*(F,n) := \big(\chi^*(F)-2\big)\binom{n}{2}/\big(\chi^*(F)-1\big)$, where $\chi^*(F)$ is defined as follows. A *generalized r-colouring* of F with $r' \in [0,r]$ cliques is a partition of $V(F)$ into r' cliques and $r-r'$ independent sets. The *colouring number* $\chi^*(F)$ is the largest integer $r+1$ such that there is an $r' \in [r]$ for which F has no generalized r-colouring with r' cliques. For example, we have $\chi^*(C_5) = 3$ and $\chi^*(C_7) = 4$.

This naturally extends to hereditary graph properties, *i.e.*, classes of graphs $\mathcal{H}$ which are closed under isomorphism and taking induced subgraphs (and may therefore be characterized by possibly infinitely many forbidden induced subgraphs). Let $\mathcal{F}(r,r')$ denote the family of all graphs that admit a generalized r-colouring with r' cliques. Then the colouring number of $\mathcal{H}$ is

$$\chi^*(\mathcal{P}) := \max\{r+1 : \mathcal{F}(r,r') \subseteq \mathcal{H} \text{ for some } r' \in [0,r]\},$$

and we set

$$\mathrm{ex}^*(\mathcal{H},n) := \big(\chi^*(\mathcal{H})-2\big)\binom{n}{2}\Big/\big(\chi^*(\mathcal{H})-1\big).$$

Observe that this definition implies $\chi^*\big(\mathcal{F}orb^*(F)\big) = \chi^*(F)$. And indeed Alekseev [1] and Bollobás and Thomason [6] generalized the result of Prömel and Steger to arbitrary hereditary graph properties $\mathcal{H}$ and showed that $|\mathcal{P}_n| = 2^{\mathrm{ex}^*(\mathcal{H},n)+o(n^2)}$.

More precise estimates for the speed were given for monotone properties $\mathcal{H}$ (properties that are closed under isomorphisms and taking subgraphs) by Balogh, Bollobás and Simonovits [5], who showed that $2^{\mathrm{ex}^*(\mathcal{H},n)} \leqslant |\mathcal{H}_n| \leqslant 2^{\mathrm{ex}^*(\mathcal{H},n)+cn\log n}$ for some constant c, and for hereditary properties $\mathcal{H}$ by Alon, Balogh, Bollobás and Morris [2], who proved $2^{\mathrm{ex}^*(\mathcal{H},n)} \leqslant |\mathcal{H}_n| \leqslant 2^{\mathrm{ex}^*(\mathcal{H},n)+n^{2-\varepsilon}}$ for some $\varepsilon = \varepsilon(\mathcal{H}) > 0$ and n sufficiently large. Prömel and Steger [18, 19] gave even more precise results for the speed of $\mathcal{F}orb_n^*(C_4)$ and $\mathcal{F}orb_n^*(C_5)$, which they determined up to a factor of $2^{O(n)}$. In fact, they showed in [19] that almost all graphs in $\mathcal{F}orb_n^*(C_5)$ are generalized split graphs, that is, graphs of a rather simple structure which are defined as follows. We say that a graph $G = (V, E)$ admits a *generalized clique partition* if there is a partition $V = V_1 \dot\cup \cdots \dot\cup V_k$ of its vertex set such that $G[V_i]$ is a clique and for $i > j > 1$ we have $e(V_i, V_j) = e(V_j, V_i) = 0$. A graph G is a *generalized split graph* if G or its complement admit a generalized clique partition.

It is illustrative to compare this result to the celebrated strong perfect graph theorem [8]. A graph G is *perfect* if $\chi(G')$ equals the clique number $\omega(G')$ for all induced subgraphs G' of G. The strong perfect graph theorem asserts that *all* graphs without induced copies of odd cycles C_{2i+1}, $i > 1$ and without induced copies of their complements $\overline{C}_{2i+1}$ are perfect. Using this characterization, it is easy to see that generalized split graphs are perfect. Consequently the result of Prömel and Steger implies that *almost all* graphs without induced C_5 are perfect (observe that C_5 is self-complementary).

In this paper, we consider induced C_5-free graphs of density c and provide bounds for their number. In the spirit of the result by Prömel and Steger, we also relate this quantity to the number of n-vertex perfect graphs and generalized split graphs with density c.

Definition 1.1. We define the following graph classes:

$$\begin{aligned}
\mathcal{C}(n,c) &:= \mathcal{F}orb_n^*(C_5, c) := \mathcal{F}orb_n^*(C_5) \cap \mathcal{G}_n(c),\\
\mathcal{P}(n,c) &:= \{G \in \mathcal{G}_n(c) : G \text{ is perfect}\},\\
\mathcal{S}(n,c) &:= \{G \in \mathcal{G}_n(c) : G \text{ is a generalized split graph}\}.
\end{aligned}$$

Observe that for all n and $c \in [0,1]$ we have $\mathcal{S}(n,c) \subseteq \mathcal{P}(n,c) \subseteq \mathcal{C}(n,c)$. Our first main result now bounds the multiplicative error term between $|\mathcal{S}(n,c)|$ and $|\mathcal{C}(n,c)|$. In order to state this we define the following function. Let

$$h(c) := \begin{cases} H(2c)/2 & \text{if } 0 < c < \frac{1}{4},\\ 1/2 & \text{if } \frac{1}{4} \leqslant c \leqslant \frac{3}{4},\\ H(2c-1)/2 & \text{otherwise}. \end{cases} \tag{1.2}$$

Note that the classes of all generalized split graphs, all perfect graphs, and all graphs without induced C_5 are closed under taking complements. Hence, for example, $|\mathcal{C}(n,c)| = |\mathcal{C}(n,1-c)|$ for all $c \in (0,1)$ and h is in fact symmetric in $(0,1)$. Further, note that $H(|2c-1|)/2 = h(c)$ for $c < 1/4$ or $c > 3/4$.

Theorem 1.2. *For all $c \in (0,1)$ we have*

$$\lim_{n\to\infty} \frac{\log_2 |\mathcal{C}(n,c)|}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 |\mathcal{P}(n,c)|}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 |\mathcal{S}(n,c)|}{\binom{n}{2}} = h(c).$$

The proof of this theorem uses Szemerédi's regularity lemma and is given in Section 2.

We remark that Bollobás and Thomason [7] studied related questions of a more general type (see also the references in [7] for earlier results in this direction). They were interested in the probability $\mathbb{P}_{\mathcal{H}} := \mathbb{P}[\mathcal{G}(n,p) \in \mathcal{H}]$ of an arbitrary hereditary property $\mathcal{H}$ in the probability space $\mathcal{G}(n,p)$, and showed that for any $\mathcal{H}$ there are very simple properties $\mathcal{H}^*$ which closely approximate $\mathcal{H}$ in the probability space $\mathcal{G}(n,p)$. In this context, our Theorem 1.2 estimates the probability of $\mathcal{H} = \mathcal{F}orb^*_n(C_5)$ in the probability space $\mathcal{G}(n,m)$ with $m = c\binom{n}{2}$, and states that $\mathcal{H} = \mathcal{F}orb^*_n(C_5)$ is approximated by the property $\mathcal{H}^*$ of being a generalized split graph in $\mathcal{G}(n,m)$. The actual value of the probability $\mathbb{P}_{\mathcal{H}}$ was estimated by Marchant and Thomason in [17] for several properties $\mathcal{H}$, such as $\mathcal{H} = \mathcal{F}orb^*_n(C_5)$ (see [17, 22]). The probabilities $\mathbb{P}[\mathcal{G}(n,p) \in \mathcal{H}]$ and $\mathbb{P}[\mathcal{G}(n, m = p\binom{n}{2}) \in \mathcal{H}]$ are related (but not identical), and we discuss their relation in Section 4.

Let us now move from the question of approximating cardinalities to determining the structure of a typical element in $\mathcal{F}orb^*_n(C_5)$. A well-known conjecture by Erdős and Hajnal [10] states that any family of graphs that does not contain a certain fixed graph F as an induced subgraph must contain a homogeneous set, *i.e.*, a clique or a stable set, which is of size at least some positive power of the number of vertices.

The conjecture is known to be true for certain graphs F, but is open, among others, for $F = C_5$ (see [12]). However, Loebl, Reed, Scott, Thomason and Thomassé [16] recently showed that for any graph F, *almost all graphs* in $\mathcal{F}orb^*_n(F)$ have a homogeneous set of size at least some positive power of n. Moreover, they ask for which graphs F it is true that almost all graphs in $\mathcal{F}orb^*_n(F)$ do indeed have a *linearly* sized homogeneous set.

It may seem at first sight that our estimates derived in Theorem 1.2, carrying an $o(n^2)$ term in the exponent, are too rough to tell us much about the structure of almost all graphs in $\mathcal{F}orb^*_n(C_5)$ or $\mathcal{F}orb^*_n(C_5,c)$. However, we can combine them with the ideas of [16] to answer the question of Loebl, Reed, Scott, Thomason and Thomassé in the affirmative for the case $F = C_5$. In fact, we can prove this assertion even in the case where we again restrict the class to graphs with a given density.

Theorem 1.3. *For $\eta > 0$, let $\mathcal{F}orb^*_{n,\eta}(F,c)$ denote the set of graphs $G \in \mathcal{F}orb^*_n(F,c)$ with $\hom(G) := \max\{\alpha(G), \omega(G)\} < \eta n$. Then, for every $0 < c < 1$ there exists $\eta > 0$ such that*

$$\frac{|\mathcal{F}orb^*_{n,\eta}(C_5,c)|}{|\mathcal{F}orb^*_n(C_5,c)|} \to 0 \quad (n \to \infty).$$

We provide the proof of this theorem in Section 3.

Statements similar to those in Theorem 1.2 and 1.3, for forbidden graphs F other than C_5, seem to require more work.

2. The proof of Theorem 1.2

In this section we prove Theorem 1.2. In Section 2.1 we start with the lower bound, by estimating the number of generalized split graphs with a given density. For the upper bound we need some preparations. We shall apply Szemerédi's regularity lemma, which is introduced in Section 2.2. In Section 2.3 we illustrate how this lemma can be used for counting graphs without a fixed (not necessarily induced) subgraph. In Section 2.4 we explain how to modify these ideas in order to deal with forbidden induced subgraphs. In Section 2.5, finally, we prove the upper bound of Theorem 1.2.

2.1. The lower bound of Theorem 1.2

In this section we estimate the number of generalized split graphs with density c and prove the following lemma, which constitutes the lower bound of Theorem 1.2.

Lemma 2.1. *For all $c, \gamma \in (0,1)$ there is an n_0 such that, for all $n \geqslant n_0$, we have*

$$|\mathcal{S}(n,c)| \geqslant 2^{h(c)\binom{n}{2}-\gamma\binom{n}{2}}.$$

We will use the following bound for binomial coefficients (see, *e.g.*, [14]). For every $\gamma > 0$ there exists n_0 such that, for every integer $m \geqslant n_0$ and for every real $c \in (0,1)$, we have

$$2^{mH(c)-\gamma m} \leqslant \binom{m}{cm} \leqslant 2^{mH(c)}. \tag{2.1}$$

We call the term $-\gamma m$ in the first exponent the *error term of equation* (2.1).

Proof of Lemma 2.1. We prove this lower bound by constructing an adequate number of generalized split graphs. Choose n_0 sufficiently large such that (2.1) holds for $m = \frac{1}{2}\binom{n}{2}$ and error term γm. Observe that it suffices to prove the lemma for $c \leqslant \frac{1}{2}$, since the complement of a split graph with density c is a split graph with density $(1-c)$.

We distinguish two cases. First, assume $c \leqslant \frac{1}{4}$. To obtain a lower bound for $|\mathcal{S}(n,c)|$ in this case, we simply count bipartite graphs with density c and with colour classes of size $n/2$. There are at least

$$\binom{\frac{n^2}{4}}{c\binom{n}{2}} \geqslant \binom{\frac{1}{2}\binom{n}{2}}{c\binom{n}{2}} \geqslant 2^{\frac{1}{2}\binom{n}{2}H(2c)-\gamma\binom{n}{2}}$$

such graphs.

Now assume that $\frac{1}{4} < c \leqslant \frac{1}{2}$. In this case we construct suitable k-partite graphs. For this purpose choose k such that

$$x := c\binom{n}{2} - (k-2)(\tfrac{n}{2}-k+2) - \binom{k-2}{2} \in \left[\frac{n^2}{8}-n, \frac{n^2}{8}+n\right]. \tag{2.2}$$

Indeed such a k exists since x monotonically decreases from $c\binom{n}{2}$ for $k=2$ to $c\binom{n}{2} - (n^2-2n-8)/8 \leqslant (n^2+8)/8$ for $k = n/2$ in steps of size at most $n/2$.

Now, construct k (independent) vertex sets $V_1, \ldots, V_k$ with $|V_1| = \frac{n}{2}$, $|V_2| = \frac{n}{2} - k + 2$ and $|V_i| = 1$ for $i = \{3, \ldots, k\}$ and insert all edges between V_i and V_j with $i, j \in [k] \setminus \{1\}$, $i \neq j$. Call the resulting graph G_0. By (2.2) we obtain a generalized split graph with density c from G_0, if we insert x edges between V_1 and $V_2 \cup \cdots \cup V_k$. Since this can be done in at least

$$\binom{\frac{n^2}{4}}{\frac{n^2}{8} - n} \geqslant 2^{\frac{1}{2}\binom{n}{2} - \gamma\binom{n}{2}}$$

ways, we obtain at least $2^{\frac{1}{2}\binom{n}{2}-\gamma\binom{n}{2}}$ generalized split graphs with exactly $c\binom{n}{2}$ edges. □

2.2. Regularity

In order to prove the upper bound from Theorem 1.2, *i.e.*,

$$|\mathcal{C}(n,c)| \leqslant 2^{h(c)\binom{n}{2}+\gamma\binom{n}{2}},$$

we will analyse the structure of graphs in $\mathcal{C}(n,c)$ by applying a variant of the regularity lemma suitable for our purposes.

Let $G=(V,E)$ be a graph. For disjoint non-empty vertex sets $A, B \subseteq V$, the *density* $d(A,B) := e(A,B)/(|A||B|)$ of the pair (A,B) is the number of edges that run between A and B divided by the number of possible edges between A and B. In the following let $\varepsilon, d \in [0,1]$. The pair (A,B) is *ε-regular* if, for all $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \geqslant \varepsilon|A|$ and $|B'| \geqslant \varepsilon|B|$, it is true that $|d(A,B) - d(A',B')| \leqslant \varepsilon$. An ε-regular pair (A,B) is called *(ε,d)-regular* if it has density at least d.

A partition $V_0\dot\cup V_1\dot\cup\cdots\dot\cup V_k$ of V is an *equipartition* if $|V_i| = |V_j|$ for all $i,j \in [k]$. An (ε,d)-*regular partition* of G with *reduced graph* $R = (V_R, E_R)$ is an equipartition $V_0\dot\cup V_1\dot\cup\cdots\dot\cup V_k$ of V with $|V_0| \leqslant \varepsilon|V|$, and $V_R = [k]$ such that (V_i, V_j) is an (ε,d)-regular pair in G if and only if $\{i,j\} \in E_R$. In this case we also call R an (ε,d)-reduced graph of G. An $(\varepsilon,0)$-regular partition $V_0\dot\cup V_1\dot\cup\cdots\dot\cup V_k$ which has at most $\varepsilon\binom{k}{2}$ pairs that are not ε-regular is also called an *ε-regular partition*. The partition classes V_i with $i \in [k]$ are called *clusters* of G and V_0 is the *exceptional set*.

With this terminology at hand we can state the celebrated regularity lemma of Szemerédi.

Lemma 2.2 (regularity lemma [21]). *For all $\varepsilon > 0$ and k_0 there is a k_1 such that every graph $G=(V,E)$ on $n \geqslant k_1$ vertices has an ε-regular partition $V = V_0\dot\cup V_1\dot\cup\cdots\dot\cup V_k$ with $k_0 \leqslant k \leqslant k_1$.* □

The strength of this lemma becomes apparent when it is complemented with corresponding embedding lemmas, such as the following (see, *e.g.*, [15]). A *homomorphism* from a graph $H = (V_H, E_H)$ to a graph $R = (V_R, E_R)$ is an edge-preserving mapping from V_H to V_R.

Lemma 2.3 (embedding lemma). *For every $d > 0$ and every integer k there exists $\varepsilon > 0$ with the following property. Let H be a graph on k vertices $v_1, \ldots, v_k$. Let G be a graph. Let $V_1, \ldots, V_k$ be clusters of an (ε,d)-regular partition of G with reduced graph $R = ([k], E_R)$. If there is a homomorphism from H to R, then H is a subgraph of G.* □

2.3. Regular partitions and counting

As a warm-up (and for the sake of completeness) we consider the problem of counting graphs of a fixed density without a given (not necessarily induced) subgraph F, and prove (1.1). For this purpose we mimic the proof given by Erdős, Frankl and Rödl in [9] for the corresponding problem without fixed density.

Proof of (1.1). Let F be a graph with $\chi(F) = r$ and $c \in (0, \frac{r-2}{r-1})$. Let $\gamma > 0$ be given. For large enough n the lower bound

$$|\mathcal{F}orb_n(F,c)| \geqslant 2^{\frac{r-2}{r-1}H(\frac{r-1}{r-2}c)\binom{n}{2}-\gamma\binom{n}{2}}$$

can easily be obtained by counting subgraphs with $c\binom{n}{2}$ edges of the complete $(r-1)$-partite graph with $n/(r-1)$ vertices in each part, and by applying (2.1).

It remains to show the upper bound

$$|\mathcal{F}orb_n(F,c)| \leqslant 2^{\frac{r-2}{r-1}H(\frac{r-1}{r-2}c)\binom{n}{2}+\gamma\binom{n}{2}}$$

for n sufficiently large. We choose d such that

$$0 < d \leqslant \min\left\{\frac{1}{16}, \frac{1}{4}\gamma^2, \left(-2\frac{r-1}{r-2}\log\left(1-c\frac{r-1}{r-2}\right)\right)^{-2}\right\}.$$

Since the entropy function is concave, we have for each $\hat{d} \in [0,d]$ that

$$\begin{aligned} H\big((c-2\hat{d})\tfrac{r-1}{r-2}\big) &\leqslant H(c\tfrac{r-1}{r-2}) - 2\hat{d}\tfrac{r-1}{r-2}H'(c\tfrac{r-1}{r-2}) \\ &\leqslant H(c\tfrac{r-1}{r-2}) - 2\hat{d}\tfrac{r-1}{r-2}\log(1-c\tfrac{r-1}{r-2}) \leqslant H(c\tfrac{r-1}{r-2}) + \sqrt{\hat{d}}, \end{aligned} \tag{2.3}$$

which we shall use later. Next, let ε be the constant returned from Lemma 2.3 for input d and with k replaced by r. Set $k_0 = \lceil 10/d \rceil$ and let k_1 be the constant returned by Lemma 2.2 for input k_0 and ε. Further let $n \geqslant k_1$.

Now we use the regularity lemma, Lemma 2.2, with parameters ε, k_0 for each graph G in $\mathcal{F}orb_n(F,c)$. For each such application the regularity lemma produces an ε-regular partition with at most k_1 clusters, for which we can construct the corresponding (ε,d)-reduced graph R. Since k_1 is finite there is only a finite number of different reduced graphs R resulting from these applications of the regularity lemma. Hence we can partition $\mathcal{F}orb_n(F,c)$ into a finite number of classes $\mathcal{R}(R,\varepsilon,d,n,F,c)$ of graphs with (ε,d)-reduced graph R. Accordingly, it suffices to show that for each R we have

$$|\mathcal{R}(R,\varepsilon,d,n,F,c)| \leqslant 2^{\frac{r-2}{r-1}H(\frac{r-1}{r-2}c)\binom{n}{2}+\gamma\binom{n}{2}}. \tag{2.4}$$

Let $R = (V_R, E_R)$ be any graph such that $\mathcal{R}(R,\varepsilon,d,n,F,c)$ is non-empty, let $k = |V_R|$, let $G \in \mathcal{R}(R,\varepsilon,d,n,F,c)$ and let P be an ε-regular partition of G corresponding to R. By the choice of k_0 at most $k\binom{n/k}{2} \leqslant \frac{d}{2}\binom{n}{2}$ edges of G are inside clusters of P, at most $d\binom{n}{2}$ edges of G are in regular pairs of P with density less than d, and at most $2\varepsilon n^2 \leqslant \frac{d}{2}\binom{n}{2}$ edges of G are in irregular pairs of P or have a vertex in the exceptional set. We conclude that at least $(c-2d)\binom{n}{2}$ edges of G lie in (ε,d)-regular pairs of P. In addition, by the choice of ε and since F has chromatic number r, Lemma 2.3 implies that $K_r \not\subseteq R$. It follows from Turán's theorem that $|E_R| \leqslant \frac{r-2}{r-1}\binom{k}{2}$. Summarizing, we can bound the number of graphs in $\mathcal{R}(R,\varepsilon,d,n,F,c)$ by bounding the number of ways to distribute at least $(c-2d)\binom{n}{2}$ edges to at most $\frac{r-2}{r-1}\binom{k}{2}$ regular pairs (corresponding to edges of R) with clusters of size at most n/k, and distributing at most $2d\binom{n}{2}$ edges arbitrarily. By the choice of n, the first of these two factors can be bounded by

$$\max_{0\leqslant\hat{d}\leqslant d}\binom{\frac{r-2}{r-1}\binom{n}{2}}{(c-2\hat{d})\binom{n}{2}} \overset{(2.1)}{\leqslant} \max_{0\leqslant\hat{d}\leqslant d} 2^{\frac{r-2}{r-1}\binom{n}{2}H(\frac{r-1}{r-2}(c-2\hat{d}))}$$
$$\overset{(2.3)}{\leqslant} \max_{0\leqslant\hat{d}\leqslant d} 2^{\frac{r-2}{r-1}\binom{n}{2}H(c\frac{r-1}{r-2})+\sqrt{\hat{d}}\binom{n}{2}}$$
$$\leqslant 2^{\frac{r-2}{r-1}\binom{n}{2}H(c\frac{r-1}{r-2})+\sqrt{d}\binom{n}{2}},$$

and the second by $\leqslant 2^{2d\binom{n}{2}}$. Since $2d+\sqrt{d}\leqslant\gamma$ this implies (2.4) as desired. □

2.4. Embedding induced subgraphs

In the last section we showed how the regularity lemma and a corresponding embedding lemma can be used to count graphs with forbidden subgraphs. In this section we provide the tools that will allow us to adapt this strategy to the setting of forbidden induced subgraphs.

We remark that the concepts and ideas presented in this section are not new. They have been used for various similar applications, *e.g.*, by Bollobás and Thomason [7] and Loebl, Reed, Scott, Thomason and Thomassé [16], as well as for different applications such as property testing, *e.g.*, by Alon, Fischer, Krivelevich and Szegedy [3] and Alon and Shapira [4].

We start with an embedding lemma for induced subgraphs, which allows us to find an induced copy of a graph F in a graph G with reduced graph R if F is an induced subgraph of R (see, *e.g.*, [3]).

Lemma 2.4 (injective embedding lemma for induced subgraphs). *For every $d>0$ and every integer k there exists $\varepsilon>0$ such that for all $f\leqslant k$ the following holds. Let $V_1,\dots,V_f$ be clusters of an ε-regular partition of a graph G such that for all $1\leqslant i<j\leqslant f$ the pair (V_i,V_j) is ε-regular. Let $F=(V_F,E_F)$ be a graph on f vertices and let $g: V_F\to[f]$ be an injective mapping from F to the clusters of G such that for all $1\leqslant i<j\leqslant f$ we have $d(V_i,V_j)\geqslant d$ if $\{g^{-1}(i),g^{-1}(j)\}\in E(F)$ and $d(V_i,V_j)\leqslant 1-d$ if $\{g^{-1}(i),g^{-1}(j)\}\notin E(F)$. Then G contains an induced copy of F.* □

In contrast to Lemma 2.3, this lemma allows us to embed only one vertex per cluster of G. Our goal in the following will be to describe an embedding lemma for induced subgraphs which allows us to embed more than one vertex per cluster. Observe first that for this purpose we must have some control over the existence of edges respectively non-edges *inside* clusters of a regular partition of G. This can be achieved by applying the following lemma to each of these clusters. It is not difficult to infer this lemma from the regularity lemma (Lemma 2.2) by applying Turán's theorem and Ramsey's theorem (see, *e.g.*, [3]).

We use the following definition. A (μ,ε,k)-*subpartition* of a graph $G=(V,E)$ is a family of pairwise disjoint vertex sets $W_1,\dots,W_k\subseteq V$ with $|W_i|\geqslant\mu|V|$ for all $i\in[k]$ such that each pair (W_i,W_j) with $\{i,j\}\in\binom{k}{2}$ is ε-regular. A (μ,ε,k)-subpartition $W_1,\dots,W_k$ of G is *dense* if $d(W_i,W_j)\geqslant\frac{1}{2}$ for all $\{i,j\}\in\binom{k}{2}$, and *sparse* if $d(W_i,W_j)<\frac{1}{2}$ for all $\{i,j\}\in\binom{k}{2}$.

Lemma 2.5. *For every k and ε there exists $\mu>0$ such that every graph $G=(V,E)$ with $n\geqslant\mu^{-1}$ vertices either has a sparse or a dense (μ,ε,k)-subpartition.* □

The idea for the embedding lemma for induced subgraphs F of G is now as follows. We first find a regular partition of G. By Lemma 2.4, if a regular pair (V_i, V_j) in this partition is very dense then we can embed edges of F into (V_i, V_j), if it is very sparse then we can embed non-edges of F, and if its density is neither very small nor very big then we can embed both edges and non-edges of F. Moreover, Lemma 2.5 asserts that each cluster either has a sparse or a dense subpartition. In the first case we can embed non-edges inside this cluster, in the second case we can embed edges.

This motivates us to tag the reduced graphs with some additional information. For this purpose we colour an edge of the reduced graph white if the corresponding regular pair is sparse, grey if it is of medium density, and black if it is dense. Moreover, we colour a cluster white if it has a sparse subpartition and black otherwise. We call a cluster graph that is coloured in this way a *type*. The following definitions make this precise.

Definition 2.6 (coloured graph, type). A *coloured graph* R is a triple (V_R, E_R, σ) such that (V_R, E_R) is a graph and $\sigma : V_R \cup E_R \to \{0, \frac{1}{2}, 1\}$ is a colouring of the vertices and the edges of this graph where $\sigma(V_R) \subseteq \{0, 1\}$. Vertices and edges with colour 0, $\frac{1}{2}$, and 1 are also called *white*, *grey*, and *black*, respectively.

Let $G = (V, E)$ be a graph and let $V = V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ be an ε-regular partition of G with reduced graph $([k], E_R)$. The $(\varepsilon, \varepsilon', d, k')$*-type* R corresponding to the partition $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ is the coloured graph $R = ([k], E_R, \sigma)$ with colouring

$$\sigma(\{i, j\}) = \begin{cases} 0 & \text{if } d(V_i, V_j) < d, \\ 1 & \text{if } d(V_i, V_j) > 1 - d, \\ \frac{1}{2} & \text{otherwise,} \end{cases}$$

for all $\{i, j\} \in E_R$, and

$$\sigma(i) = \begin{cases} 0 & \text{if } G[V_i] \text{ has a sparse } (\mu, \varepsilon', k')\text{-subpartition,} \\ 1 & \text{if } G[V_i] \text{ has a dense } (\mu, \varepsilon', k')\text{-subpartition,} \end{cases}$$

for all $i \in [k]$, where μ is the constant from Lemma 2.5 for input k' and ε'. In this case we also simply say that G has $(\varepsilon, \varepsilon', d, k')$-type R.

By the discussion above a combination of the regularity lemma, Lemma 2.2, and Lemma 2.5 gives the following.

Lemma 2.7 (type lemma). *For every* $\varepsilon, \varepsilon' \in (0, \frac{1}{2})$ *and for all integers* k', k_0, *there are integers* k_1 *and* n_0 *such that, for every* $d > 0$, *every graph* G *on at least* n_0 *vertices has an* $(\varepsilon, \varepsilon', d, k')$*-type* $R = ([k], E_R, \sigma)$ *with* $k_0 \leqslant k \leqslant k_1$ *and with at most* εk^2 *non-edges.*

Proof. Given ε, ε' and k', k_0, we let k_1 be the constant returned from Lemma 2.2 for input ε and k_0, and let μ be the constant returned from Lemma 2.5 for input ε' and k'. Set $n_0 := 2\mu^{-1}k_1$.

Now let d be given and let G be a graph on at least n_0 vertices. By Lemma 2.2 the graph G has an ε-regular partition $V = V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ such that $k_0 \leqslant k \leqslant k_1$. By definition at most $\varepsilon\binom{k}{2} \leqslant \varepsilon k^2$ pairs (V_i, V_j) are not ε-regular. Let $R' = ([k], E_R)$ be the ε-reduced graph of $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$.

It follows that R' has at most εk^2 non-edges. Let $i \in [k]$. Since $|V_i| \geqslant (1-\varepsilon)n/k \geqslant n/(2k_1) \geqslant \mu^{-1}$, we can apply Lemma 2.5 and conclude that V_i either has a sparse or a dense (μ, ε', k')-subpartition. Accordingly we obtain an $(\varepsilon, \varepsilon', d, k')$-type $R = ([k], E_R, \sigma)$ for G by colouring the edges and vertices of R' as specified in Definition 2.6. □

For formulating our embedding lemma we need one last preparation. We generalize the concept of a graph homomorphism to the setting of coloured graphs.

Definition 2.8 (coloured homomorphism). Let $F = (V_F, E_F)$ be a graph and let $R = (V_R, E_R, \sigma)$ be a coloured graph. A *coloured homomorphism* from F to R is a mapping $h\colon V_F \to V_R$ with the following properties.

(a) If $u, v \in V_F$ and $h(u) \neq h(v)$ then $\{h(u), h(v)\} \in E_R$.
(b) If $\{u, v\} \in E_F$ then $h(u) = h(v)$ and $\sigma(h(u)) = 1$, or $h(u) \neq h(v)$ and $\sigma\big(\{h(u), h(v)\}\big) \in \{\frac{1}{2}, 1\}$.
(c) If $\{u, v\} \notin E_F$ then $h(u) = h(v)$ and $\sigma(h(u)) = 0$, or $h(u) \neq h(v)$ and $\sigma\big(\{h(u), h(v)\}\big) \in \{0, \frac{1}{2}\}$.

If there is a coloured homomorphism from F to R we also write $F \xrightarrow{\sigma} R$.

The following embedding lemma states that a graph F is an induced subgraph of a graph G with type R if there is a coloured homomorphism from F to R. This lemma is inherent in [4], for example. For completeness we provide its proof below.

Lemma 2.9 (embedding lemma for induced graphs). *For every pair of integers k, k' and for every $d \in (0,1)$ there are $\varepsilon, \varepsilon' > 0$ such that the following holds. Let $f \leqslant k'$ and G be a graph on n vertices with $(\varepsilon, \varepsilon', d, k')$-type R on k vertices. Let F be an f-vertex graph such that there is a coloured homomorphism from F to R. Then F is an induced subgraph of G.*

Proof. Let $k, k' \in \mathbb{N}$ and $d \in (0,1)$ be given. Let ε' be given by Lemma 2.4 for input $\frac{d}{2}$ and k'. Let μ be the constant from Lemma 2.5 for input k' and ε'. Set $\varepsilon := \min\{d/2, \mu\varepsilon'\}$.

Let G, $R = ([k], E_R, \sigma)$ and F be as required, let $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ be an ε-regular partition of G corresponding to R and let $h\colon F \xrightarrow{\sigma} R$ be a coloured homomorphism from F to R. For each $i \in [k]$ we have by definition that if $\sigma(V_i) = 0$ then V_i has a sparse (μ, ε', k')-subpartition $W_{i,1}, \ldots, W_{i,k'}$, and if $\sigma(V_i) = 1$ then V_i has a dense (μ, ε', k')-subpartition $W_{i,1}, \ldots, W_{i,k'}$.

Observe that $\bigcup_{i \in [k], j \in [k']} W_{i,j}$ has the following properties, since $\varepsilon' \geqslant \varepsilon/\mu$ and $\varepsilon \leqslant d/2$. If $\{i, i'\} \in E_R$, then for all $j, j' \in [k']$ the pair

$$(W_{i,j}, W_{i',j'}) \text{ is } \varepsilon'\text{-regular}, \tag{2.5}$$

and has density

$$d(W_{i,j}, W_{i',j'}) \in \begin{cases} [0, 2d) & \text{if } \sigma(\{i, i'\}) = 0, \\ (1-2d, 1] & \text{if } \sigma(\{i, i'\}) = 1, \\ (\frac{d}{2}, 1-\frac{d}{2}) & \text{if } \sigma(\{i, i'\}) = \frac{1}{2}. \end{cases} \tag{2.6}$$

Moreover, for all $i \in [k]$ and all $j, j' \in [k']$ with $j \neq j'$ the pair

$$(W_{i,j}, W_{i,j'}) \text{ is } \varepsilon'\text{-regular}, \tag{2.7}$$

and has density

$$d(W_{i,j},W_{i,j'}) \in \begin{cases} [0,\frac{1}{2}) & \text{if } \sigma(i)=0, \\ [\frac{1}{2},1] & \text{if } \sigma(i)=1. \end{cases} \tag{2.8}$$

Now, we define an injective mapping $g: V(F) \to [k] \times [k']$ as follows. For $i \in [k]$ let $F_i := \{x \in V(F) : h(x)=i\}$ and name the vertices in F_i arbitrarily by $\{x_{i,1},\ldots,x_{i,f_i}\}=F_i$. Set $g(x_{i,j}) := (i,j)$ for all $j \in [f_i]$. This is well-defined since $|F_i| = f_i \leqslant f \leqslant k'$. Let $I \subseteq [k]$ with $|I| \leqslant f$ be the set of indices $i \in [k]$ such that $F_i \neq \emptyset$.

We claim that G, $\bigcup_{i\in I, j\in[f_i]} W_{i,j}$, F, and g satisfy the conditions of Lemma 2.4 with parameters $d/2$, k', ε', and f. Indeed, by (2.7) each cluster pair $(W_{i,j},W_{i,j'})$ with $i \in I$ and $j,j' \in [k']$, $j \neq j'$ is ε'-regular. Moreover, for each $i,i' \in I$ with $i \neq i'$ we have that there are $x \in F_i$ and $y \in F_{i'}$. By the definition of a coloured homomorphism (Definition 2.8) we have that $\{V_i,V_{i'}\} = \{h(x),h(y)\} \in E_R$. Hence (2.5) implies that $(W_{i,j},W_{i',j'})$ is ε'-regular too. It remains to show that if x,y are two vertices in $V(F)$ and $(i,j)=g(x)$ and $(i',j')=g(y)$, then $d(W_{i,j},W_{i',j'}) \geqslant d/2$ if $\{x,y\} \in E(F)$ and $d(W_{i,j},W_{i',j'}) \leqslant 1-d/2$ otherwise. To see this, assume first that $\{x,y\} \in E(F)$. Then, by the definition of a coloured homomorphism, either $h(x)=h(y)$ and $\sigma(h(x))=1$, which implies $d(W_{i,j},W_{i',j'}) \geqslant \frac{1}{2} \geqslant d/2$ by (2.8). Or $h(x) \neq h(y)$ and $\sigma(\{h(x),h(y)\}) \geqslant \frac{1}{2}$, and hence we have $d(W_{i,j},W_{i',j'}) \geqslant d/2$ by (2.6). If $\{x,y\} \notin E(F)$, on the other hand, then either $h(x)=h(y)$ and $\sigma(h(x))=0$, and so $d(W_{i,j},W_{i',j'}) \leqslant \frac{1}{2} \leqslant 1-d/2$ by (2.8). Or $h(x) \neq h(y)$ and $\sigma(\{h(x),h(y)\}) \leqslant \frac{1}{2}$ and thus $d(W_{i,j},W_{i',j'}) \leqslant 1-d/2$ by (2.6).

It follows that we can indeed apply Lemma 2.4 and conclude that F is an induced subgraph of G as desired. □

2.5. The upper bound of Theorem 1.2

Now we are ready to prove the upper bound of Theorem 1.2, that is, we establish the following lemma.

Lemma 2.10. *For all $c,\gamma \in (0,1)$, there is an n_0 such that for all $n \geqslant n_0$ we have*

$$|\mathcal{C}(n,c)| \leqslant 2^{h(c)\binom{n}{2}+\gamma\binom{n}{2}}.$$

The idea of the proof of Lemma 2.10 is as follows. We proceed in three steps. Firstly, as in the proof of (1.1) in Section 2.3, we start by applying the regularity lemma to all graphs in $\mathcal{C}(n,c)$. For each of the regular partitions obtained in this way there is a corresponding type, and in total we only get a constant number K of different types. Secondly, we continue with a structural analysis of the possible types R for graphs from $\mathcal{C}(n,c)$ and infer from Lemma 2.9 that R cannot contain a triangle all of whose edges are grey (see Lemma 2.11). Thirdly, we prove that a coloured graph without such a grey triangle can only serve as a type for at most $\mathrm{UB}(n)$ graphs on n vertices (see Lemma 2.12). Multiplying $\mathrm{UB}(n)$ with K then gives the desired bound.

We start with the second step.

Lemma 2.11. *For every integer $k' \geqslant 5$ and every $d > 0$, there exist $\varepsilon_{L2.11}, \varepsilon'_{L2.11} > 0$ such that for every $0 < \varepsilon \leqslant \varepsilon_{L2.11}$ and every $0 < \varepsilon' \leqslant \varepsilon'_{L2.11}$ the following is true. If G is a graph whose $(\varepsilon,\varepsilon',d,k')$-type R contains three grey edges forming a triangle, then G has an induced C_5.*

Proof. Given $d > 0$ and $k' \geqslant 5$, set $k := 3$ and let $\varepsilon_{\mathrm{L2.11}}, \varepsilon'_{\mathrm{L2.11}} > 0$ be the constants given by Lemma 2.9 for d, k, and k'. Let positive constants $\varepsilon \leqslant \varepsilon_{\mathrm{L2.11}}$ and $\varepsilon' \leqslant \varepsilon'_{\mathrm{L2.11}}$ be given. Let G be a graph with $(\varepsilon, \varepsilon', d, k')$-type R such that R contains a triangle T with three grey edges.

By Lemma 2.9 the graph G contains an induced C_5 if there exists a coloured homomorphism from C_5 to T, in which case we are done. We claim that such a coloured homomorphism h always exists (regardless of the colours of the vertices of T). Indeed, if T has at least two black vertices V_1, V_2 then we can construct h by mapping a pair of adjacent vertices of C_5 to V_1, a disjoint pair of adjacent vertices of C_5 to V_2, and the remaining vertex of C_5 to the remaining vertex of T. If T has at least two white vertices V_1, V_2, on the other hand, then we can construct h by mapping one pair of non-adjacent vertices of C_5 to V_1, a disjoint pair of non-adjacent vertices of C_5 to V_2 and the remaining vertex of C_5 to the remaining vertex of T. □

Next, we show an upper bound on the number of graphs on n vertices with a fixed type R, where R does not contain a triangle with three grey edges. We use the following definition:

$$\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c) := \{G \in \mathcal{G}(n, c) : G \text{ has } (\varepsilon, \varepsilon', d, k')\text{-type } R\}. \tag{2.9}$$

We stress that $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ and $\mathcal{R}(R', \varepsilon, \varepsilon', d, k', n, c)$ may have non-empty intersection for $R \neq R'$.

Lemma 2.12. *For every c with $0 < c \leqslant \frac{1}{2}$, and every $\gamma > 0$, there exist $\varepsilon_{L2.12}, d_0 > 0$ and integers $n_{L2.12}$, k_0 such that for all positive $d \leqslant d_0$, $\varepsilon \leqslant \varepsilon_{L2.12}$, ε', and all integers $k \geqslant k_0$, k', $n \geqslant \max\{k, n_{L2.12}\}$, the following holds. If R is a coloured graph of order k which has at most εk^2 non-edges and does not contain a triangle with three grey edges, then*

$$|\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)| \leqslant 2^{h(c)\binom{n}{2} + \gamma\binom{n}{2}}.$$

Proof. Let c, γ be given. Choose $\varepsilon_{\mathrm{L2.12}}$, d_0, k_0 such that

$$\max\left\{4\varepsilon_{\mathrm{L2.12}}, H(d_0), \frac{1}{k_0}\right\} \leqslant \frac{\gamma}{5}.$$

Let $n_{\mathrm{L2.12}}$ be large enough to guarantee $\log(n_{\mathrm{L2.12}} + 1) \leqslant \frac{\gamma}{5}(n_{\mathrm{L2.12}} - 1)$. Let $\varepsilon \leqslant \varepsilon_{\mathrm{L2.12}}$, ε', $d \leqslant d_0$, $k \geqslant k_0$, k', $n \geqslant \max\{n_{\mathrm{L2.12}}, k\}$ be given.

Let $R = ([k], E_R, \sigma)$ be a coloured graph which has at most εk^2 non-edges and does not contain a triangle with three grey edges. We shall count the graphs in $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ by estimating the number of equipartitions $V_0 \dot\cup \cdots \dot\cup V_k$ of $[n]$, the number of choices for edges with one end in the exceptional set V_0 and edges in pairs (V_i, V_j) such that $\{i, j\} \notin E_R$, the number of choices for edges in clusters V_i such that i is white or black in R, and the number of choices for at most $c\binom{n}{2}$ edges in pairs (V_i, V_j) such that $\{i, j\}$ is a white, black, or grey edge of R.

The number of equipartitions $V_0 \dot\cup \cdots \dot\cup V_k$ of $[n]$ is bounded by

$$(k+1)^n = 2^{n\log(k+1)} \leqslant 2^{n\log(n+1)} \leqslant 2^{\frac{\gamma}{5}\binom{n}{2}}. \tag{2.10}$$

Let us now fix such an equipartition. There are at most εn^2 possible edges that have at least one end in V_0 and at most $\frac{\varepsilon}{2} n^2$ possible edges in pairs (V_i, V_j) such that $\{i, j\} \notin E_R$. Thus there are at most

$$2^{\frac{3}{2}\varepsilon n^2} \leqslant 2^{4\varepsilon\binom{n}{2}} \tag{2.11}$$

possible ways to distribute such edges. In addition, the number of ways to distribute edges in clusters V_i corresponding to white or black vertices of R is at most

$$2^{k\binom{n/k}{2}} \leqslant 2^{\frac{1}{k}\binom{n}{2}}. \tag{2.12}$$

By definition, white edges of an $(\varepsilon, \varepsilon', d, k')$-type correspond to pairs with density at most d and black edges correspond to pairs with density at least $(1-d)$. Hence, by the symmetry of the binomial coefficient the number of ways to distribute edges in pairs (V_i, V_j) such that $\{i, j\}$ is a white or a black edge of R is at most

$$\binom{\left(\frac{n}{k}\right)^2}{d\left(\frac{n}{k}\right)^2}^{\binom{k}{2}} \leqslant 2^{\binom{k}{2}\left(\frac{n}{k}\right)^2 H(d)} \leqslant 2^{H(d)\binom{n}{2}}. \tag{2.13}$$

For later reference we now sum up the estimates obtained so far. The product of (2.10)–(2.13) gives less than

$$2^{\left(\frac{\gamma}{5}+4\varepsilon+H(d)+\frac{1}{k}\right)\binom{n}{2}} \leqslant 2^{\frac{4}{5}\gamma\binom{n}{2}} \tag{2.14}$$

choices for the partition $V_0\dot\cup\cdots\dot\cup V_k$ and for the distribution of edges inside such a partition, except for the pairs (V_i, V_j) corresponding to grey edges of R.

It remains to take the grey edges E_g of R into account. By assumption, E_g does not contain a triangle. Hence, by Turán's theorem (see, *e.g.*, [23]) we have $|E_g| \leqslant \frac{k^2}{4}$. It follows that there are at most $\frac{k^2}{4}\left(\frac{n}{k}\right)^2 = \frac{n^2}{4}$ possible places for edges in E_g-pairs (V_i, V_j), *i.e.*, pairs such that $\{i, j\} \in E_g$. Hence the number N_g of possible ways to distribute at most $c\binom{n}{2}$ edges to E_g-pairs is at most $c\binom{n}{2}\binom{n^2/4}{c\binom{n}{2}}$. If $c < \frac{1}{4}$, then this gives

$$N_g \leqslant 2^{\frac{1}{2}\binom{n}{2}H(2c)+\gamma\binom{n}{2}}, \tag{2.15}$$

and if $\frac{1}{4} \leqslant c \leqslant \frac{1}{2}$ then

$$N_g \leqslant 2^{\frac{n^2}{4}} \leqslant 2^{\frac{1}{2}\binom{n}{2}+\frac{\gamma}{5}\binom{n}{2}}. \tag{2.16}$$

Combining (2.15) and (2.16) and recalling the definition of $h(c)$ in (1.2) gives

$$N_g \leqslant 2^{h(c)\binom{n}{2}+\frac{\gamma}{5}\binom{n}{2}}. \tag{2.17}$$

Multiplying (2.14) and (2.17) gives the desired upper bound

$$|\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)| \leqslant 2^{h(c)\binom{n}{2}+\frac{\gamma}{5}\binom{n}{2}+\frac{4}{5}\gamma\binom{n}{2}} = 2^{h(c)\binom{n}{2}+\gamma\binom{n}{2}}. \qquad \square$$

With this we are in position to prove Lemma 2.10.

Proof of Lemma 2.10. Observe first that it suffices to prove the lemma for $c \leqslant \frac{1}{2}$, as the complement of a graph without induced C_5 is induced C_5-free and hence $|\mathcal{C}(n, c)| = |\mathcal{C}(n, 1-c)|$.

Now let $c \in (0, \frac{1}{2}]$ and $\gamma > 0$ be given. Set $k' = 5$. Lemma 2.12 with input c and $\gamma/2$ provides constants $\varepsilon_{\text{L2.12}}$, d_0, $n_{\text{L2.12}}$, k_0. Set $d = d_0$. From Lemma 2.11 with input d we obtain constants

$\varepsilon_{L2.11}$ and $\varepsilon'_{L2.11}$. Set $\varepsilon := \min\{\varepsilon_{L2.12}, \varepsilon_{L2.11}\}$ and $\varepsilon' := \varepsilon'_{L2.11}$. The type lemma, Lemma 2.7, finally, with input ε, ε', and k_0, k' gives constants k_1 and $n_{L2.7}$. We set $n_0 = \max\{n_{L2.7}, n_{L2.12}, \frac{3}{\sqrt{\gamma}}k_1\}$.

Now, for each graph $G \in \mathcal{C}(n,c)$ we apply the type lemma, Lemma 2.7, with parameters ε, ε', k_0, k' and d and obtain an $(\varepsilon, \varepsilon', d, k')$-type R of G on $k \leqslant k_1$ vertices and with at most εk^2 non-edges. Let $\tilde{\mathcal{R}}$ be the set of types obtained from these applications of Lemma 2.7. It follows that $|\tilde{\mathcal{R}}| \leqslant 4^{\binom{k_1}{2}} 2^{k_1} \leqslant 2^{k_1^2}$. By Lemma 2.11 applied with d, ε, and ε', no coloured graph in $\tilde{\mathcal{R}}$ contains a triangle with three grey edges. Hence, by Lemma 2.12 applied with c, $\gamma/2$, ε, ε' and d we have $|\mathcal{R}(R, \varepsilon, \varepsilon', d, 5, n, c)| \leqslant 2^{h(c)\binom{n}{2} + \frac{1}{2}\gamma\binom{n}{2}}$. Since, by Lemma 2.7,

$$\mathcal{C}(n,c) \subseteq \bigcup_{R \in \tilde{\mathcal{R}}} \mathcal{R}(R, \varepsilon, \varepsilon', d, 5, c, n).$$

We conclude from the choice of n_0 that

$$|\mathcal{C}(n,c)| \leqslant 2^{k_1^2} \cdot 2^{h(c)\binom{n}{2} + \frac{1}{2}\gamma\binom{n}{2}} \leqslant 2^{h(c)\binom{n}{2} + \gamma\binom{n}{2}}. \qquad \square$$

3. The proof of Theorem 1.3

Our proof of Theorem 1.3 consists of the following steps. We start, as in the proof of Theorem 1.2, by constructing for each graph G in $\mathcal{F}orb^*_{n,\eta}(C_5, c)$ a type R of size independent of n with the help of the type lemma, Lemma 2.7. Next, we consider each cluster V_i of a partition of $V(G)$ corresponding to R separately. We shall show that the fact that G does not contain homogeneous sets of size ηn implies that $G[V_i]$ has many vertex-disjoint induced copies of P_3, the path on three vertices, or many vertex-disjoint induced copies of the anti-path $\overline{P}_3$, the complement of P_3 (see Lemma 3.1). Many induced copies of P_3 or $\overline{P}_3$ in two clusters V_i and V_j, however, limit the number of possibilities for inserting edges between V_i and V_j without inducing a C_5 (see Lemma 3.2). Combining this with the proof strategy from Theorem 1.2 will give us an upper bound for the number of graphs from $\mathcal{F}orb^*_{n,\eta}(C_5, c)$ with type R (see Lemma 3.3). Finally, comparing this upper bound with the lower bound on $|\mathcal{F}orb^*_n(C_5, c)|$ from Theorem 1.2 will lead to the desired result.

We start by proving that graphs without big homogeneous sets contain many vertex-disjoint induced P_3 or $\overline{P}_3$.

Lemma 3.1. *Let G be a graph of order n with $\hom(G) \leqslant n/6$. Then one of the following is true:*

(i) *G contains $n/6$ vertex-disjoint induced copies of P_3, or*
(ii) *G contains $n/6$ vertex-disjoint induced copies of $\overline{P}_3$.*

Proof. Let G be an n-vertex graph with $\hom(G) \leqslant n/6$. Select a maximal set of disjoint copies of P_3. If this set consists of less than $n/6$ paths then there is a subgraph $G' \subseteq G$ with $v(G') = n/2$ that has no induced P_3 and thus is a vertex-disjoint union of cliques $Q_1, \ldots, Q_\ell$. We claim that in G' we can find $n/6$ vertex-disjoint induced $\overline{P}_3$, which proves the lemma.

Indeed, since $\hom(G) \leqslant n/6$ we have $\ell \leqslant n/6$, and for each $i \in [\ell]$ we have $q_i := |Q_i| \leqslant n/6$. This implies

$$\sum_{i\in[\ell]} \lfloor q_i/2 \rfloor \geqslant \frac{1}{2}\left(\frac{n}{2} - \frac{n}{6}\right) = n/6.$$

It follows that we can find a set of $n/6$ vertex-disjoint edges $E = \{e_1, \dots, e_{n/6}\}$ in these cliques in the following way. We first choose as many vertex-disjoint edges in Q_1 as possible, then in Q_2, and so on, until we have chosen $n/6$ edges in total. Let Q_k be the last clique used in this process. Then, for each clique Q_i with $i < k$ at most one vertex was unused in this process, and in Q_k possibly several vertices were unused. Let X be the set of all these unused vertices together with all vertices from $\bigcup_{k<i\leqslant\ell} Q_i$. Clearly $|X| = n/6$.

We consider the auxiliary bipartite graph $B = (X \cup E, E_B)$ with $\{x, e\} \in E_B$ for $x \in X$ and $e \in E$ if and only if x and e do not lie in the same clique of G'. We verify Hall's condition for B. So let $Y \subseteq X$. If $Y \not\subseteq Q_i$ for all i we have $|N(Y)| = |E| = |X| \geqslant |Y|$. Otherwise, if $Y \subseteq Q_i$ for some i, then $|N(Y)| \geqslant |E| - (|Q_i| - |Y|)/2 \geqslant n/6 - n/12 + |Y|/2 \geqslant |Y|$ since $|Y| \leqslant n/6$. It follows that B has a perfect matching, which means that there are $n/6$ vertex-disjoint induced $\overline{P}_3$ in G', as claimed. $\square$

Now suppose we are given a graph G with vertex set $V_1 \dot\cup V_2$ and no edges between V_1 and V_2. Further, let H_1 and H_2 be such that for $i \in [2]$ the graph H_i induces a copy of P_3 or $\overline{P}_3$ in $G[V_i]$. Observe that, no matter which combination of P_3 or $\overline{P}_3$ we choose, we can create an induced C_5 in G by adding appropriate edges between H_1 and H_2. Since we are interested in graphs without induced C_5, this motivates us to call (H_1, H_2) a *dangerous pair* of (V_1, V_2).

Our next goal is to use these dangerous pairs in order to derive an upper bound on the number of possibilities for inserting edges between V_1 and V_2 without creating an induced copy of C_5 if we know that (V_1, V_2) contains many dangerous pairs. In order to quantify this upper bound, in Lemma 3.2 we use the following technical definition. We define $R(c) = c^4(1-c)^4$ and the function $r : (0,1) \to \mathbb{R}^+$ with

$$r(c) = \frac{1}{72}\begin{cases} R(2c) & \text{if } c < \frac{1}{4}, \\ (1/4)^4 & \text{if } c \in [\frac{1}{4}, \frac{3}{4}], \\ R(2c-1) & \text{otherwise.} \end{cases} \tag{3.1}$$

Recall in addition the definition of the function $h(c)$ from (1.2).

Lemma 3.2. *For every $0 < c_0 \leqslant \frac{1}{2}$ there is an n_0 such that, for all c with $c_0 \leqslant 2c \leqslant 1 - c_0$ and $n \geqslant n_0$, the following holds. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two n-vertex graphs, each of which contains $n/6$ vertex-disjoint induced copies of P_3 or $n/6$ vertex-disjoint induced copies of $\overline{P}_3$. Let $G = (V_1 \dot\cup V_2, E)$ be the disjoint union of G_1 and G_2. Then there are at most*

$$2^{2n^2(h(c)-r(c))}$$

ways to add exactly $2cn^2$ edges to G that run between V_1 and V_2 without inducing a C_5 in G.

We remark that in the proof of this lemma we are going to make use of the following probabilistic principle: we can *count* the number of elements in a finite set X which have some property P, by determining the *probability* that an element which is chosen from X uniformly at random has property P.

Proof of Lemma. 3.2. Given $c_0 \in (0, \frac{1}{2}]$, let n_0 be sufficiently large that

$$n_0 e^{-2r(c_0/2)n_0^2} \leqslant 2^{-2r(c_0/2)n_0^2}. \tag{3.2}$$

Now let c be such that $c_0 \leqslant 2c \leqslant 1 - c_0$. Observe first that it suffices to prove the lemma for $2c \leqslant \frac{1}{2}$, since induced C_5-free graphs are self-complementary and P_3 is the complement of $\overline{P}_3$. Hence, we assume from now on that $2c \leqslant \frac{1}{2}$. Observe moreover that (3.2) remains valid if c_0 is replaced by c, since $r(c)$ is monotone increasing in $[0, \frac{1}{2}]$. Let G_1, G_2, and G be as required.

Our first goal is to estimate the probability P^* of inducing no C_5 in G when choosing uniformly at random exactly $2cn^2$ edges between V_1 and V_2. Instead of dealing with P^* directly, we consider the following binomial random graph $\mathcal{G}(V_1, V_2, p)$ with $p = 2c$: we start with G and add each edge between V_1 and V_2 independently with probability p.

Now, let A be the event that $\mathcal{G}(V_1, V_2, p)$ contains exactly $2cn^2$ edges between V_1 and V_2, and let B be the event that $\mathcal{G}(V_1, V_2, p)$ contains no induced C_5. Observe that each graph with $2cn^2$ edges between V_1 and V_2 is equally likely in $\mathcal{G}(V_1, V_2, p)$, and thus

$$P^* = \mathbb{P}[B|A] \leqslant \frac{\mathbb{P}[B]}{\mathbb{P}[A]}. \tag{3.3}$$

Hence it suffices to estimate $\mathbb{P}[A]$ and $\mathbb{P}[B]$.

We first bound $\mathbb{P}[B]$. By assumption there are at least $n^2/36$ dangerous pairs in (V_1, V_2). Now fix such a dangerous pair (H_1, H_2). The probability that (H_1, H_2) induces a C_5 in $\mathcal{G}(V_1, V_2, p)$ is at least $p^2(1-p)^4$ unless H_1 and H_2 are both $\overline{P}_3$, and at least $p^4(1-p)^2$ unless H_1 and H_2 are both P_3. Thus (H_1, H_2) induces a copy of C_5 with probability at least

$$p^4(1-p)^4 = (2c)^4(1-2c)^4 \overset{(3.1)}{\geqslant} 72 \cdot r(c).$$

Since we can upper-bound the probability of B by the probability that none of the $n^2/36$ dangerous pairs in (V_1, V_2) induces a C_5 in $\mathcal{G}(V_1, V_2, p)$, we obtain

$$\mathbb{P}[B] \leqslant \big(1 - 72 \cdot r(c)\big)^{n^2/36} \leqslant e^{-2r(c)n^2}.$$

Note that the number of edges between V_1 and V_2 in $\mathcal{G}(V_1, V_2, p)$ is binomially distributed. Thus, by Stirling's formula, $\mathbb{P}[A] \geqslant 1/(\sqrt{2\pi p(1-p)}n) \geqslant 1/n$. By the choice of n_0, combining this with (3.3) gives

$$P^* \leqslant n\, e^{-2r(c)n^2} \overset{(3.2)}{\leqslant} 2^{-2r(c)n^2}. \tag{3.4}$$

It remains to estimate the number N of ways to choose exactly $2cn^2$ edges between V_1 and V_2. We have

$$N \leqslant \binom{n^2}{2cn^2} \overset{(2.1)}{\leqslant} 2^{2h(c)n^2}. \tag{3.5}$$

This implies that the number of ways to add exactly $2cn^2$ edges to G that run between V_1 and V_2 without inducing a C_5 is

$$P^* \cdot N \overset{(3.4),(3.5)}{\leqslant} 2^{-2r(c)n^2} \cdot 2^{2h(c)n^2}. \qquad \square$$

Next, we want to show that Lemma 3.2 allows us to derive an upper bound on the number of graphs G such that (a) G has no large homogeneous sets and (b) G has a fixed type R which does not contain a triangle with three grey edges. Our aim is to obtain an upper bound which is much smaller than the bound provided in Lemma 2.12 for the corresponding problem without restriction (a). Lemma 3.3 states that this is possible. Recall for this purpose the definition of $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ from (2.9).

Lemma 3.3. *For every c with $0 < c \leqslant \frac{1}{2}$, and every $\gamma > 0$, there exist $\varepsilon_0, d_0 > 0$ and an integer k_0 such that, for all integers $k_1 \geqslant k_0$, k', there is an integer n_0 such that, for all positive $d \leqslant d_0$, $\varepsilon \leqslant \varepsilon_0$, ε', and all integers $n \geqslant n_0$ and $k_0 \leqslant k \leqslant k_1$, the following holds. If R is a coloured graph of order k which has at most εk^2 non-edges and does not contain a triangle with three grey edges, and $\eta = 1/(6k_1)$, then*

$$|\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c) \cap \mathcal{F}orb^*_{n,\eta}(C_5, c)| \leqslant 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\gamma\binom{n}{2}}.$$

In the proof of this lemma we combine the strategy of the proof of Lemma 2.12 with an application of Lemma 3.1 to all clusters of a partition corresponding to R, and an application of Lemma 3.2 to regular pairs of medium density. We shall make use of the following observation.

Using the definition of $r(c)$ from (3.1), it is easy to check that $f(c) := h(c) - r(c)$ is a concave function for $c \in (0, 1)$. Thus f enjoys the following property, which is a special form of Jensen's inequality (see, *e.g.*, [13]).

Proposition 3.4 (Jensen's inequality). *Let f be a concave function, $0 < c < 1$, $0 < c_i < 1$ for $i \in [m]$ and let $\sum_{i=1}^{m} c_i = mc$. Then*

$$\sum_{i=1}^{m} f(c_i) \leqslant m \cdot f(c). \qquad \square$$

Proof of Lemma 3.3. Let $c, \gamma > 0$ be given and choose ε_0, d_0, k_0 and n_0 as in the proof of Lemma 2.12. Let $d \leqslant d_0$ and k_1 be given, and possibly increase n_0 so that $n_0 \geqslant 2k_1 n_{\mathrm{L3.2}}$, where $n_{\mathrm{L3.2}}$ is the constant from Lemma 3.2 with parameter d, and so that

$$\frac{3}{4}k_1^2 \log n_0 + 2n_0 \leqslant \frac{\gamma}{10}\binom{n_0}{2}. \tag{3.6}$$

If necessary decrease ε_0 so that

$$3\varepsilon_0 \log \frac{1}{c} \leqslant \frac{\gamma}{10}. \tag{3.7}$$

Let $\varepsilon \leqslant \varepsilon_0$, ε', $n \geqslant n_0$, and k with $k_0 \leqslant k \leqslant k_1$ be given.

Let $R = ([k], E_R, \sigma)$ be a coloured graph which has at most εk^2 non-edges and does not contain a triangle with three grey edges. In the proof of Lemma 2.12 we counted the number of graphs in $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ by estimating the number of equipartitions $V_0 \dot\cup \cdots \dot\cup V_k$ of $[n]$, the number of choices for edges with one end in the exceptional set V_0 and edges in pairs (V_i, V_j) such that $\{i, j\} \notin E_R$, the number of choices for edges inside clusters V_i (such that i is white or black in R), and the number of choices for at most $c\binom{n}{2}$ edges in pairs (V_i, V_j) such that $\{i, j\}$ is a white, black, or grey edge of R. Now we are interested in the number of graphs in $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c) \cap \mathcal{F}orb^*_{n,\eta}(C_5, c)$. Clearly we can use the same strategy, and it is easy to verify that the estimates in (2.10)–(2.13) and thus in (2.14) from the proof of Lemma 2.12 remain valid in this setting. From now on, as in the proof of Lemma 2.12, we fix a partition $V_0 \dot\cup \cdots \dot\cup V_k$ of $[n]$ and observe that also $m_g := |E_g| \leqslant \frac{k^2}{4}$ still holds for the grey edges E_g in R. However, we shall now use Lemma 3.2 to obtain an improved bound on the number of possible choices for edges in E_g-pairs (V_i, V_j), and use this to replace (2.17) by a smaller bound on the number N_g of possible ways to distribute at most $c\binom{n}{2}$ edges to E_g-pairs. Since in the following we do not rely on any interferences between different E_g-pairs, clearly N_g will be maximal if m_g is maximal, and hence we assume from now on that

$$m_g = \frac{k^2}{4}. \tag{3.8}$$

Let $s := |V_1| = \cdots = |V_k|$ and observe that

$$\frac{n}{k} \geqslant s \geqslant (1-\varepsilon)\frac{n}{k} \geqslant n_{\mathrm{L}3.2}. \tag{3.9}$$

By Lemma 3.1, for each cluster V_i of a partition P of a graph in

$$\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c) \cap \mathcal{F}orb^*_{n,\eta}(C_5, c)$$

such that P corresponds to R, we have that V_i contains either $s/6$ copies of P_3 or $s/6$ copies of $\overline{P}_3$. Hence we will assume from now on that in our fixed partition the clusters V_i have this property.

We now upper-bound N_g by multiplying the possible ways A to assign at most $c\binom{n}{2}$ edges to one E_g-pair each, and the maximum number B of ways to choose all these assigned edges in the corresponding pairs, without inducing a C_5. First observe that we have

$$A \leqslant \left(c\binom{n}{2}\right)^{m_g+1} \leqslant n^{3m_g} \overset{(3.8)}{\leqslant} 2^{\frac{3}{4}k^2 \log n}. \tag{3.10}$$

For estimating B, we now assume that we have fixed an assignment in which each pair (V_i, V_j) with $\{i, j\} \in E_g$ is assigned $2c_{i,j}s^2$ edges. Let $\hat{c}$ be such that

$$\sum_{\{i,j\} \in E_g} 2c_{i,j}s^2 =: \hat{c}n^2 \leqslant c\binom{n}{2}. \tag{3.11}$$

Observe further that, since $\{i, j\} \in E_g$ is a grey edge of R, and we are interested in counting graphs with a partition corresponding to R, we can assume that $d \leqslant 2c_{i,j} \leqslant (1-d)$. Hence, by (3.9) we can apply Lemma 3.2 with $c_0 = d$ to infer that for each $\{i, j\} \in E_g$ there are at most

$$B_{ij} \leqslant 2^{2s^2\left(h(c_{i,j}) - r(c_{i,j})\right)} \tag{3.12}$$

possible ways to choose the $2c_{i,j}s^2$ edges in (V_i,V_j) without inducing a C_5. Now let $2\tilde{c} := \sum_{\{i,j\}\in E_g} 2c_{i,j}/m_g$ and observe that

$$\tilde{c} \overset{(3.11)}{=} \hat{c}\frac{n^2}{2s^2m_g} \overset{(3.9)}{\leqslant} \hat{c}\frac{k^2}{2(1-\varepsilon)^2m_g} \overset{(3.8)}{\leqslant} 2(1+3\varepsilon)\hat{c} \overset{(3.11)}{\leqslant} (1+3\varepsilon)c \leqslant \frac{3}{4}.$$

Therefore, since $f(x) := h(x)-r(x)$ is a concave function for $x\in(0,1)$, which is moreover non-decreasing for $x\leqslant \frac{3}{4}$, we can infer from Lemma 3.4 that

$$\sum_{\{i,j\}\in E_g} f(c_{i,j}) \leqslant m_g\cdot f(\tilde{c}) \leqslant m_g\cdot f\big(c(1+3\varepsilon)\big) \overset{(3.8)}{=} \frac{k^2}{4} f\big(c(1+3\varepsilon)\big). \tag{3.13}$$

As $h(x)$ is a convex function with $h'(x)\leqslant \log(1/x)$ and $r(x)$ is non-decreasing for $x\leqslant 3/4$, we have

$$f\big(c(1+3\varepsilon)\big) \leqslant h(c+3\varepsilon)-r(c) \leqslant h(c)+3\varepsilon h'(c)-r(c) \overset{(3.7)}{\leqslant} h(c)-r(c)+\frac{\gamma}{10}.$$

Together with (3.12) and (3.13), this implies

$$B=\prod_{\{i,j\}\in E_g} B_{ij} \leqslant 2^{2s^2(k^2/4)\left(h(c)-r(c)+\gamma/10\right)} \overset{(3.9)}{\leqslant} 2^{(n^2/2)\left(h(c)-r(c)+\gamma/10\right)},$$

which in turn, together with (3.10), gives

$$N_g \leqslant 2^{\frac{3}{4}k^2\log n}\cdot 2^{(n^2/2)\left(h(c)-r(c)+\gamma/10\right)} \overset{(3.6)}{\leqslant} 2^{\left(h(c)-r(c)\right)\binom{n}{2}+(\gamma/5)\binom{n}{2}}.$$

By multiplying this with (2.14) from the proof of Lemma 2.12, we obtain

$$|\mathcal{R}(R,\varepsilon,\varepsilon',d,k',n,c)\cap \mathcal{F}orb^*_{n,1/6k}(C_5,c)| \leqslant 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\gamma\binom{n}{2}},$$

as claimed. □

Lemma 3.3 with the type lemma, Lemma 2.7, implies an upper bound on $|\mathcal{F}orb^*_{n,\eta}(C_5,c)|$. Now we can combine this with the lower bound on $|\mathcal{F}orb^*_n(C_5,c)|$, which follows from Lemma 2.1, in order to prove Theorem 1.3.

Proof of Theorem 1.3. Observe first that, since C_5 is self-complementary, it suffices to prove Theorem 1.3 for $c\leqslant 1/2$. Hence we assume $c\leqslant 1/2$ from now on.

We first need to set up some constants. Given $c\in(0,\frac{1}{2}]$, we choose $\gamma>0$ such that $2\gamma<r(c)$. For input c and γ Lemma 2.1 supplies us with a constant $n_{\mathrm{L2.1}}$. We apply Lemma 3.3 with input c and $\gamma/2$ to obtain $\varepsilon_{\mathrm{L3.3}}$, d_0, and k_0. Next, we apply Lemma 2.11 with input d_0 and obtain constants $\varepsilon_{\mathrm{L2.11}}$ and ε'. Let $\varepsilon := \min\{\varepsilon_{\mathrm{L3.3}},\varepsilon_{\mathrm{L2.11}}\}$. For input ε, ε', and k_0, Lemma 2.7 returns constants k_1 and $n_{\mathrm{L2.7}}$. With this parameter k_1 we continue the application of Lemma 3.3 and obtain $n_{\mathrm{L3.3}}$. Choose $n_0 := \max\{n_{\mathrm{L2.1}}, n_{\mathrm{L2.7}}, n_{\mathrm{L3.3}}, \frac{3}{\sqrt{\gamma}}k_1\}$, assume that $n\geqslant n_0$, and set $\eta := 1/(6k_1)$.

Now, for each graph $G\in\mathcal{F}orb^*_{n,\eta}(C_5,c)$ we apply the type lemma, Lemma 2.7, with parameters ε, ε', k_0, k' and d, and obtain an $(\varepsilon,\varepsilon',d,k')$-type R of G on k vertices with $k_0\leqslant k\leqslant k_1$ and with at most εk^2 non-edges. Let $\tilde{\mathcal{R}}$ be the set of types obtained from these applications of Lemma 2.7. It follows that $|\tilde{\mathcal{R}}|\leqslant 4^{\binom{k_1}{2}}2^{k_1}\leqslant 2^{k_1^2}$. By Lemma 2.11 applied with d_0, ε, and ε', no coloured graph in $\tilde{\mathcal{R}}$ contains a triangle with three grey edges. Hence, by Lemma 3.3 applied with c, $\gamma/2$, ε, ε',

n, and k we have

$$|\mathcal{R}(R,\varepsilon,\varepsilon',d,k',n,c)\cap \mathcal{F}orb^*_{n,\eta}(C_5,c)| \leqslant 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\frac{1}{2}\gamma\binom{n}{2}}.$$

Since, by Lemma 2.7,

$$\mathcal{F}orb^*_{n,\eta}(C_5,c) \subseteq \bigcup_{R\in\tilde{\mathcal{R}}} \left(\mathcal{R}(R,\varepsilon,\varepsilon',d,k',c,n)\cap \mathcal{F}orb^*_{n,\eta}(C_5,c)\right),$$

we conclude from the choice of n_0 that

$$|\mathcal{F}orb^*_{n,\eta}(C_5,c)| \leqslant 2^{k_1^2}\cdot 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\frac{1}{2}\gamma\binom{n}{2}} \leqslant 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\gamma\binom{n}{2}}.$$

On the other hand, by Lemma 2.1 and the choice of n_0 we have

$$|\mathcal{F}orb^*_{n}(C_5,c)| \geqslant 2^{h(c)\binom{n}{2}-\gamma\binom{n}{2}}.$$

Since $2\gamma < r(c)$, by the choice of γ, this implies that almost all graphs in $\mathcal{F}orb^*_n(C_5,c)$ satisfy $\hom(G)\geqslant \eta n$. □

4. Concluding remarks

$\mathcal{G}(n,m)$ versus $\mathcal{G}(n,p)$

Our counting results can be interpreted as probabilities in the random graph model $\mathcal{G}(n,m = c\binom{n}{2})$. We can reformulate Theorem 1.2 as

$$\mathbb{P}\left[\mathcal{G}\left(n,m=c\binom{n}{2}\right)\in \mathcal{F}orb^*_n(C_5)\right] = \frac{|\mathcal{F}orb^*_n(C_5,c)|}{|\mathcal{G}(n,c)|} = 2^{(h(c)-H(c)+o(1))\binom{n}{2}}.$$

We now compare $\mathcal{G}(n,m=c\binom{n}{2})$ to the standard Erdős–Rényi model studied by Marchant and Thomason in [17]. They showed that

$$\mathbb{P}[\mathcal{G}(n,p)\in \mathcal{F}orb^*_n(C_5)] = 2^{c_p\binom{n}{2}},$$

where $c_p = \frac{1}{2}\max\{\log p, \log(1-p)\}$. We can now derive the same estimate via Theorem 1.2. Obviously $\mathbb{P}[\mathcal{G}(n,p)\in \mathcal{F}orb^*_n(C_5)]$ equals

$$\max_c\ \mathbb{P}\left[e(\mathcal{G}(n,p)) = c\binom{n}{2}\right]\cdot \mathbb{P}\left[\mathcal{G}\left(n,m=c\binom{n}{2}\right)\in \mathcal{F}orb^*_n(C_5)\right]\cdot 2^{o(n^2)}.$$

Setting $g_p(c) = H(c)+c\log p+(1-c)\log(1-p)$, we obtain

$$\mathbb{P}\left[e(\mathcal{G}(n,p)) = c\binom{n}{2}\right] = 2^{(g(c)+o(1))\binom{n}{2}},$$

and thus, by Theorem 1.2,

$$\mathbb{P}[\mathcal{G}(n,p)\in \mathcal{F}orb^*_n(C_5)] = 2^{(\max_c g_p(c)+h(c)-H(c)+o(1))\binom{n}{2}}.$$

The maximum is attained at $c = p/2$ for $p < 1/2$ and $c = (p+1)/2$ for $p > 1/2$. For $p = 1/2$ all values $c \in [1/4, 3/4]$ are optimal. Inserting the optimal value for c shows that the exponent is indeed equal to c_p as computed in [22]. This indicates that, for example, a graph from $\mathcal{G}(n,1/4)$ that happens to be induced C_5-free will a.a.s. have $(1/8+o(1))\binom{n}{2}$ edges. Thus the $\mathcal{G}(n,m)$ model can also be used to derive that a typical element having a certain property might be far from a typical element in $\mathcal{G}(n,p)$.

Extensions

The final thing to consider is whether more can be said about $\mathcal{F}orb_n^*(F,c)$ in the $\mathcal{G}(n,m)$ model. Indeed, there are at least three natural ways to enhance our results. First one might want to count $\mathcal{F}orb_n^*(F,c)$ for graphs F other than C_5. But already the case of a forbidden induced C_7 is more challenging, as tight upper bounds as in Lemma 2.12 are not so easy to derive. Furthermore, $h(c)$ does not seem to be unimodal for C_7.

Second, it would be interesting to obtain even sharper asymptotic bounds for the speed of $\mathcal{F}orb_n^*(F,c)$. Alon, Balogh, Bollobás and Morris [2] determine the speed of some hereditary properties up to a subquadratic term in the exponent. We believe their techniques can be extended to the case of restricted density.

Finally one might want to prove a much better constant in the size of the linear homogeneous sets that can be found in almost all graphs in $\mathcal{F}orb_n^*(C_5,c)$. It easily follows from the fact that almost all graphs in $\mathcal{F}orb_n^*(C_5)$ are generalized split graphs (see [19]) that almost all of them also have a homogeneous set of size $(1/2-o(1))n$. We believe that the same is true for the density-restricted case, and that this can be proved as in Theorem 1.3 combined with a stability-type argument. We plan to return to this in the near future.

Acknowledgements

The authors want to thank an anonymous referee for a careful reading and valuable suggestions.

References

[1] Alekseev, V. E. (1993) Range of values of entropy of hereditary classes of graphs. *Discrete Math. Appl.* **3** 191–199.

[2] Alon, N., Balogh, J., Bollobás, B. and Morris, R. (2011) The structure of almost all graphs in a hereditary property. *J. Combin. Theory Ser. B* **101** 85–110.

[3] Alon, N., Fischer, E., Krivelevich, M. and Szegedy, M. (2000) Efficient testing of large graphs. *Combinatorica* **20** 451–476.

[4] Alon, N. and Shapira, A. (2008) A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.* **37** 1703–1727.

[5] Balogh, J., Bollobás, B. and Simonovits, M. (2004) The number of graphs without forbidden subgraphs. *J. Combin. Theory Ser. B* **91** 1–24.

[6] Bollobás, B. and Thomason, A. (1997) Hereditary and monotone properties of graphs. In *The Mathematics of Paul Erdős II* (R. L. Graham and J. Nešetřil, eds), Vol. 14 of *Algorithms and Combinatorics*, Springer, pp. 70–78.

[7] Bollobás, B. and Thomason, A. (2000) The structure of hereditary properties and colourings of random graphs. *Combinatorica* **20** 173–202.

[8] Chudnovsky, M., Robertson, N., Seymour, P. and Thomas, R. (2006) The strong perfect graph theorem. *Ann. Math.* **164** 51–229.

[9] Erdős, P., Frankl, P. and Rödl, V. (1986) The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.* **2** 113–121.

[10] Erdős, P. and Hajnal, A. (1989) Ramsey-type theorems. *Discrete Appl. Math.* **25** 37–52.

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

[12] Gyárfás, A. (1997) Reflections on a problem of Erdős and Hajnal. In *The Mathematics of Paul Erdős II* (R. L. Graham and J. Nešetřil, eds), Vol. 14 of *Algorithms and Combinatorics*, Springer, pp. 93–98.

[13] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1988) *Inequalities*, Cambridge Mathematical Library, Cambridge University Press, reprint of the 1952 edition.
[14] Jukna, S. (2001) *Extremal Combinatorics*, Springer.
[15] Komlós, J., Shokoufandeh, A., Simonovits, M. and Szemerédi, E. (2000) The regularity lemma and its applications in graph theory. In *Theoretical Aspects of Computer Science* (G. B. Khosrovshahi, A. Shokoufandeh and M. A. Shokrollahi, eds), Vol. 2292 of *Lecture Notes in Computer Science*, Springer, pp. 84–112.
[16] Loebl, M., Reed, B., Scott, A., Thomason, A. and Thomassé, S. (2010) Almost all F-free graphs have the Erdős–Hajnal property. In *An Irregular Mind*, Vol. 21 of *Bolyai Society Mathematical Studies*, Springer, pp. 405–414.
[17] Marchant, E. and Thomason, A. (2011) The structure of hereditary properties and 2-coloured multigraphs. *Combinatorica* **31** 85–93.
[18] Prömel, H. J. and Steger, A. (1991) Excluding induced subgraphs: Quadrilaterals. *Random Struct. Alg.* **2** 55–72.
[19] Prömel, H. J. and Steger, A. (1992) Almost all Berge graphs are perfect. *Combin. Probab. Comput.* **1** 53–79.
[20] Prömel, H. J. and Steger, A. (1992) Excluding induced subgraphs III: A general asymptotic. *Random Struct. Alg.* **3** 19–31.
[21] Szemerédi, E. (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica* **27** 199–245.
[22] Thomason, A. (2011) Graphs, colours, weights and hereditary properties. In *Surveys in Combinatorics 2011*, Vol. 392 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, pp. 333–364.
[23] Turán, P. (1941) On an extremal problem in graph theory. *Matematicko Fizicki Lapok* **48** 436–452.