Small subgraphs in random graphs and the power of multiple choices

Torsten Mütze   Reto Spöhel†   Henning Thomas

Institute of Theoretical Computer Science
ETH Zürich, 8092 Zürich, Switzerland
{muetzet|rspoehel|hthomas}@inf.ethz.ch

ABSTRACT. The standard paradigm for online power of two choices problems in random graphs is the Achlioptas process. Here we consider the following natural generalization: Starting with $G_0$ as the empty graph on $n$ vertices, in every step a set of $r$ edges is drawn uniformly at random from all edges that have not been drawn in previous steps. From these, one edge has to be selected, and the remaining $r-1$ edges are discarded. Thus after $N$ steps, we have seen $rN$ edges, and selected exactly $N$ out of these to create a graph $G_N$.

In a recent paper by Krivelevich, Loh, and Sudakov [10], the problem of avoiding a copy of some fixed graph $F$ in $G_N$ for as long as possible is considered, and a threshold result is derived for some special cases. Moreover, the authors conjecture a general threshold formula for arbitrary graphs $F$. In this work we disprove this conjecture and give the complete solution of the problem by deriving explicit threshold functions $N_0(F,r,n)$ for arbitrary graphs $F$ and any fixed integer $r$. That is, we propose an edge selection strategy that a.a.s. (asymptotically almost surely, i.e. with probability $1-o(1)$ as $n \to \infty$) avoids creating a copy of $F$ for as long as $N = o(N_0)$, and prove that any online strategy will a.a.s. create such a copy once $N = \omega(N_0)$.

1. Introduction

Consider the following graph process. Starting with the empty graph on $n$ vertices, in every step a new edge is drawn uniformly at random from all non-edges and inserted into the graph. It is natural to ask how many edges will typically appear until the resulting graph satisfies some monotone property, e.g., contains a triangle, is non-3-colorable, or is Hamiltonian.

It is not hard to see that the resulting graph after $N$ steps is distributed uniformly over all graphs on $n$ vertices with exactly $N$ edges. Thus the analysis of this process is closely related to analyzing the ‘static’ random graph $G_{n,m}$ (a graph drawn uniformly at random from all graphs on $n$ vertices with $m$ edges) introduced by Erdős and Rényi in the 60’s [7]. Since then, a large body of work has been devoted to the subject, and threshold results have been proved for many natural graph properties.

In this paper we study the property of containing a copy of some fixed subgraph $F$. For the Erdős-Rényi model introduced above, the following threshold result was proved by Bollobás [4] in 1981. Throughout, we denote the number of vertices and edges of a graph $F$ by $v(F)$ and $e(F)$, respectively. We tacitly assume that all graphs contain at least one vertex, and call a graph empty if it has no edges, nonempty otherwise. We write $f(n) \ll g(n)$ if $f(n) = o(g(n))$, and $f(n) \gg g(n)$ if $f(n) = \omega(g(n))$.

**Theorem 1** ([4]). Let $F$ be a fixed nonempty graph. Then

$$\lim_{n \to \infty} \Pr[G_{n,m} \text{ contains a copy of } F] = \begin{cases} 1 & \text{if } m \gg n^{2-1/m(F)} \\ 0 & \text{if } m \ll n^{2-1/m(F)} \end{cases}$$

†The author was supported by the Swiss National Science Foundation, grant 200021-108158.
Assume now that the graph process is modified as follows: In every step \( r \) edges are drawn uniformly at random from all edges that have not been drawn in previous steps. From these, one has to be selected, and the other \( r - 1 \) are discarded. Thus after \( N \) steps, we have seen \( rN \) edges, and selected exactly \( N \) out of these to create a graph \( G_N \). How does this freedom of choice affect the thresholds that are known for the first model? Can we significantly increase or decrease the number of steps it typically takes until \( G_N \) satisfies some monotone property?

This process is known in the literature as an Achlioptas process (where usually \( r = 2 \)) and has become the standard paradigm for online power of two choices problems in the context of random graphs. Among many natural properties of random graphs, the one that has received most attention so far is the property of containing a linear-sized component. It turns out that the emergence of such a ‘giant component’ can be both accelerated or slowed down by a constant factor if appropriate edge selection strategies are used \([11, 12, 13, 14, 15]\).

Only recently, other graph properties have been studied in this context \([10, 11]\). Motivated by Theorem 1, Krivelevich, Loh, and Sudakov considered the problem of avoiding a copy of some fixed graph \( F \) in \( G_N \) for as long as possible. For some special cases, they proved an explicit threshold function \( N_0 = N_0(F, r, n) \) in the following sense: For any \( N \ll N_0 \), there exists an edge selection strategy that a.a.s. (asymptotically almost surely, i.e. with probability tending to 1 as \( n \) tends to infinity) does not create a copy of \( F \) in the first \( N \) steps of the process. On the other hand, if \( N \gg N_0 \) then any online strategy will a.a.s. create a copy of \( F \) within the first \( N \) steps. (Such an online threshold exists for any monotone graph property, as can be shown along the lines of the multi-round exposure proof of the well-known offline result by Bollobás and Thomason \([6]\), see \([14, 16]\).) The results of \([10]\) can be summarized as follows.

**Theorem 2**  \([10]\). Let \( F \) be a cycle, a complete graph or a complete bipartite graph with parts of equal size, and let \( r \geq 2 \) be a fixed integer. Then the threshold for avoiding \( F \) in the generalized Achlioptas process with parameter \( r \) is

\[
N_0(F, r, n) = n^{2 - 1/\tilde{d}^*(F)},
\]

where

\[
\tilde{d}^*(F) := \max_{s \geq 1} \frac{r^s(e(F) - s) + r^{s-1}}{r^s(v(F) - 2) + 2}.
\]  \( (1) \)

The maximum in \([1]\) is attained for

- \( s = 1 \) if \( F = C_\ell \) is a cycle of length \( \ell \geq 3 \),
- \( s = \lceil \log_r((r-1)\ell + 1) \rceil \) if \( F = K_\ell \) is a complete graph of size \( \ell \geq 4 \),
- \( s = \lceil \log_r((r-1)\ell + 1) \rceil \) if \( F = K_{\ell, \ell} \) is a complete bipartite graph with parts of size \( \ell \geq 3 \).

Moreover, they conjectured that the general threshold of the problem is \( N_0(F, r, n) = n^{2 - 1/\tilde{m}^*(F)} \), where

\[
\tilde{m}^*(F) := \max_{H \subseteq F} \tilde{d}^*(H).
\]  \( (2) \)

In this work we disprove this conjecture and give the general solution of the problem.

1.1. **Our result.** In order to state our result, we need to introduce some terminology. Throughout this work, we will use the notion of ordered graphs. An ordered graph is a pair \((H, \pi)\), where \( H \) is a graph, \( h := e(H) \), and \( \pi : E(H) \to \{1, \ldots, h\} \) is an ordering of the edges of \( H \), conveniently denoted by its preimages, \( \pi = (\pi^{-1}(1), \ldots, \pi^{-1}(h)) \). In the context of the Achlioptas process, we interpret
the ordering \( \pi =: (e_1, \ldots, e_h) \) as the order in which the edges of \( H \) appeared in the process, where \( e_h \) is the edge that appeared first (the ‘oldest’ edge) and \( e_1 \) is the edge that appeared last (the ‘youngest’ edge).

We denote by \( \Pi(E(H)) \) the set of all possible orderings of the edges of \( H \), and by

\[
S(F) := \{(H, \pi) \mid H \subseteq F \land \pi \in \Pi(E(H))\}
\]

the set of all ordered subgraphs of \( F \). For some ordered graph \((H, \pi)\) and some subgraph \( J \subseteq H \), we denote by \( \pi|_J \) the order on the edges of \( J \) induced by \( \pi \). Given an ordered graph \((H, \pi)\), \( \pi = (e_1, \ldots, e_h) \), we denote by \( H \setminus e_1 \) the graph obtained from \( H \) by removing the edge \( e_1 \). As we do not remove vertices (even if they become isolated), we always have \( v(H \setminus e_1) = v(H) \). We use \( \pi \setminus e_1 \) as a shorthand notation for \( \pi|_{H \setminus e_1} \), and \( e_1 \in J \) as a shorthand notation for \( e_1 \in E(J) \).

For a fixed integer \( r \geq 2 \) and a fixed real value \( 0 \leq \theta \leq 2 \), we introduce for any graph \( H \) and any \( \pi = (e_1, \ldots, e_h) \in \Pi(E(H)) \) the parameter \( \lambda_{r,\theta}(H, \pi) \), defined recursively by

\[
\lambda_{r,\theta}(H, \pi) := \begin{cases} v(H), & \text{if } e(H) = 0, \\ \lambda_{r,\theta}(H \setminus e_1, \pi \setminus e_1) - \theta + (r - 1) \cdot \min_{J \subseteq H: e_1 \in J} (\lambda_{r,\theta}(J \setminus e_1, \pi|_{J \setminus e_1}) - 2), & \text{otherwise.} \end{cases}
\]

As we shall see, using an optimal edge selection strategy ensures that a.a.s. the number of copies of \((H, \pi)\) in \( G_N \) with \( N = n^{2-\theta} \) is of order \( n^{\lambda_{r,\theta}(H, \pi)} \) at most.

Furthermore, we set for \( r \) and \( \theta \) as before and any graph \( F \)

\[
\Lambda_{r,\theta}(F) := \max_{\pi \in \Pi(E(F))} \min_{H \subseteq F} \lambda_{r,\theta}(H, \pi|_H)
\]

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)

(4)
instead of placed back in the pool of available edges. This is essential for our approach since otherwise
the graph formed by the first $N$ many $r$-sets drawn in the process is not uniformly distributed
over all graphs with exactly $rN$ edges (and neither over all multigraphs with exactly $rN$ edges).
Nevertheless, the threshold given by Theorem 3 carries over to the setting with replacement; we
briefly outline the proof at the end of this paper.

The online problem studied here gives rise to the following offline problem: given a random $r$-
matched graph $G_{n,m}^r$, can we select one edge from each $r$-set to obtain a graph that does not
contain a copy of $F$? As was shown in [12], for ‘most’ graphs $F$ and any fixed integer $r \geq 2$, the
threshold of this offline problem is $n^{2-1/m_2(F)}$, where $m_2(F) := \max_{H \subseteq F}(e(H) - 1)/(v(H) - 2)$ is
the well-known 2-density. As it turns out, for any nonempty graph $F$ the parameter $\theta^*(F, r)$ as
defined in Theorem 3 satisfies

$$\lim_{r \to \infty} \theta^*(F, r) = 1/m_2(F) ,$$

i.e., the online threshold approaches the offline threshold as the parameter $r$ increases.

1.2. The edge selection strategy. We prove the lower bound in Theorem 3 by analyzing the
following generic edge selection strategy. We consider an ordered graph $(H, \pi) \in \mathcal{S}(F)$ the more
‘dangerous’ the lower $\lambda_{r, \theta^*}(H, \pi)$ is (intuitively, a high value $\lambda_{r, \theta^*}(H, \pi)$ means that we will have
many copies of $(H, \pi)$ anyway, cf. the remark after [3]). Thus for a concrete forbidden graph $F$, an optimal strategy can be described as a simple priority list of ordered
graphs $(H, \pi) \in \mathcal{S}(F)$.

Let us address one possible objection straight away. Our strategy remembers the order in which
ingraphs $(H, \pi)$ are considered. While we do not remember the order in which $r$-sets appear, but instead
remember the edge order in which the edges were inserted into the graph $G_N$, and it considers ordered subgraphs of $G_N$ to make its
decisions. This might seem pointless at first glance – after all, once a copy of $H$ is present in $G_N$,
why should it matter in which order $\pi \in \Pi(E(H))$ its edges appeared? However, in our approach
the order $\pi$ is not of interest per se, but conveniently encodes information about how a given copy of
$H$ overlaps with other copies of ordered graphs in $G_N$. Presumably, there is a ‘memoryless’ strategy
that does not remember the order in which edges appear, but we suspect that such a strategy would
have to look considerably more complicated.

1.3. Alternative formulation and special cases. We now present an equivalent formulation of
Theorem 3 that is more related to the maximum density perspective known from Theorem 1 and
Theorem 2.

Given an ordered graph $(H, \pi), \pi = (e_1, \ldots, e_h)$, we denote by $H \setminus \{e_1, \ldots, e_{i-1}\}$ the graph obtained
from $H$ by removing the edges $e_1, \ldots, e_{i-1}$. As before no vertices are removed. For any nonempty
ordered graph $(H_1, \pi), \pi = (e_1, \ldots, e_h)$, any sequence of subgraphs $H_2, \ldots, H_h \subseteq H_1$ with $H_i \subseteq
H_1 \setminus \{e_1, \ldots, e_{i-1}\}$ and $e_i \in H_i$ for all $2 \leq i \leq h$, and any integer $r \geq 2$, define coefficients $c_i := c_i((H_1, \pi), H_2, \ldots, H_h, r)$ recursively by

$$c_1 := r ,$$

$$c_i := (r - 1) \cdot \sum_{j=1}^{i-1} c_j 1_{\{e_i \in H_j\}} , \quad 2 \leq i \leq h ,$$

(7)
Corollary 7. For any $\ell \geq 1$ and $r \geq 2$, the threshold for avoiding $K_{1,\ell}$ in the generalized Achlioptas process with parameter $r$ is $N_0(K_{1,\ell}, r, n) = n^{1-1/e_{\star}(K_{1,\ell})}$, where

$$e_{\star}(K_{1,\ell}) = \frac{r^\ell - 1}{r - 1}. \tag{13}$$
From [12] and [13] we see that for fixed $r\geq 2$ the parameter $e_r^*(P_\ell)$ grows polynomially in $\ell$, while $e_r^*(K_{1,2})$ grows exponentially in $\ell$. This is in contrast with Theorem 5 which gives a threshold of $n^{1-1/\ell}$ for any tree on $\ell$ edges regardless of its structure.

1.4. Organization of this paper. Before actually proving Theorem 3 we outline the key ideas of our approach in Section 2. In Sections 3 and 4 we then prove the lower and upper bound in Theorem 5 respectively. Our upper bound proof relies on a somewhat technical deterministic approach. We then illustrate with several examples how this result is applied in our proofs, and try to convey some intuition for the resulting threshold formulas.

2. The random $r$-matched graph approach

2.1. $r$-matched graphs. An $r$-matched graph is a pair $H = (V, K)$ such that $K \subseteq \binom{V}{2}$ is a family of disjoint $r$-sets of edges. For any $r$-matched graph $H$ we use $V(H)$ and $K(H)$ to denote the set of vertices and the family of $r$-sets of $H$, and we denote the cardinality of these sets by $v(H)$ and $\kappa(H)$, respectively. Observe that subgraph containment, isomorphism and similar notions have natural extensions to $r$-matched graphs.

The generalized Achlioptas process can be described by $r$-matched graphs $(G^r(n, N))_{0 \leq N \leq \binom{V}{2}/r}$, where $G^r(n, 0) = (V(K_n), \emptyset)$ and $G^r(n, N)$ is obtained from $G^r(n, N-1)$ by adding an $r$-set $E^N$ drawn uniformly at random from all edges in $\binom{V(K_n)}{2} \setminus (E^1 \cup \cdots \cup E^{N-1})$. Thus we have $G^r(n, N) = (V(K_n), \{E^1, \ldots, E^N\})$, and our goal is to select one edge $f^N$ from each $r$-set $E^N$ immediately in such a way that $G_N = (V(K_n), \{f^1, \ldots, f^N\})$ does not contain a copy of $F$ for as long as possible. Note that, by symmetry, $G^r(n, N)$ is distributed like $G_{n,m}$ with $m = rN$ if we ignore the partition into $r$-sets (and the order in which the edges appear).

By $G^r_{n,m}$ we denote a random $r$-matched graph obtained by first generating a normal random graph $G_{n,m}$ and then choosing a random partition of its edge set into sets of size $r$ uniformly at random (w.l.o.g. we assume that $m$ is divisible by $r$). Again by symmetry, $G^r(n, N)$ is distributed like $G^r_{n,m}$ with $m = rN$ if we take into account the partition into $r$-sets (but ignore the order in which those appear). As mentioned in the introduction, this allows us to analyze the process $(G^r(n, N))_{0 \leq N \leq \binom{V}{2}/r}$ by studying the ‘static’ object $G^r_{n,m}$.

All our asymptotic results are with respect to $n$, the number of vertices of $G^r_{n,m}$ or $G_N$. We write $f(n) \asymp g(n)$ if $f(n) = \Theta(g(n))$.

Theorem 8 generalizes to $r$-matched graphs as follows.

**Theorem 8.** Let $r \geq 1$ be a fixed integer, and let $F$ be a fixed nonempty $r$-matched graph. Then

$$\lim_{n \to \infty} \Pr[G^r_{n,m} \text{ contains a copy of } F] = \begin{cases} 1 & \text{if } m \gg n^{2-1/m(F)} \\ 0 & \text{if } m \ll n^{2-1/m(F)} \end{cases},$$

where

$$m(F) := \max_{H \subseteq F} \frac{\kappa(H)}{v(H) - 2(r-1)\kappa(H)}.$$

In our proofs we will need two related statements that are given by the following lemma.
Lemma 9. Let \( r \geq 1 \) be a fixed integer, and let \( F \) be a fixed nonempty \( r \)-matched graph.

i) If \( m \gg 1 \), the expected number of copies of \( F \) in \( G_{n,m}^r \) is \( \Theta \left( n^{v(F)}(mn^{-2r})^{\kappa(F)} \right) \).

ii) If \( m \gg n^{2-1/m(F)} \), the number of copies of \( F \) in \( G_{n,m}^r \) is \( \Theta \left( n^{v(F)}(mn^{-2r})^{\kappa(F)} \right) \) a.a.s.

Proof. We only prove part i) of the lemma. Part ii) and Theorem 8 then follow analogously to the textbook first and second moment proof of Theorem 1 (see e.g. [5] or [9]).

Let Aut(\( F \)) denote the number of isomorphisms of \( F \). There are
\[
\binom{n}{v(F)} \binom{v(F)!}{\text{Aut}(F)} \asymp n^{v(F)}
\]
possible copies of \( F \) in \( K_n \), and each of them is present in \( G_{n,m}^r \) with probability
\[
\frac{\binom{n}{v(F)} - \kappa(F)r}{\binom{n}{v(F)}} \cdot \frac{1}{\binom{m}{v(F)} - \kappa(F)(r-1)} \asymp (mn^{-2})^{\kappa(F)}.
\]

Here the first term in the first line is the probability that all \( \kappa(F)r \) edges of a fixed copy of \( F \) are present, and the remaining terms are the probability that these edges are partitioned into \( r \)-sets in the right way.

For any \( r \)-matched graph \( H \) and \( 0 \leq \theta \leq 2 \), let
\[
\mu_{r,\theta}(H) := v(H) - 2(r-1)\kappa(H) - \kappa(H) \cdot \theta.
\]

Note that by Lemma 9, the expected number of copies of \( H \) in \( G_{n,m}^r \) with \( m = n^{2-\theta} \) (for some \( 0 < \theta < 2 \)) is \( \Theta(n^{\mu_{r,\theta}(H)}) \), and that the threshold given by Theorem 8 can alternatively be written as \( n^{2-\bar{\theta}} \), where \( \bar{\theta} = \bar{\theta}(F,r) \) is the unique solution of \( \min_{H \subseteq F} \mu_{r,\theta}(H) = 0 \).

2.2. Combinatorial interpretation of Theorem 4. We now outline how the results from the previous section are applied in our approach. Our point of view in this section is mainly a lower bound perspective: how can we analyze the performance of a given edge selection strategy, and which strategies yield good lower bounds?

We start by discussing two examples for which the threshold was already found in [10] (cf. Theorem 2). Our approach via \( r \)-matched graphs allows us to interpret the density \( d^* \) as defined in [1] combinatorially. In this combinatorial interpretation, it then becomes quite apparent that the formula in Theorem 2 fails to capture several features of the general problem. Our point of reference in this section is Theorem 4 which can be compared to Theorem 2 more directly than Theorem 3.

2.2.1. The gluing intuition. Consider the example \( F = K_4 \) and \( r = 3 \), and suppose we use the simplest strategy imaginable: In every step \( N \), we select an arbitrary edge \( f^N \) from \( E^N \) that does not close a copy of \( F = K_4 \) if such an edge is available. We shall refer to this strategy as the naive strategy in the following. Then clearly we lose in step \( N \) if and only if every edge from \( E^N \) would close copy of \( K_4 \) in \( G_N \). Typically this means that \( G^r(n,N) \) contains an \( r \)-matched subgraph \( H \) as shown on the left hand side of Figure 1. (In all figures of this section, the dashed ovals indicate \( r \)-sets, and edges that were presented but not selected are drawn grey. Note that for many black edges the \( r-1 \) grey partner edges are omitted since they are irrelevant in our considerations.) In
some sense, the \( r \)-matched graph \( H \) captures a possible ‘history of failure’ for the naive strategy. Theorem 8 allows us to conclude that as long as \( N \ll n^{2-1/d(H)} \), where

\[
d(H) := \frac{\kappa(H)}{v(H) - 2(r - 1)\kappa(H)},
\]

\( H \) a.a.s. does not appear as a subgraph of \( G_{n,r,N} \), and that consequently we do not lose as shown in Figure 1. (In general, it is possible that the three copies of \( F = K_4 \) overlap to some extent, giving rise to several more possible ‘histories of failure’. However, we can argue in exactly the same way for each of them and obtain that none of these appears as long as \( N \ll n^{2-1/d(H)} \). For the rest of this section, we ignore this issue and focus on ‘typical’ histories only, i.e. those in which no overlappings occur.) Since \( d(H) = 16/8 = 2 \), it follows that the naive strategy ‘survives’ a.a.s. as long as \( N \ll n^{2-1/2} = n^{3/2} \). Theorem 2 asserts that this is indeed the threshold for this example.

Observe that the terms \( \kappa(H) \) and \( v(H) - 2(r - 1)\kappa(H) \) in the definition of \( d(H) \) have the following combinatorial interpretation: If we contract the \( r \) edges of each \( r \)-set of \( H \) into a single edge, we obtain a graph \( \bar{H} \) with exactly \( \kappa(H) \) edges and \( v(H) - 2(r - 1)\kappa(H) \) vertices (see Figure 1). We also refer to this as ‘gluing together’ the three copies of \( F = K_4 \). Thus we have

\[
d(H) = \frac{e(\bar{H})}{v(\bar{H})},
\]

where \( \bar{H} \) is the glued version of \( H \).

Summarizing, the gluing intuition allowed us to analyze the naive strategy by constructing the associated history graph \( H \) and counting the edges and vertices of its glued version \( \bar{H} \). As we shall see next, this approach also works for more complicated strategies.

2.2.2. Maximization over subgraph sequences. For \( F = K_4 \) and \( r = 2 \) the naive strategy can be analyzed analogously: the glued version \( \bar{H} \) of the corresponding history graph (two copies of \( K_4 \) glued together at an edge) yields \( e(\bar{H})/v(\bar{H}) = 11/6 \), and thus the lower bound guaranteed by the naive strategy is \( n^{2-6/11} = n^{16/11} \). As it turns out, this is not optimal: A better strategy is to not
only avoid copies of $K_4$, but to also consider $K_4 \setminus e$ (a lower-priority) threat. (Here $K_4 \setminus e$ denotes the graph obtained by removing one edge from $K_4$.) A ‘typical’ history graph $H$ corresponding to this strategy is shown in Figure 1 (every copy of $K_4$ or $K_4 \setminus e$ was closed because the strategy was forced to do so by a second copy of the same type). We obtain $\frac{e(H)}{v(H)} = 19/10 > 11/6$, i.e., this two-stage strategy is indeed better than the naive strategy. According to Theorem 2 this is best possible, i.e., $n^{2-10/19} = n^{28/19}$ is in fact the threshold for this example.

Note that in this example the history of failure goes back two steps in time. In all our examples we indicate the order in which the last $r$-sets appeared by $e_1$ and $e_2$, where $e_1$ denotes the last $r$-set and $e_2$ the $r$-sets that appeared just before $e_1$.

Similarly we can analyze strategies that go over even more stages. As a result, determining the best strategy for a given graph $F$ essentially corresponds to determining a sequence of subgraphs $H_1, \ldots, H_h \subseteq F$ that should be avoided. This is reflected by the maximizations in (8) and (9). In fact, the numerator and denominator on the right hand side of (8) count exactly the number of edges and vertices of the glued version $\bar{H}$ of the history graph $H$ corresponding to the sequence $H_1, \ldots, H_h$. Note that the formalism allows to set $H_i = F$, which results in subgraph sequences that have strictly less than $h$ relevant entries. For the examples $F = K_4$ and $r = 3$ or $r = 2$ discussed above, maximizing subgraph sequences are given by $H_1 = K_4, H_2 = \cdots = H_6 = K_2$ and $H_1 = K_4, H_2 = K_4 \setminus e, H_3 = \cdots = H_6 = K_2$, respectively.

Let $F \setminus ie$ denote any graph obtained by removing exactly $i$ edges from $F$. The formula in Theorem 2 corresponds to the special case of (8) when

$$H_i = F \setminus (i-1)e \ , \ 1 \leq i \leq s \ ,$$
$$H_i = K_2 \ , \ s + 1 \leq i \leq e(F) \ ,$$

and thus $c_i = r^i - r^{i-1}$ for $2 \leq i \leq s$ (note that the values $c_{s+1}, \ldots, c_{e(F)}$ are irrelevant). Observe that setting $s = 1$ yields the lower bound resulting from the naive strategy.
In general, a maximizing subgraph sequence may look different from (15). Consider for instance the ‘bow tie’ graph depicted in Figure 2. Here the maximizing sequence is $H_1 = F$, $H_2 = K_3$ (and $H_3 = \cdots = H_6 = K_2$), yielding $m^r(F) = \frac{15}{10}$ (and hence a threshold of $n^{2-10/15} = n^{2/3}$), whereas (2) evaluates to the wrong value $\tilde{m}^r(F) = \tilde{d}^r(F) = \max\{\frac{11}{8}, \frac{19}{14}, \ldots\} = \frac{11}{8}$ (cf. Figure 3).

2.2.3. Minimization over edge orderings. An essential feature of the problem that is out of control of a strategy is the order in which the edges arrive to close certain substructures. We illustrate this with the example shown in Figure 3. Suppose our strategy is as follows: avoid copies of $F$ with high priority, and copies of $C_3$ and $C_4$ with lower priority. Then there are two ‘typical’ ways of losing: either we lose because we first closed copies of $C_4$ and then a copy of $F$ ($H_1 = F$, $H_2 = C_4$), or because we first closed copies of $C_3$ and then a copy of $F$ ($H_1 = F$, $H_2 = C_3$). Since we lose as soon as either one of these scenarios occurs, the resulting lower bound is given by the minimum of the two values $19/14$ and $17/12$ (cf. Figure 3).

This is reflected by the minimization over all edge orderings $\pi \in \Pi(E(F))$ in (9). In general, different edge orderings of $F$ yield different maximizing subgraph sequences, cf. (8). Note that an optimal strategy cannot just focus on the most dangerous ordering, but will have to take care of all orderings (i.e. all possible ways of losing) simultaneously.

2.2.4. Trees. Let us point out one last phenomenon, which is particularly important for sparse graphs like trees: It is possible that $F$ does not grow step by step as in the previous examples, but is obtained from several components that first evolve separately and only grow together later in the process. This is illustrated in Figure 4 to the left, where the last black edge connects two $P_3$’s that have evolved independently before.

In general, if the forbidden graph $F$ is a tree, then the glued version $\tilde{H}$ of a ‘typical’ $r$-matched history graph $H$ is a tree as well, satisfying $v(\tilde{H}) = e(\tilde{H}) + 1$ (cf. Figure 4). The formula in Theorem 5 is the result of directly considering $e^{r*}(F) = e(\tilde{H})$ instead of $m^{r*}(F) = e^{r*}(F)/(e^{r*}(F) + 1)$. 

Figure 3. Minimization over edge orderings.
2.3. Equivalence of Theorem 3 and Theorem 4. We conclude this section by providing an intuitive justification for the two dual formulations of our main result.

In the above examples, our goal was to find strategies that guaranteed that \( F \) appeared as late as possible in the Achlioptas process. This boiled down to maximizing \( d(H) = \frac{e(\bar{H})}{v(\bar{H})} \), which essentially corresponded to maximizing \( H_1, \ldots, H_h \) in (8) and (9).

Suppose now that instead we want to minimize the number of copies of \( F \) after exactly \( N = n^{2-\theta} \) steps of the process, for some fixed \( 0 < \theta < 2 \). In this setting we can evaluate the performance of a given strategy quite similarly to before, except that now we have to minimize the expected number of copies of some history graph \( H \) in \( G^r(n, N) \) (instead of maximizing \( d(H) \)). By Lemma 9, this expectation is of order

\[
 n^{v(H) - 2(r-1)\kappa(H) - \kappa(H) \cdot \theta} \cdot \mu_{r, \theta}(H)
\]

where, using the gluing intuition, the exponent \( \mu_{r, \theta}(H) \) can be written as

\[
 \mu_{r, \theta}(H) = v(H) - 2(r-1)\kappa(H) - \kappa(H) \cdot \theta = v(\bar{H}) - e(\bar{H}) \cdot \theta.
\]

So rather than maximizing the ratio \( e(\bar{H})/v(\bar{H}) \) we now minimize the linear function \( v(\bar{H}) - e(\bar{H}) \cdot \theta \) for some fixed \( 0 < \theta < 2 \). If the fixed parameter \( \theta \) is such that \( \mu_{r, \theta}(H) \leq 0 \), then for \( N \ll n^{2-\theta} \) the history graph \( H \) a.a.s. does not appear in \( G^r(n, N) \) by Markov’s inequality. By the same arguments as before, this implies that no copy of \( F \) is closed in the first \( N \) steps, and the smallest \( \theta \) (corresponding to the largest \( N = n^{2-\theta} \)) for which we can argue like this is of course \( \theta^* \) as defined in Theorem 3.

The advantage of this slightly roundabout way of carrying out the same arguments as in Section 2.2 is that the problem of minimizing subgraph counts for some fixed \( N = n^{2-\theta} \) has a recursive structure. In particular, for \( 0 < \theta < 2 \) fixed, a sequence \( H_1, \ldots, H_h \subseteq F \) minimizing \( \mu_{r, \theta}(H) \) of the associated history graph \( H \) can be found recursively (instead of by an unwieldy global optimization as in (8)).

This leads to the recursive definition of the parameter \( \lambda_{r, \theta}(\cdot) \) in (3), which determines our general lower bound strategy.
3. Lower bound

3.1. Preliminaries. We will need the following technical lemma.

Lemma 10. Let \( r \geq 2 \) be an integer, \( 0 \leq \theta \leq 2 \) fixed, and let \( \mathcal{F} \) be a family of ordered graphs with the property that if some \( (H, \pi) \), \( \pi = (e_1, \ldots, e_3) \), is in \( \mathcal{F} \), then for every subgraph \( J \subseteq H \) with \( e_1 \in J \), also \( (J, \pi|_J) \) is in \( \mathcal{F} \). Then for \( \lambda_{r,\theta}(\cdot) \) as defined in (3) we have

\[
\arg \min_{(H,\pi) \in \mathcal{F}} \lambda_{r,\theta}(H, \pi) = \arg \min_{(H,\pi) \in \mathcal{F}} \lambda_{r,\theta}(H \setminus e_1, \pi \setminus e_1) \subseteq \mathcal{F},
\]

and all ordered graphs \( (\tilde{J}, \tilde{\pi}) \), \( \tilde{\pi} = (\tilde{e}_1, \ldots, \tilde{e}_2) \), in the family \( \mathcal{F} \)

\[
\lambda_{r,\theta}(\tilde{J}, \tilde{\pi}) = 2 - \theta + r \cdot (\lambda_{r,\theta}(\tilde{J} \setminus \tilde{e}_1, \tilde{\pi} \setminus \tilde{e}_1) - 2).
\]

Proof. As the family \( \mathcal{F} \) is closed under taking subgraphs, the inner minimization on the right hand side of

\[
\min_{(H,\pi) \in \mathcal{F}} \lambda_{r,\theta}(H, \pi) \equiv \min_{(H,\pi) \in \mathcal{F}} \left\{ \lambda_{r,\theta}(H \setminus e_1, \pi \setminus e_1) - \theta + (r - 1) \cdot \min_{J \subseteq H: e_1 \in J} \left( \lambda_{r,\theta}(J \setminus e_1, \pi|_{J \setminus e_1}) - 2 \right) \right\}
\]

can be dropped. Rearranging terms yields

\[
\min_{(H,\pi) \in \mathcal{F}} \lambda_{r,\theta}(H, \pi) = \min_{(H,\pi) \in \mathcal{F}} \left\{ 2 - \theta + r \cdot (\lambda_{r,\theta}(H \setminus e_1, \pi \setminus e_1) - 2) \right\}
\]

\[
= 2 - \theta + r \cdot \left( \min_{(H,\pi) \in \mathcal{F}} \lambda_{r,\theta}(H \setminus e_1, \pi \setminus e_1) - 2 \right),
\]

where the minimum is attained by the same graphs on both sides. Both (16) and (17) follow. \( \square \)

3.2. Lower bound proof. We give a more formal description of the edge selection strategy outlined in the introduction. Let \( r \geq 2 \) and \( 0 \leq \theta \leq 2 \) be arbitrary but fixed. By saying that an edge \( f \in E_N \) closes a copy of \( (H, \pi) \in \mathcal{S}(F) \) we mean that \( f \) completes a copy of \( (H \setminus e_1, \pi \setminus e_1) \) in \( G_{N-1} \) to a copy of \( (H, \pi) \) in \( G_N \) (where \( e_1 \) is the first edge of \( \pi \)). In step \( N \) of the Achlioptas process, we calculate for each edge \( f \in E_N \) the value

\[
d(f) := \min \left\{ \lambda_{r,\theta}(H, \pi) \mid f \text{ closes a copy of } (H, \pi) \in \mathcal{S}(F) \right\},
\]

and select \( f^N \) as the edge for which this value is maximal.

Ties are broken according to the following rule: Consider the directed graph \( \mathcal{G} = \mathcal{G}(F) \) with vertex set \( \mathcal{S}(F) \) and arcs given by proper (ordered) subgraph inclusion; i.e., from every vertex \( (H, \pi) \) there are arcs to all vertices \( (J, \pi|_J) \) with \( J \subseteq H \). Clearly, \( \mathcal{G} \) contains no directed cycles. We extend \( \mathcal{G} \) to a graph \( \mathcal{G}' = \mathcal{G}'(F, r, \theta) \) by first connecting every pair of distinct vertices \( (H_1, \pi_1), (H_2, \pi_2) \) for which \( \lambda_{r,\theta}(H_1, \pi_1) = \lambda_{r,\theta}(H_2, \pi_2) \) with an (undirected) edge, and then orienting these additional edges in such a way that the directed graph \( \mathcal{G}' \) remains acyclic. (It is easy to see that this is always possible.) Note that for every fixed \( \lambda_0 \in \mathbb{R} \) this yields a total ordering on all graphs \( (H, \pi) \) with \( \lambda_{r,\theta}(H, \pi) = \lambda_0 \). We say that \( (H_1, \pi_1) \) is higher than \( (H_2, \pi_2) \) in this ordering if the corresponding arc in \( \mathcal{G}' \) is directed from \( (H_1, \pi_1) \) to \( (H_2, \pi_2) \). Our strategy breaks ties according to this ordering: Whenever we have the choice between different edges with the same value \( d(f) \), then for each such edge \( f \) we consider the set of ordered graphs

\[
\mathcal{J}(f) := \left\{ \left( (H, \pi) \in \mathcal{S}(F) \mid f \text{ closes a copy of } (H, \pi) \wedge \lambda_{r,\theta}(H, \pi) = d(f) \right) \right\},
\]

and, among these, we let \( J(f) \in \mathcal{J}(f) \) denote the graph which is lowest in the total ordering for \( \lambda_0 := d(f) \). Then we select the edge \( f \) for which \( J(f) \) is highest in the total ordering for \( \lambda_0 \).

A careful analysis of this strategy along the lines of our informal considerations in Section 2 yields the following lemma. Note that its statement is purely deterministic and holds even if the \( r \) edges presented in each step of the Achlioptas process are selected by an adversary.
Lemma 11. Let $r \geq 2$ be an integer and $0 \leq \theta \leq 2$ be fixed. Following the above edge selection strategy ensures that the following invariant is maintained throughout for some $v_{\text{max}} = v_{\text{max}}(F, r, \theta)$: The graph $G^r(n, N)$ contains an $r$-matched subgraph $K'$ with $v(K') \leq v_{\text{max}}$ and

$$
\mu_{r, \theta}(K') < 0 \quad (20)
$$

or for every $(H, \pi) \in S(F)$ we have that every copy of $(H, \pi)$ in $G_N$ is contained in an $r$-matched subgraph $H'$ of $G^r(n, N)$ with $v(H') \leq v_{\text{max}}$ and

$$
\mu_{r, \theta}(H') \leq \lambda_{r, \theta}(H, \pi) \quad (21)
$$

where $\lambda_{r, \theta}()$ and $\mu_{r, \theta}()` are defined in (3) and (14), respectively.

With Lemma 11 in hand, the proof of the lower bound in Theorem 3 is straightforward.

Proof of Theorem 3 (lower bound). Let $\theta^* = \theta^*(F, r)$ be defined as in the theorem. We show that the above edge selection strategy with $\theta := \theta^*$ a.a.s. avoids $F$ as long as $N \ll N_0(F, r, n) = n^{2-\theta^*}$.

By the definition of $\theta^*$ (cf. (4) and (5)), for each possible ordering $\pi$ of the edges of $F$ there exists an ordered subgraph $(H, \sigma), \sigma := \pi|_H$, with $\lambda_{r, \theta^*}(H, \sigma) \leq 0$. According to Lemma 11, the following holds for each such $(H, \sigma)$: If $G_N$ contains a copy of $(H, \sigma)$, then $G^r(n, N)$ contains an $r$-matched graph $K'$ with $v(K') \leq v_{\text{max}}$ satisfying (20), or an $r$-matched graph $H'$ with $v(H') \leq v_{\text{max}}$ satisfying

$$
\mu_{r, \theta^*}(H') \leq \lambda_{r, \theta^*}(H, \sigma) \leq 0 \quad (21)
$$

This yields a family $\mathcal{W} = \mathcal{W}(F, \pi, r)$ of $r$-matched graphs $W'$ satisfying $\mu_{r, \theta^*}(W') \leq 0$ and $v(W') \leq v_{\text{max}}$ such that, deterministically, $G^r(n, N)$ contains a graph from $\mathcal{W}$ if $G_N$ contains a copy of $(F, \pi)$ (and hence also a copy of $(H, \sigma)$). It follows that $G^r(n, N)$ contains a graph from $\mathcal{W}^* = \mathcal{W}^*(F, r) := \bigcup_{\pi \in \Pi(E(F))} \mathcal{W}(F, \pi, r)$ if $G_N$ contains a copy of $F$. Moreover, since no graph in $\mathcal{W}^*$ has more than $v_{\text{max}}(F, r, \theta^*(F, r))$ vertices, the size of $\mathcal{W}^*$ is bounded by a constant only depending on $F$ and $r$.

Thus by Lemma 6, the definition of $\mu_{r, \theta^*}()$ in (14), and the fact that $\mu_{r, \theta^*}(W') \leq 0$ for all $W' \in \mathcal{W}^*$, the expected number of copies of graphs from $\mathcal{W}^*$ in $G^r(n, N)$ with $N \ll n^{2-\theta^*}$ is of order

$$
\sum_{W' \in \mathcal{W}^*} n^{v(W')}(Nn^{-2r})^{s(W')} \ll \sum_{W' \in \mathcal{W}^*} n^{\mu_{r, \theta^*}(W')} \leq |\mathcal{W}^*| \cdot n^0 = \Theta(1) .
$$

It follows with Markov’s inequality that a.a.s. $G^r(n, N)$ contains no graph from $\mathcal{W}^*$. Consequently, a.a.s. $G_N$ contains no copy of $F$.

It remains to prove Lemma 11. It is not too hard to prove by induction on $N$ that our strategy and the recursive definition of $\lambda_{r, \theta}(H, \pi)$ in (3) guarantee the existence of $K'$ satisfying (20) or $H'$ satisfying (21) (cf. (9) below). The more difficult part is to guarantee that $K'$ or $H'$ does not become arbitrarily large. This requires some rather technical work, and is also the point where the tie-breaking rule introduced above comes into play.

Proof of Lemma 11. To simplify notation, we drop subscripts and write $\lambda = \lambda_{r, \theta}$ and $\mu = \mu_{r, \theta}$ in the following. For the reader’s convenience, Figure 5 illustrates the notations used throughout the proof. For the first part of the argument, only the top part of Figure 5 is relevant.

Let

$$
\varepsilon = \varepsilon(F, r, \theta) := \min \{ |\lambda(H_1, \pi_1) - \lambda(H_2, \pi_2)| \mid (H_1, \pi_1), (H_2, \pi_2) \in S(F) \wedge \lambda(H_1, \pi_1) \neq \lambda(H_2, \pi_2) \} > 0 \quad (22)
$$

and

$$
v_{\text{max}} = v_{\text{max}}(F, r, \theta) := r^{v(F)/\varepsilon + 1}|S(F)|+2v(F) . \quad (23)
$$
We proceed by induction on \( N \), showing that the statement about graphs \((H, \pi) \in S(F)\) is true for as long as \( G'(n, N) \) does not contain an \( r \)-matched subgraph \( K' \) with \( v(K') \leq v_{\max} \) satisfying (20). Note that the statement holds trivially if \( H \) has no edges: For any copy of \((H, \pi)\) in \( G_N \) we define \( H' \) to be exactly this copy, yielding \( \mu(H') = v(H) = \lambda(H, \pi) \) and \( v(H') = v(H) \leq v_{\max} \). This also takes care of the induction base \( N = 0 \).

Let \( E^N = \{ f_1, \ldots, f_r \} \), and assume w.l.o.g. that \( f_1 \) is the edge \( f_N \) that gets selected by our strategy. Furthermore, fix some ordered graph \((H, \pi) \in S(F)\) with \( e(H) \geq 1 \), and assume that \( f_1 \) completes a copy of \((H \setminus \bar{e}_1, \pi \setminus \bar{e}_1)\) to a copy of \((H, \pi)\) in \( G_N \) (where \( \bar{e}_1 \) is the first edge of \( \pi \)). By induction, this copy is contained in an \( r \)-matched graph \( H'_1 = (V_1, K_1) \) in \( G''(n, N - 1) \) with

\[
\mu(H'_1) \leq \lambda(H \setminus \bar{e}_1, \pi \setminus \bar{e}_1)
\]

and

\[
v(H'_1) \leq v_{\max}.
\]

By the definition of our strategy, the edges \( f_i \) would have closed copies of some ordered graphs \((J_i, \pi_i) \in S(F)\) satisfying

\[
\lambda(J_i, \pi_i) = d(f_i), \quad 1 \leq i \leq r,
\]

Figure 5. Notations used in the proof of Lemma 11. The arcs of \( T(H') \) drawn grey are either grey or red in the proof.
Let \( \hat{\mathcal{F}} \) be some graph from \( \arg \min_{J \subseteq H \setminus \hat{e}_i} \lambda(J \setminus \hat{e}_1, \pi_{|J \setminus \hat{e}_1}) \), and note that \( f_1 \) also completed a copy of \((\hat{\mathcal{J}} \setminus \hat{e}_1, \pi_{|\hat{\mathcal{J}} \setminus \hat{e}_1})\) to a copy of \((\hat{\mathcal{J}}, \pi_{|\hat{\mathcal{J}}})\), implying that

\[
\lambda(J_1, \pi_1) \leq \lambda(\hat{\mathcal{J}}, \pi_{|\hat{\mathcal{J}}})
\]  
(28)

(note that \((J_1, \pi_1)\) and \((\hat{\mathcal{J}}, \pi_{|\hat{\mathcal{J}}})\) might actually be the same graph). Combining (26), (27) and (28) yields

\[
\lambda(J_i, \pi_i) \leq \lambda(J_1, \pi_1) \leq \lambda(\hat{\mathcal{J}}, \pi_{|\hat{\mathcal{J}}}) , \quad 2 \leq i \leq r .
\]  
(29)

Applying Lemma 10 to the families \( \hat{\mathcal{F}} = \{(J, \sigma) \in \mathcal{S}(F) \mid f_i \) closes a copy of \((J, \sigma)\}, 1 \leq i \leq r, \) and \( \hat{\mathcal{F}} = \{(J, \pi_{|J}) \mid J \subseteq H \land \hat{e}_1 \in J\} \) yields that the transformation (17) holds for both \((J_i, \pi_i)\), 1 \leq i \leq r, and \((\hat{\mathcal{J}}, \pi_{|\hat{\mathcal{J}}})\). Thus it follows from inequality (29) that

\[
\lambda(J_i \setminus e_1, \pi_i \setminus e_1) \leq \lambda(\hat{\mathcal{J}} \setminus \hat{e}_1, \pi_{|\hat{\mathcal{J}} \setminus \hat{e}_1}) , \quad 2 \leq i \leq r .
\]  
(30)

(With slight abuse of notation, we write \((J_i \setminus e_1, \pi_i \setminus e_1)\) and tacitly assume that the variable \(e_1\) represents the first edge of \(\pi_i\), adapting to the context.) By induction, the copies of the graphs \((J_i \setminus e_1, \pi_i \setminus e_1)\) that were completed by the edges \(f_i\) to copies of \((J_i, \pi_i)\), 2 \leq i \leq r, are contained in \(r\)-matched graphs \(J'_i = (V_i, K_i)\) in \(G^r(n, N - 1)\) with

\[
\mu(J'_i) \leq \lambda(J_i \setminus e_1, \pi_i \setminus e_1) \leq \lambda(\hat{\mathcal{J}} \setminus \hat{e}_1, \pi_{|\hat{\mathcal{J}} \setminus \hat{e}_1}) , \quad 2 \leq i \leq r ,
\]  
(31)

and

\[
v(J'_i) \leq v_{\text{max}} , \quad 2 \leq i \leq r .
\]  
(32)

Recall that \(H'_1 = (V_1, K_1)\) is the \(r\)-matched graph containing the copy of \((H \setminus \hat{e}_1, \pi \setminus \hat{e}_1)\) that was completed by \(f_1 = f^N\) to the copy of \((H, \pi) \in \mathcal{S}(F)\) we are considering. If \(\mu(H'_1) < 0\) or \(\mu(J'_i) < 0\) for some \(2 \leq i \leq r\), then we have found an \(r\)-matched graph \(K'\) with \(\mu(K') < 0\) and \(v(K') \leq v_{\text{max}}\) (cf. (25) and (32)). Otherwise we have \(\mu(H'_1) \geq 0\) and \(\mu(J'_i) \geq 0\) for all \(2 \leq i \leq r\). We will argue later that this implies even stronger bounds on the number of vertices of \(H'_1\) and \(J'_i\), namely

\[
v(H'_1) \leq v_{\text{max}}/r ,
\]

\[
v(J'_i) \leq v_{\text{max}}/r , \quad 2 \leq i \leq r .
\]  
(33)

We define the \(r\)-matched graph

\[
H' := \left( \bigcup_{1 \leq i \leq r} V_i, \{E^N\} \cup \bigcup_{1 \leq i \leq r} K_i \right) .
\]  
(34)

Furthermore, we define for \(2 \leq i \leq r\) the \(r\)-matched graphs

\[
K'_i := \left( V_i \cap \bigcup_{1 \leq j \leq i - 1} V_j, K_i \cap \bigcup_{1 \leq j \leq i - 1} K_j \right) .
\]  
(35)

Standard inductive arguments yield

\[
v(H') = v(H'_1) + \sum_{i=2}^{r} v(J'_i) - \sum_{i=2}^{r} v(K'_i) ,
\]  
(36)

\[
\kappa(H') = \kappa(H'_1) + \sum_{i=2}^{r} \kappa(J'_i) - \sum_{i=2}^{r} \kappa(K'_i) + 1 .
\]  
(37)
From (33) and (35) we conclude that \( v(K'_i) \leq v_{\max}/r \leq v_{\max} \), \( 2 \leq i \leq r \). Therefore, if \( \mu(K'_i) < 0 \) for some \( 2 \leq i \leq r \), then we have found an \( r \)-matched graph \( K' \) with \( \mu(K') < 0 \) and \( v(K') \leq v_{\max} \). Otherwise we have

\[
\mu(K'_i) \geq 0 \ , \quad 2 \leq i \leq r \ .
\]

Combining our previous observations we obtain that

\[
\mu(H') = \mu(H'_1) + \sum_{i=2}^{r} \mu(J'_i) - \sum_{i=2}^{r} \mu(K'_i) - 2r + (2 - \theta)
\]

\[
\leq \lambda(H \setminus \bar{e}_1, \pi \setminus \bar{e}_1) - \theta + (r - 1) \cdot (\lambda(\hat{J} \setminus \bar{e}_1, \pi | \hat{J} \setminus \bar{e}_1) - 2)
\]

\[
\leq \lambda(H, \pi) ,
\]

which proves (21). From (33) and (34) we conclude that \( v(H') \leq v_{\max} \).

It remains to show (33), i.e. that for every \( r \)-matched graph \( H' \) as defined in (34) with \( \mu(H') \geq 0 \) we have \( v(H') \leq v_{\max}/r \).

In the above argument we constructed the \( r \)-matched graph \( H' \) containing the copy of \((H, \pi)\) inductively from the \( r \)-matched graph \( H'_1 \) containing the copy of \((H \setminus \bar{e}_1, \pi \setminus \bar{e}_1)\) and the \( r \)-matched graphs \( J'_i \) containing the copies of \((J_i \setminus \bar{e}_1, \pi_i \setminus \bar{e}_1)\), \( 2 \leq i \leq r \). We associate this inductive construction with an edge-colored directed rooted tree \( T(H') \) as follows (cf. Figure 5): The vertices of \( T(H') \) correspond to copies of graphs from \( S(F) \) in \( G_N \) (the same copy may appear as a vertex multiple times). If \( H \) has no edges (recall that in this case \( H' \) is empty as well), \( T(H') \) consists only of this copy of \((H, \pi)\) as an isolated vertex. Otherwise \( T(H') \) consists of the copy of \((H, \pi)\) as the root vertex joined to \( r \) subtrees, \( T(H'_1) \) and \( T(J'_1) \) for all \( 2 \leq i \leq r \), where \( T(H'_1) \) is connected to the root by a black arc, and every \( T(J'_1) \) is connected to the root by an arc that is either grey or red according to the following criterion: Each such arc corresponds to an instance of the inequalities in (29) somewhere along the induction. The arc is grey if both these inequalities are tight, i.e., \( \lambda(J_i, \pi_i) = \lambda(\hat{J}, \pi | \hat{J}) \), and red if at least one of them is strict, i.e., \( \lambda(J_i, \pi_i) < \lambda(\hat{J}, \pi | \hat{J}) \). All arcs are oriented away from the root. Note that the tree \( T(H') \) captures only the logical structure of the inductive history of \( H' \). Overlaps (captured by the graphs \( K'_i \), \( 2 \leq i \leq r \)) are completely neglected.

Every red arc of \( T(H') \) corresponds to a strict inequality in (29). In this case, as a consequence of (17), inequality (30) (and thus also (31)) is strict as well, with a difference of at least \( \varepsilon \) between the right and the left hand side (cf. (22)). Consequently, each red arc contributes a term of \(-\varepsilon\) to the right hand side of (39) in the corresponding induction step. Accumulating these terms along the induction yields that

\[
\mu(H') \leq \lambda(H, \pi) - \ell(H') \cdot \varepsilon ,
\]

where \( \ell(H') \) denotes the number of red arcs in \( T(H') \).

Note that \( \lambda(H, \pi) \leq v(F) \) for all \((H, \pi) \in S(F)\). Thus if \( \mu(H') \geq 0 \), then by (40) the tree \( T(H') \) has at most \( \lambda(H, \pi)/\varepsilon \leq v(F)/\varepsilon \) many red arcs. We will show that, due to our tie-breaking rule involving the auxiliary graph \( G' \), this bound on the number of red arcs implies the claimed bound of \( v_{\max}/r \) on the number of vertices of \( H' \). To that end, we first show that if two vertices of \( T(H_1) \) are connected by a (directed, i.e. descending) path \( P \) that contains no red arcs, then these two vertices correspond to copies of different ordered graphs \((H_1, \pi_1), (H_2, \pi_2) \in S(F)\).

Observe that if both inequalities in (29) are tight for some \( 2 \leq i \leq r \) (i.e., the corresponding arc is grey), then by our tie-breaking rule there is an arc from \((\hat{J}, \pi | \hat{J})\) to \((J_1, \pi_1)\) and an arc from \((J_1, \pi_1)\) to \((J_i, \pi_i)\) in the auxiliary graph \( G' \) (unless \((\hat{J}, \pi | \hat{J}) = (J_1, \pi_1) \) or \((J_1, \pi_1) = (J_i, \pi_i) \) of course). Here
we assume w.l.o.g. that for $1 \leq i \leq r$, $(J_i, \pi_i)$ is lowest among all graphs in $\mathcal{J}(f_i)$ with respect to the total ordering given by $\mathcal{G}'$, cf. [19].

Using this observation we can associate the path $P$ with a directed walk $P'$ (of possibly different length) in $\mathcal{G}'$ as follows: The initial vertex of $P'$ is $(H_1, \pi_1)$. For each black arc on $P$ from a copy of some $(H, \pi) \in \mathcal{S}(F)$ to a copy of $(H \setminus e_1, \pi \setminus e_1)$ we extend $P'$ by an arc from $(H, \pi)$ to $(H \setminus e_1, \pi \setminus e_1)$. For each grey arc on $P$ from a copy of some $(H, \pi) \in \mathcal{S}(F)$ to a copy of some $(J_i \setminus e_1, \pi_i \setminus e_1) \in \mathcal{S}(F)$ for some $2 \leq i \leq r$ we extend $P'$ by up to four arcs, from $(H, \pi)$ to $(\hat{J}, \pi|_J)$, from there to $(J_i, \pi_i)$, from there to $(J_i \setminus e_1, \pi_i \setminus e_1)$ (if some of these graphs happen to be the same, then the extension is less than four arcs long). By the definition of the graph $\mathcal{G}'$ and our tie-breaking rule, all these arcs exist in $\mathcal{G}'$. Proceeding in this manner we obtain a directed walk $P'$ in $\mathcal{G}'$ from $(H_1, \pi_1)$ to $(H_2, \pi_2)$. As $\mathcal{G}'$ is acyclic we must have $(H_1, \pi_1) \neq (H_2, \pi_2)$.

It follows that a (directed) path in $\mathcal{T}(H')$ that contains no red arcs has at most $|\mathcal{S}(F)|$ vertices. Since in total there are at most $v(F)/\varepsilon$ many red arcs in $\mathcal{T}(H')$, it follows that the depth of $\mathcal{T}(H')$ is bounded by

$$v(F)/\varepsilon + 1)^{|\mathcal{S}(F)|},$$

and that consequently

$$v(\mathcal{T}(H')) \leq 1 + r + r^2 + \cdots + r^{(v(F)/\varepsilon+1)|\mathcal{S}(F)|} \leq r^{(v(F)/\varepsilon+1)|\mathcal{S}(F)|+1}.$$ 

Observing that every vertex of $\mathcal{T}(H')$ corresponds to at most $v(F)$ vertices of $H'$, we finally obtain that

$$v(H') \leq r^{(v(F)/\varepsilon+1)|\mathcal{S}(F)|+1} \cdot v(F) \overset{[33]}{=} v_{\max}/r.$$ 

This justifies [33] and concludes the proof. 

\[ \square \]

4. Upper bound

4.1. Preliminaries. A grey-black $r$-matched graph is a triple $H = (V, K, B)$ such that $(V, K)$ is an $r$-matched graph, and $B$ is an edge set containing exactly one edge from every $r$-set in $K$. The $|K|$ many edges in $B$ are considered black, and the remaining $|K| \cdot (r - 1)$ edges are considered grey. For any grey-black $r$-matched graph $H$ we use $V(H)$, $K(H)$ and $B(H)$ to denote the set of vertices, the family of $r$-sets and the set of black edges of $H$, respectively. We sometimes ignore the coloring and tacitly identify $H$ with the underlying $r$-matched graph $(V(H), K(H))$.

We may naturally interpret the edge that is selected in each step of the Achlioptas process as being colored black, and the $r - 1$ edges that are discarded as being colored grey. More formally, we denote by $\mathcal{G}_N$ the grey-black $r$-matched graph $(V(K_n), \{E^1, \ldots, E^N\}, \{f^1, \ldots, f^N\})$, i.e., the grey-black version of $\mathcal{G}'(n, N)$ in which exactly the edges of $G_N$ are colored black.

For any natural number $t$ and any (grey-black $r$-matched) graph $H$ we denote by $t \cdot H$ the disjoint union of $t$ copies of $H$. The following definition is illustrated in Figure 6 for the case $r = 2$.

**Definition 12.** For any integer $r \geq 2$ and any ordered graph $(F, \pi)$, $\pi = (e_1, \ldots, e_f)$, we define the grey-black $r$-matched graph $F_r^\pi$, and a distinguished black copy of the (ordered) graph $(F, \pi)$ in $F_r^\pi$, referred to as the central copy of $(F, \pi)$ in $F_r^\pi$, recursively as follows:

- If $F$ has no edges, we set $F_r^\pi := F$ and call this the central copy of $(F, \pi)$ in $F_r^\pi$.
- Otherwise, $F_r^\pi$ is obtained as follows: Let $H_i = (V_i, K_i, B_i)$, $1 \leq i \leq r$, be $r$ disjoint copies of $(F \setminus e_1)_{r \setminus e_1}$, and let $K = \{f_1, \ldots, f_r\}$ be an $r$-set of edges such that for all $1 \leq i \leq r$ the edge

\[ e_i \]

in $H_i$ is colored black, and in $F_r^\pi$ it is colored black if it occurs in $F \setminus e_1$, and in $F \setminus e_1$, and let $K = \{f_1, \ldots, f_r\}$ be an $r$-set of edges such that for all $1 \leq i \leq r$ the edge
$K = \{f_1, f_2\}$

$F_{\pi}^{r}$

$H_1$ $H_2$

Figure 6. Construction of $F_{\pi}^{r}$ for the case $r = 2$. We use the abbreviation $F_{i-} := F\{e_1, \ldots, e_i\}$, where $\pi = (e_1, \ldots, e_f)$. A label of $F_{i-}$ indicates that the corresponding black edges form an (ordered) copy of $F_{i-}$. Note that each of the two copies $H_1, H_2$ of $(F_{i-})_{\pi}^{r} e_1 = (F_{i-})_{\pi}^{r} f_1^{-}$ (indicated by dashed lines) is formed by two copies of $(F_{i-})_{\pi}^{r} e_2^{-}$ (indicated by dotted lines), and so on.

$f_i$ completes the central copy of $(F_{i-} e_1, \pi e_1)$ in $H_i$ to a copy of $(F, \pi)$. We define

$$V(F_{\pi}^{r}) := \bigcup_{1 \leq i \leq r} V(H_i) ,$$

$$K(F_{\pi}^{r}) := \{K\} \cup \bigcup_{1 \leq i \leq r} K(H_i) ,$$

$$B(F_{\pi}^{r}) := \{f_1\} \cup \bigcup_{1 \leq i \leq r} B(H_i)$$

and define the central copy of $(F, \pi)$ in $F_{\pi}^{r}$ to be the copy that is formed by the central copy of $(F_{i-} e_1, \pi e_1)$ in $H_1$ and the black edge $f_1$.

We refer to the $r$-set $K$ added in the recursive step as the central $r$-set in $F_{\pi}^{r}$.

Note that even though ordered graphs are used in the above construction, $F_{\pi}^{r}$ is considered an unordered (grey-black $r$-matched) graph.

As it turns out, the threshold given by Theorem 3 and Theorem 4 is in fact the threshold for the appearance of the $r$-matched graph $F_{\pi}^{r}$ corresponding to the ‘most dangerous’ $\pi \in \Pi(E(F))$ in $G_{n,m}^{r}$ (or $G^{r}(n, N)$), as given by Theorem 8. In other words, we have

$$m^{r*}(F) = \min_{\pi \in \Pi(E(F))} m(F_{\pi}^{r}) .$$

In essence, our upper bound proof relies on the same variance calculations that are used in the proof of Theorem 8 (or Theorem 1). However, in order to be able to control which edges will be
colored black during the process, we shall do this variance calculation in \( e(F) \) rounds, each round corresponding to one step of the recursive definition of \( F^\pi_r \).

We shall show that if the number of steps \( N \) in the Achlioptas process is such that \( N \gg n^{2-\theta'} \), where \( \theta' = \theta'(F, \pi, r) \) is defined in (41) below, then a.a.s. a copy of the ordered graph \( (F, \pi) \) will be created in \( G_N \), regardless of the edge selection strategy employed. The upper bound in Theorem 3 then follows immediately by considering an optimal edge ordering \( \pi \in \Pi(E(F)) \), i.e., one that yields the lowest possible upper bound (cf. (4)). The next lemma essentially states that for \( N \gg n^{2-\theta'} \), the expected number of copies of any subgraph \( J \subseteq F^\pi_r \) in \( G'(n, N) \) is \( \omega(1) \). This is exactly what is needed for the mentioned variance calculation to work out. The proof of Lemma 13 is quite technical and deferred to Section 5.

**Lemma 13.** Let \( r \geq 2 \) be an integer, and let \( (F, \pi) \) be a nonempty ordered graph. Let \( F^\pi_r \) be as in Definition 12 and let \( \theta' = \theta'(F, \pi, r) \) be the unique solution of

\[
\min_{H \subseteq F} \lambda_{r,\theta'}(H, \pi|_H) \geq 0 ,
\]

where \( \lambda_{r,\theta'}() \) is defined in (3). Then every \( r \)-matched subgraph \( J \subseteq F^\pi_r \) satisfies

\[
\mu_{r,\theta'}(J) \geq 0 ,
\]

where \( \mu_{r,\theta'}() \) is defined in (14).

4.2. **Upper bound proof.** We extend the notion of induced subgraph containment to \( r \)-matched graphs. Let \( G \) be an \( r \)-matched graph and \( H \) a subgraph of \( G \). We say that \( H \) is an induced subgraph of \( G \) if the two endvertices of any edge in any \( r \)-set in \( K(G) \setminus K(H) \) are not both in \( V(H) \) (i.e., if the underlying unmatched graph of \( H \) is an induced subgraph of the underlying unmatched graph of \( G \)).

The upper bound in Theorem 3 is a straightforward consequence of the next lemma. Recall that \( \tilde{G}_N \) denotes the grey-black version of \( G'(n, N) \) in which exactly the edges of \( G_N \) are colored black.

**Lemma 14.** Let \( r \geq 2 \) be an integer, and let \( (F, \pi) \) be a nonempty ordered graph. Let \( t \geq 1 \) be an integer, and let \( F^\pi_r := t \cdot F^\pi_r \), where \( F^\pi_r \) is as in Definition 12. If \( n^2 \gg N \gg n^{2-\theta'} \), where \( \theta' = \theta'(F, \pi, r) \) is the unique solution of

\[
\min_{H \subseteq F} \lambda_{r,\theta'}(H, \pi|_H) \geq 0 ,
\]

and \( \lambda_{r,\theta'}() \) is defined in (3), then a.a.s. the number of induced copies of \( F^\pi_r \) in \( \tilde{G}_N \) is of order

\[
n^{\nu(F^\pi_r)}(Nn^{-2r})^{\kappa(F^\pi_r)} ,
\]

regardless of the edge selection strategy employed.

Note that the order of magnitude of the number of induced copies of \( F^\pi_r \) guaranteed by Lemma 14 is the order of magnitude of the expected number of (uncolored) copies of \( F^\pi_r \) in \( G'(n, N) \) and hence best possible, cf. Lemma 9.

**Proof of Theorem 3 (upper bound).** We show that if \( N \gg N_0(F, r, n) = n^{2-\theta^*} \), then a.a.s. \( G_N \) contains a black copy of \( F \), regardless of the edge selection strategy employed. Let \( \pi \in \Pi(E(F)) \) be an edge ordering that maximizes the right hand side of (41) for \( \theta = \theta^* \), such that \( \theta^*(F, r) = \theta'(F, \pi, r) \) for \( \theta' \) as in Lemma 13 and Lemma 14. By Lemma 14 (applied with \( t = 1 \)), the definition of \( \mu_{r,\theta^*}() \) in (14), and Lemma 13 a.a.s. the number of (induced) copies of \( F^\pi_r \) in \( \tilde{G}_N \) with \( n^2 \gg N \gg n^{2-\theta^*} \) is of order

\[
n^{\nu(F^\pi_r)}(Nn^{-2r})^{\kappa(F^\pi_r)} \gg n^{\mu_{r,\theta^*}(F^\pi_r)} \geq 1 .
\]
Thus $G_N$ contains at least one black copy of $F$ (the central copy of $(F, \pi)$ in any copy of $F^\pi$), regardless of the edge selection strategy employed.

It remains to prove Lemma 14. For the rest of this section, we assume that $r \geq 2$ is fixed, and drop the corresponding subscript from $F^\pi$. Moreover, we abbreviate $F \setminus e_1$ by $F_-$, $\pi \setminus e_1$ by $\pi_-$, and $(F \setminus e_1)^{\pi \setminus e_1}$ by $F^\pi_-$.

The proof of Lemma 14 proceeds by induction on $e(F)$. Essentially, the induction step consists in proving that a.a.s. the right number of copies of $r \cdot F^\pi$ evolve into copies of $F^\pi$, i.e., that the right number of $r$-sets presented in the process are such that all $r$ edges complete the central copy of $(F_-, \pi_-)$ in some copy of $F^\pi$ to a copy of $(F, \pi)$. By Definition 12 and the fact that one of these edges needs to be colored black in $\tilde{G}_N$, this creates a copy of $F$ regardless of the edge selection strategy used.

To ensure that we have enough disjoint copies of $F^\pi$ in the next induction step, we carry out this argument $t$ times in parallel, showing that the right number of copies of $F^\pi := tr \cdot F^\pi = t \cdot (r \cdot F^\pi)$ evolve into copies of $F^\pi := t \cdot F^\pi$. Furthermore, we need these copies to be induced since by our definition of the process, every edge gets presented at most once, so edges that appear ‘too early’ might spoil our argument.

Proof of Lemma 14. We proceed by induction on $e(F)$. Even though the preconditions of the lemma exclude graphs with no edges, the conclusion of the lemma also holds in this case: If $F$ is empty, then $\tilde{G}_N$ contains $\Theta(\sqrt{(n^r)}) = \Theta(n^{e(F)})$ induced copies of $t \cdot F^\pi = t \cdot F$ for any $n^2 \gg N \geq 0$. This serves as our induction basis. For the induction step we employ a two-round approach. That is, we divide the process into two rounds of equal length $N/2$ and analyze these two rounds separately. Specifically, we apply the induction hypothesis and some known facts about $\tilde{G}^\pi_{n,m}$ to the first round and then show by a variance calculation that, conditional on a ‘good’ first round, in the second round the claimed number of copies of $F^\pi$ are created.

Recall that $F^\pi = tr \cdot F^\pi_-$, and note that

$$v(F^\pi) = v(F^\pi_-), \quad \kappa(F^\pi) = \kappa(F^\pi_-) + t$$

(cf. Definition 12). Let $M$ denote the number of induced copies of $F^\pi$ in $\tilde{G}_{N/2}$. Due to $\theta'(F_-, \pi_-, r) \leq \theta'(F, \pi, r)$ (recall that $\lambda_{r,\theta}(H)$ is decreasing in $\theta$ for $r$ and $H$ fixed), the induction hypothesis is applicable and yields (with $t \leftarrow tr$) that a.a.s.

$$M \asymp n^{v(F^\pi)}(N n^{-2r})^{\kappa(F^\pi)}.$$  \hspace{1cm} (43)

We label these copies $F^\pi_{-i}$, $1 \leq i \leq M$. For a given copy $F^\pi_{-i}$, consider the $tr$ copies of $F^\pi_-$, and, for each of these, fix a non-edge that completes the central copy of $(F_-, \pi_-)$ to a copy of $(F, \pi)$. Call the set of all these non-edges $T_i$. Furthermore, fix an arbitrary partition $K_i$ of $T_i$ into sets of size exactly $r$. Thus for all $i$ we have $|T_i| = tr$ and $|K_i| = t$.

For $1 \leq i \leq M$, let $Z_i$ be the indicator variable for the event that the $t$ many $r$-sets of $K_i$ are revealed during the second round of the Achlioptas process, and that no other edge of $(\binom{V}{2})$ appears in the second round. Let

$$Z := \sum_{i=1}^{M} Z_i.$$  \hspace{1cm} (44)

We shall prove by the methods of first and second moment that a.a.s.

$$Z \asymp n^{v(F^\pi)}(N n^{-2r})^{\kappa(F^\pi)}.$$  \hspace{1cm} (45)

Note that this implies that a.a.s. this many copies of $F^\pi$ evolve into induced copies of $F^\pi$, regardless of the edge selection strategy used.
The probability that all edges of $T_i$ are present is of order $(Nn^{-2})^{tr}$, the probability that they are partitioned as required is of order $(N^{-r}n)^t$ (cf. the proof of Lemma 9), and, due to the assumption $N \ll n^2$, the requirement that no other edges of $(V(F_i))$ appear contributes only a factor of $\Theta(1)$. Thus we have

$$\Pr[Z_i = 1] \asymp (Nn^{-2})^{tr} \cdot (N^{-r}n)^t = (Nn^{-2r})^t$$

and, conditioning on the first round satisfying (44),

$$E[Z] \asymp M \cdot (Nn^{-2r})^t \cdot n^{v(F_i)}(Nn^{-2r})^{\kappa(F_i)+t} \cdot n^{v(F_i)}(Nn^{-2r})^{\kappa(F_i)} \ .$$

To calculate the variance consider two copies $F_{i-1}^\pi, F_{j-1}^\pi$ and the corresponding sets $T_i, T_j$. In order for $Z_i$ and $Z_j$ to be equal to one simultaneously, we need to have that $K_i \cap K_j$ is a partition of $T_i \cap T_j$ into sets of size exactly $r$. We denote by $I \subseteq \{1, \ldots, M\}$ the set of pairs $(i, j)$ for which this is the case and for which moreover $t_{ij} := |K_i \cap K_j| \geq 1$. We obtain similarly to (46) that

$$\Pr[Z_i = 1 \land Z_j = 1] \asymp (Nn^{-2r})^{2t_{ij}} \ .$$

As for pairs with $t_{ij} = 0$ the variables $Z_i$ and $Z_j$ are negatively correlated (for any $N \ll n^2$), such pairs can be omitted, and we have

$$\text{Var}[Z] = \sum_{i,j=1}^M \left( E[Z_i Z_j] - E[Z_i] E[Z_j] \right) \lesssim \sum_{(i,j) \in I} E[Z_i Z_j] \lesssim \sum_{(i,j) \in I} (Nn^{-2r})^{2t_{ij}} \ .$$

We split this sum with respect to the type of intersection. For any $r$-matched subgraph $J \subseteq F_i$, let $0 \leq t_J \leq t$ denote its number of central $r$-sets, and let $J_- \subseteq F_i$ denote the $r$-matched graph obtained by removing these $t_J$ central $r$-sets from $J$. Note that $t_J$ is not a function of the isomorphism class of $J \subseteq F_i$, but depends on the precise position of $J$ in $F_i$. For this reason, in the following we distinguish between different subgraphs $J \subseteq F_i$ that are isomorphic to each other.

For $i = 1, \ldots, M$, let $F_i^\pi$ denote the copy of $F_i$ formed by $F_i^\pi$, and the $r$-sets of $K_i$, and for $(i, j) \in I$, let $J_{ij} \subseteq F_i^\pi$ denote the $r$-matched subgraph formed by $F_i^\pi \cap F_j^\pi$ in $F_i^\pi$. Note that $t_{ij} = t_{J_{ij}}$. For any $r$-matched subgraph $J \subseteq F_i^\pi$, let $M_J$ denote the number of pairs $(i, j) \in I$ with $J_{ij} = J$. Note that $M_J$ is bounded by some constant $C = C(F, r, t)$ times the number $M_J'$ of copies of $F_i \cup_\pi F_i^\pi$ in $G(n, N/2)$, where by $F_i \cup_\pi F_i^\pi$ we denote an arbitrary (uncolored) $r$-matched graph formed by two copies of $F_i^\pi$ that intersect in a copy of $J_-$. We only have to consider nonempty graphs $J \subseteq F_i^\pi$ due to $t_{ij} \geq 1$ for all $(i, j) \in I$. Splitting such a $J$ into $t$ disjoint subgraphs $J_1, \ldots, J_t$ such that $J_k \subseteq F_i^\pi$, $1 \leq k \leq t$, and using that $N \gg n^{2-\rho'}$, we obtain with the definition of $\mu_{r, \rho'}(\cdot)$ in (14) and Lemma 13 that

$$n^{v(J)}(Nn^{-2r})^{\kappa(J)} = \prod_{k=1}^t \left( n^{v(J_k)}(Nn^{-2r})^{\kappa(J)} \right) \gg \prod_{k=1}^t n^{\mu_{r, \rho'}(J_k)} \geq 1 \ .$$

Moreover, due to $v(J) = v(J_-)$ and $\kappa(J) = \kappa(J_-) + t_J$ we have

$$v(F_i \cup_\pi F_i^\pi) = 2v(F_i^\pi) - v(J) ,$$

$$\kappa(F_i \cup_\pi F_i^\pi) = 2\kappa(F_i^\pi) - \kappa(J) + t_J .$$

Thus by Lemma 9, the expected number of copies of $F_i \cup_\pi F_i^\pi$ in $G(n, N/2)$ is of order

$$n^{2v(F_i^\pi)-v(J)}(Nn^{-2r})^{2\kappa(F_i^\pi)-\kappa(J)+t_J} \lesssim n^{2v(F_i^\pi)}(Nn^{-2r})^{2\kappa(F_i^\pi)+t_J} ,$$

and Markov’s inequality implies that

$$M_J' \ll n^{2v(F_i^\pi)}(Nn^{-2r})^{2\kappa(F_i^\pi)+t_J} \ .$$
a.a.s. As moreover the number of $r$-matched subgraphs $J \subseteq F^\pi$ is a constant depending only on $F$, $r$ and $t$, a.a.s. (51) holds for all $J$ simultaneously.

Conditioning on the first round satisfying (44) and (51) for all $J \subseteq F^\pi$, we may continue (49) as follows:

$$
\text{Var}[Z] \leq \sum_{J \subseteq F^\pi} \sum_{(i,j) \in I: J_{ij} = J} (Nn^{-2r})^{2t - t_j} = \sum_{J \subseteq F^\pi: \sum_{i,j} (i,j) \in I : J = J} M_J \cdot (Nn^{-2r})^{2t - t_J}
$$

$$
\leq \sum_{J \subseteq F^\pi: \sum_{i,j} (i,j) \in I : J = J} CM_J' \cdot (Nn^{-2r})^{2t - t_J}
$$

\[\leq \sum_{J \subseteq F^\pi: \sum_{i,j} (i,j) \in I : J = J} \left( n^{v(F^\pi)} (Nn^{-2r})^{\kappa(F^\pi) + t} \right)^2 \leq 2 \left(3 \right)<E[Z]^2.
\]

Chebyshev’s inequality now yields that a.a.s. the second round satisfies (15), which, as discussed, implies that there is at least the claimed number of induced copies of $F^\pi$ in $	ilde{G}_N$. As already mentioned, it follows from the second part of Lemma 9 that a.a.s. there are not more copies than that.

\[\square\]

5. Proofs of technical lemmas

In this section we prove the remaining open claims. We begin by deriving Theorem 4 from Theorem 3. Reusing some of the arguments from that proof, we then show Lemma 13. Lastly, we deduce Theorem 5 from Theorem 3.

5.1. Proof of Theorem 4 and Lemma 13. Theorem 4 is an immediate consequence of the next lemma.

**Lemma 15.** For any nonempty graph $F$ and any integer $r \geq 2$, $\theta^*(F, r)$ as defined in Theorem 3 and $m^*(F)$ as in Theorem 4 satisfy

$$
m^*(F) = \frac{1}{\theta^*(F, r)}.
$$

**Proof.** For any nonempty ordered graph $(F, \pi)$, set

$$
\tilde{H}(F, \pi) := \{ \tilde{H} = (H_1, \sigma), H_2, \ldots, H_h) \mid H_1 \subseteq F \land \sigma = \pi|_{H_1} = (e_1, \ldots, e_h) \land \forall i \geq 2 : (H_i \subseteq H_1 \setminus \{e_1, \ldots, e_{i-1}\} \land e_i \in H_i) \}
$$

(cf. the maximizations in (8) and (9)). For all $\tilde{H} \in \tilde{H}(F, \pi)$ we define

$$
e^*(\tilde{H}) := 1 + \sum_{i=1}^h c_i(e(H_i) - 1),
$$

$$
v^*(\tilde{H}) := 2 + \sum_{i=1}^h c_i(v(H_i) - 2),
$$

where the coefficients $c_i = c_i(\tilde{H}, r)$ are defined in (7). Furthermore, we set

$$
\mu^*_{r, \theta}(\tilde{H}) := v^*(\tilde{H}) - e^*(\tilde{H}) \cdot \theta.
$$

\[\square\]
Note that
\[
\max_{H_1 \subseteq F} d^* (H_1, \pi | H_1) \stackrel{8}{=} \max_{H \in H(F, \pi)} e^*(\tilde{H}) \frac{e^{**}(\tilde{H})}{v^{**}(\tilde{H})} = \frac{1}{\theta'(F, r)} ,
\]
where \(\theta'(F, r)\) is the unique solution of
\[
\min_{\tilde{H} \in \tilde{H}(F, \pi)} \mu^*_{\theta, r}(\tilde{H}) = 0 .
\]
Thus by (4) and (9), to prove the lemma it suffices to show that for any nonempty ordered graph \((F, \pi)\) and any \(r \geq 2\) and \(0 \leq \theta \leq 2\) we have
\[
\min_{\tilde{H} \in \tilde{H}(F, \pi)} \mu^*_{\theta, r}(\tilde{H}) = \min_{\tilde{H} \subseteq F} \lambda_{\theta, r}(H, \pi | H) .
\](55)

In the remainder of the proof we will show (55). To simplify notation we consider \(r\) and \(\theta\) fixed and drop all corresponding sub- and superscripts. Furthermore, in the following definitions we write \(\tilde{c}, \tilde{v}, \tilde{\mu}\) where in principle we should write \(\tilde{c}(H_1, \sigma), \tilde{v}(H_1, \sigma), \tilde{\mu}(H_1, \sigma)\), since all these quantities depend on the ordering \(\sigma \in \Pi(E(H_1))\).

We claim that \(e^*(\tilde{H})\) and \(v^*(\tilde{H})\) as defined in (53) can be written as
\[
\begin{align*}
e^*(\tilde{H}) &= 1 + r \cdot \tilde{c}(H_1, \ldots, H_h) , \quad (56a) \\
v^*(\tilde{H}) &= 2 + r \cdot \tilde{v}(H_1, \ldots, H_h) , \quad (56b)
\end{align*}
\]
where for \(1 \leq i \leq h\) the values \(\tilde{c}(H_i, \ldots, H_h)\) and \(\tilde{v}(H_i, \ldots, H_h)\) are recursively defined by
\[
\begin{align*}
\tilde{c}(H_i, \ldots, H_h) &= c(e(H_i) - 1) + (r - 1) \cdot \sum_{j=i+1}^h 1_{\{e_j \in H_i\}} \tilde{c}(H_j, \ldots, H_h) , \quad (57a) \\
\tilde{v}(H_i, \ldots, H_h) &= v(H_i) - 2 + (r - 1) \cdot \sum_{j=i+1}^h 1_{\{e_j \in H_i\}} \tilde{v}(H_j, \ldots, H_h) . \quad (57b)
\end{align*}
\]
This can be checked by induction, noting that for \(1 \leq k \leq h\) we have
\[
\begin{align*}
e^*(\tilde{H}) &= 1 + \sum_{i=1}^k c_i (e(H_i) - 1) + (r - 1) \cdot \sum_{j=k+1}^h \left( \sum_{i=1}^k c_i 1_{\{e_j \in H_i\}} \right) \tilde{c}(H_j, \ldots, H_h) , \\
v^*(\tilde{H}) &= 2 + \sum_{i=1}^k c_i (v(H_i) - 2) + (r - 1) \cdot \sum_{j=k+1}^h \left( \sum_{i=1}^k c_i 1_{\{e_j \in H_i\}} \right) \tilde{v}(H_j, \ldots, H_h) ,
\end{align*}
\]
which for \(k = h\) is equivalent to (53) and for \(k = 1\) is equivalent to (56). Combining (56) (57) via (54) also yields that
\[
\mu^*(\tilde{H}) = 2 - \theta + r \cdot \tilde{\mu}(H_1, \ldots, H_h) , \quad (58)
\]
where
\[
\tilde{\mu}(H_1, \ldots, H_h) := (v(H_1) - 2) - (e(H_i) - 1) \cdot \theta + (r - 1) \cdot \sum_{j=i+1}^h 1_{\{e_j \in H_i\}} \tilde{\mu}(H_j, \ldots, H_h) . \quad (59)
\]
It follows that for any fixed subgraph \(H_1 \subseteq F\) and \(\pi | H_1 = \sigma = (e_1, \ldots, e_h)\) the following holds: For \(1 \leq i \leq h\) and any graph \(H_i \subseteq H_1 \setminus \{e_1, \ldots, e_{i-1}\}\) with \(e_i \in H_i\), the value
\[
\lambda_{\pi(H_1, \sigma)}(H_i, i) := \min_{H_{i+1}, \ldots, H_h} \tilde{\mu}(H_i, H_{i+1}, \ldots, H_h) \quad (60)
\]
Furthermore, we can get rid of the big sum noting that (64) is equivalent to

\[ \tilde{\lambda}_{(H_1, \sigma)}(H_i, i) = (v(H_i) - 2) - (e(H_i) - 1) \cdot \theta \]

\[ + (r - 1) \cdot \sum_{j=i+1}^{h} 1_{\{e_j \in H_i\}} \min_{H_j \subseteq H_i \setminus \{e_1, \ldots, e_{j-1}\}: e_j \in H_j} \tilde{\lambda}_{(H_1, \sigma)}(H_j, j) . \]  

(61)

In the following we simplify this recursion step by step with the goal of relating it to \( \lambda() \) as defined in (59), cf. (66) below. Equation (55) will then follow.

Note that so far (61) is a recursion along a fixed edge ordering \( \sigma = (e_1, \ldots, e_h) \) of a fixed graph \( H_1 \), and the parameter \( \lambda_{(H_1, \sigma)}(H_i, i) \) is only defined for graphs \( H_i \subseteq H_1 \setminus \{e_1, \ldots, e_{i-1}\} \) with \( e_i \in H_i \).

We show that this context is irrelevant and that the value \( \tilde{\lambda}_{(H_1, \sigma)}(H_i, i) \) is in fact a function of the isomorphism class of \((H_i, \sigma|_{H_i})\) only. Towards that goal, we prove that for any fixed ordered graph \((H_1, \sigma)\) there exists a sequence of graphs \( H_2, \ldots, H_h \subseteq H_1 \) as in (52) minimizing \( \tilde{\mu}(H_1, H_2, \ldots, H_h) \) with the additional property that

\[ e_j \in H_i \implies H_j \subseteq H_i . \]  

(62)

Let \( H_2, \ldots, H_h \subseteq H_1 \) be graphs minimizing \( \tilde{\mu}(H_1, H_2, \ldots, H_h) \) such that every \( H_i \) is inclusion-maximal, and assume for the sake of contradiction that there exist indices \( 2 \leq i < j \) with \( e_j \in H_i \) but \( H_j \not\subseteq H_i \). Our choice of \( H_2, \ldots, H_h \) implies that for \( H'_i := H_i \cup H_j \) and \( H'_j := H_i \cap H_j \) we have

\[ \tilde{\mu}(H'_i, \ldots, H_h) - \tilde{\mu}(H_i, \ldots, H_h) > 0 , \]

\[ \tilde{\mu}(H_j, \ldots, H_h) - \tilde{\mu}(H'_j, \ldots, H_h) \leq 0 , \]

where the first inequality is strict because we assumed \( H_i \) to be inclusion-maximal. However, plugging in the definition (59) yields that both terms are exactly

\[ (v(H_j) - v(H_i \cap H_j)) - (e(H_j) - e(H_i \cap H_j)) \cdot \theta + (r - 1) \cdot \sum_{k=j+1}^{h} 1_{\{e_k \in H_i \setminus H_j\}} \tilde{\mu}(H_k, \ldots, H_h) , \]

which is a contradiction. Thus we may assume w.l.o.g. that (62) holds, and in the recursion (61) it suffices to minimize over graphs \( H_j \subseteq H_i \setminus \{e_1, \ldots, e_{j-1}\} \) (instead over graphs \( H_j \subseteq H_1 \setminus \{e_1, \ldots, e_{j-1}\} \)).

Observe that the context \((H_1, \sigma)\) is now irrelevant and that only \( \sigma|_{H_i} \), the induced order on the edges of \( H_i \), is required on the right hand side of (61). Thus by setting

\[ \tilde{\lambda}_{(H_1, \sigma)}(H_i, i) =: \tilde{\lambda}(H_i, \sigma|_{H_i}) \]

(63)

and changing notations accordingly, we obtain the ‘context-free’ recursion

\[ \tilde{\lambda}(H, \tau =: (e_1, \ldots, e_h)) = (v(H) - 2) - (e(H) - 1) \cdot \theta \]

\[ + (r - 1) \cdot \sum_{j=2}^{h} 1_{\{e_j \in H \setminus \{e_1, \ldots, e_{j-1}\}: e_j \in H_j\}} \min_{J_j \subseteq H \setminus \{e_1, \ldots, e_{j-1}\}: e_j \in J_j} \tilde{\lambda}(J_j, \tau|_{J_j}) . \]

(64)

Furthermore, we can get rid of the big sum noting that (64) is equivalent to

\[ \tilde{\lambda}(H, \tau) = \tilde{\lambda}(H \setminus e_1, \tau \setminus e_1) - \theta + (r - 1) \cdot \min_{J \subseteq H \setminus e_1: e_2 \in J} \tilde{\lambda}(J, \tau|_{J}) , \]

(65)
due to
\[
\tilde{\lambda}(H \setminus e_1, \tau \setminus e_1) = (v(H \setminus e_1) - 2) - (e(H \setminus e_1) - 1) \cdot \theta \\
+ (r - 1) \cdot \sum_{j=3}^{h} \min_{J_j \subseteq H \setminus \{e_1, \ldots, e_{j-1}\}: e_j \in J_j} \tilde{\lambda}(J_j, \tau | J_j)
\]
\[
= (v(H) - 2) - (e(H) - 2) \cdot \theta \\
+ (r - 1) \cdot \sum_{j=2}^{h} \min_{J_j \subseteq H \setminus \{e_1, \ldots, e_{j-1}\}: e_j \in J_j} \tilde{\lambda}(J_j, \tau | J_j)
\]
\[
- (r - 1) \cdot \min_{J_2 \subseteq H \setminus \{e_1\}: e_2 \in J_2} \tilde{\lambda}(J_2, \tau | J_2)
\]
\[
\tilde{\lambda}(H, \tau) + \theta - (r - 1) \cdot \min_{J \subseteq H \setminus \{e_2\}} \tilde{\lambda}(J, \tau | J).
\]
Substituting
\[
\tilde{\lambda}(H, \tau) = \lambda(H \setminus e_1, \tau \setminus e_1) - 2,
\]
we see that (65) is equivalent to
\[
\lambda(H, \tau) = \lambda(H \setminus e_1, \tau \setminus e_1) - \theta + (r - 1) \cdot \min_{J \subseteq H \setminus \{e_1\}} (\lambda(J \setminus e_1, \tau | J \setminus e_1) - 2),
\]
which is exactly the recursive step of (3). Moreover, if \((\bar{H}, \bar{\tau}) = (H \setminus e_1, \tau \setminus e_1)\) is empty we have
\[
\tilde{\lambda}(ar{H}, \bar{\tau}) = \lambda(H, \tau) + 2v(H) \lambda(\bar{H}, \bar{\tau}).
\]
This takes care of the base case and implies that
\[
\tilde{\lambda}(H, \tau) = \lambda(H, \tau)
\]
for all ordered graphs \((H, \tau)\). Thus we have for every fixed \((H_1, \sigma) = (e_1, \ldots, e_h)\), that
\[
\min_{H_2, \ldots, H_h} \mu^*(\tilde{\mu}) \geq 2 - \theta + r \cdot \min_{H_2, \ldots, H_h} \mu^*(\tilde{\mu})(H_1, 1)
\]
\[
\lambda(H_1, \sigma) \geq \lambda(H_1 \setminus e_1, \sigma \setminus e_1) - 2 \lambda(H_1 \setminus e_1, \sigma \setminus e_1) - 2.
\]
Still using the notation \(\pi|_{H_1} = \sigma = (e_1, \ldots, e_h)\) (cf. (52)), equation (55) now follows from
\[
\min_{\tilde{\mu}(H) \in \tilde{\mu}(F, \pi) \cap H \subseteq F} \left\{ 2 - \theta + r \cdot \min_{H_2, \ldots, H_h} \mu^*(\tilde{\mu})(H_1, 1) \right\}
\]
\[
\geq \min_{H_1 \subseteq F} \left\{ 2 - \theta + r \cdot (\lambda(H_1 \setminus e_1, \sigma \setminus e_1) - 2) \right\}
\]
\[
\geq \min_{H_1 \subseteq F} \lambda(H_1, \sigma) = \min_{H_1 \subseteq F} \lambda(H_1, \pi|_H),
\]
where in the last line we applied Lemma 10. This concludes the proof of Lemma 15.

Proof of Lemma 13. We extend the notion of connectedness to \(r\)-matched graphs. We call an \(r\)-matched graph \(\bar{H} = (V, \bar{K})\) connected if for any two vertices \(u, v \in V\) there is a sequence of \(r\)-sets \(K_1, \ldots, K_t \subseteq \bar{K}\) with \(u \in V(K_1), v \in V(K_t)\) and \(V(K_i) \cap V(K_{i+1}) \neq \emptyset\) for all \(1 \leq i \leq t - 1\). Since for a disconnected \(r\)-matched graph \(H\) the value \(\mu_{r, \theta}(H)\) is simply the sum of the \(\mu_{r, \theta}(\cdot)\)-values of all components of \(H\), it suffices to prove the claim for all connected \(r\)-matched subgraphs \(J \subseteq F^r\).
Moreover, we assume that \( J \) contains at least one \( r \)-set, since the claim holds trivially if \( \kappa(J) = 0 \) (cf. \( \Box \)). Due to \( \Box \) we may also assume that the minimization in \( \Box \) is over nonempty subgraphs \( H \subseteq F \). Hence it suffices to show that for any integer \( r \geq 2 \) and any \( 0 \leq \theta \leq 2 \) we have

\[
\min_{J \subseteq F^\pi: \kappa(J) \geq 1 \land J \text{ connected}} \mu_{r, \theta}(J) = \min_{H \subseteq F: e(H) \geq 1} \lambda_{r, \theta}(H, \pi|H) .
\]  

(69)

For the rest of the proof we consider \( r \) and \( \theta \) fixed and drop all corresponding subscripts. Let \( \pi := (e_1, \ldots, e_f) \) and observe that by the recursive structure of \( F^\pi \), for any \( 0 \leq i \leq f \) there are \( r^i \) copies of \((F_{-i})_i\) contained in \( F^\pi \), where \( F_{-i} := F \setminus \{e_1, \ldots, e_i\} \) and \( \pi_{-i} := \pi|F_{-i} \). Let \( J \) be a nonempty connected subgraph of \( F^\pi \), and let \( 0 \leq i \leq f - 1 \) be the largest index such that \( J \) is also contained in a copy of \((F_{-i})_i\) in \( F^\pi \). Then, by the maximal choice of \( i \) and the connectedness of \( J \), the graph \( J \) contains the central \( r \)-set of this copy (cf. Figure \( \Box \)). Thus \( \Box \) can be written equivalently as

\[
\min_{0 \leq i \leq f-1} \min_{J \subseteq (F_{-i})_i: K(e_{i+1}) \in J \land J \text{ connected}} \mu(J) = \min_{0 \leq i \leq f-1} \min_{e_i+1 \in H} \lambda(\pi|H) ,
\]  

(70)

where we use \( K(e_{i+1}) \in J \) as a shorthand notation to indicate that \( J \) contains the central \( r \)-set \( K(e_{i+1}) \) of \((F_{-i})_i\). We now show that the inner minimizations in \( \Box \) are equivalent. By changing variables \((F \leftarrow F_{-i} \text{ and } \pi \leftarrow \pi_{-i})\) this reduces to showing that for any nonempty ordered graph \((F, \pi) \), \( \pi = (e_1, \ldots, e_f) \), we have

\[
\min_{J \subseteq F^\pi: K(e_1) \in J \land J \text{ connected}} \mu(J) = \min_{H \subseteq F: e_1 \in H} \lambda(\pi|H) .
\]  

(71)

For any \( r \)-matched graph \( H \) we refer to a subgraph \( J \subseteq H \) that minimizes \( \mu(J) \) as a rarest subgraph of \( H \). Let \( H \) be an \( r \)-matched graph with an \( r \)-set \( K \subseteq H \) such that \( H \setminus K \) consists of \( r \) disjoint subgraphs \( H_1, \ldots, H_r \), each of which contains exactly the two endvertices of one of the edges from \( K \) (we refer to these two vertices as the attachment vertices of the corresponding subgraph). By the linearity of \( \mu(H) \) in \( v(H) \) and \( \kappa(H) \) (cf. \( \Box \)), a rarest connected subgraph of \( H \) containing the \( r \)-set \( K \) can be found by determining a rarest connected subgraph of \( H_i \) that contains the attachment vertices of \( H_i \) for all \( 1 \leq i \leq r \), and taking the union of all these rarest subgraphs together with the \( r \)-set \( K \). Using this general fact and exploiting the recursive structure of \( F^\pi \) (cf. Definition \( \Box \) and Figure \( \Box \)), we can determine a rarest subgraph of \( F^\pi \) that is connected and contains the central \( r \)-set recursively (‘outside-in’ in Figure \( \Box \)) as follows.

By \((F_{-i})_i\) we denote any copy of \((F_{-i})_i\) in \( F^\pi \). Moreover, \( \hat{F}_{-i} \) denotes the central copy of \((F_{-i})_i\) in \( \hat{F}_{-i} \), and \( \hat{e}_i \) denotes the edge of \( F^\pi \) that completes \( \hat{F}_{-i} \) to a copy of \((F_{-i-1})_i\). For \( i = f, f-1, \ldots, 1 \), we determine a rarest subgraph \( J_i \subseteq (\hat{F}_{-i})_i \) containing the two endvertices of \( \hat{e}_i \) (which play the role of attachment vertices) by determining an optimal choice of \( H_i := (J_i \cap \hat{F}_{-i})_i \cup \{\hat{e}_i\} \). By recursion, for each (black) edge \( e'_j \) that can be included into \( H_i \), \( i+1 \leq j \leq f \), and for each of the \( r - 1 \) grey edges \( \hat{e}_j \) in the same \( r \)-set as \( e'_j \), we already know the \( \mu() \)-value of a rarest subgraph \( J_j \subseteq (\hat{F}_{-i})_i \) (in the ‘branch’ of \( F^\pi \) corresponding to \( \hat{e}_j \)) containing the endvertices of \( \hat{e}_j \). For \( i+1 \leq j \leq f \), let \( \hat{J}_j \) denote such a rarest subgraph. Since the graph \( J_i \) resulting from choosing \( H_i \) contains exactly \( v(H_i) \) vertices and \( e(H_i) - 1 \) many \( r \)-sets not contained in any of the graphs \( \hat{J}_j \), \( i+1 \leq j \leq f \), we have

\[
\mu(J_i) = v(H_i) - (e(H_i) - 1) \cdot (2(r - 1) + \theta) + \sum_{j=i+1}^f 1_{\{e_j \in H_i\}}(r - 1)\mu(\hat{J}_j)
\]  

(72)
Setting \( \mu(J_i) - 2 =: \tilde{\lambda}_{(F, \pi)}(H_i, i) \), equation (72) yields for \( 1 \leq i \leq f \) the recursion

\[
\tilde{\lambda}_{(F, \pi)}(H_i, i) = (v(H_i) - 2) + (e(H_i) - 1) \cdot \theta + (r - 1) \sum_{j=i+1}^f 1_{\{e_j \in H_i\}} \min_{H_j \subseteq F_{j-1}} \tilde{\lambda}_{(F, \pi)}(H_j, j),
\]

which is essentially the same recursion as (61) in the proof of Lemma 15 (the difference is that \( H_1 \) was considered fixed and the underlying edge ordering was given by some \( \sigma \in \Pi(E(H_1)) \)). Analogously to the proof of Lemma 15 one can show that

\[
\tilde{\lambda}_{(F, \pi)}(H_1, 1) = \lambda(H_1 \setminus e_1, \sigma \setminus e_1) - 2,
\]

where \( \sigma := \pi|_{H_1} \) (cf. (68)). Similarly to (72) one also obtains

\[
\min_{J \subseteq F^\pi: K(e_1) \in J, J \text{ connected}} \mu(J) = -(2(r - 1) + \theta) + r \cdot \mu(\tilde{J}_1)
\]

\[
= 2 - \theta + r \cdot (\mu(\tilde{J}_1) - 2)
\]

\[
= 2 - \theta + r \cdot \min_{H_1 \subseteq F: e_1 \in H_1} \tilde{\lambda}_{(F, \pi)}(H_1, 1)
\]

where the term \( 2(r - 1) + \theta \) accounts for the central \( r \)-set \( K(e_1) \) and the factor \( r \) for the \( r \) copies of \( (F_1)_{\pi_1} \) that \( F^\pi \setminus K(e_1) \) consists of. We obtain

\[
\min_{J \subseteq F^\pi: K(e_1) \in J, J \text{ connected}} \mu(J) = \min_{H_1 \subseteq F: e_1 \in H_1} \{2 - \theta + r \cdot (\lambda(H_1 \setminus e_1, \sigma \setminus e_1) - 2)\}
\]

\[
\min_{H_1 \subseteq F: e_1 \in H_1} \lambda(H_1, \sigma) = \min_{H \subseteq F: e_1 \in H} \lambda(H, \pi|_H).
\]

where in the last line we applied Lemma 10 to the family of all ordered subgraphs of \((F, \pi)\) that contain the edge \( e_1 \). This proves (71) and thus the lemma. \( \square \)

5.2. **Proof of Theorem 5.** Theorem 5 is an immediate consequence of the next lemma.

**Lemma 16.** For any nonempty forest \( F \) and any integer \( r \geq 2, \theta^*(F, r) \) as defined in Theorem 3 and \( e^{**}(F) \) as in Theorem 5 satisfy

\[
\theta^*(F, r) = 1 + \frac{1}{e^{**}(F)}.
\]

In order to prove Lemma 16 we extend (10) and (11) to ordered trees and forests. By \( \pi_0 \) we denote the empty edge ordering. For any ordered tree \( (T, \pi) \), \( \pi = (e_1, \ldots, e_t) \), we define recursively

\[
e^{**}(K_1, \pi_0) := 0
\]

\[
e^{**}(T, \pi) := 1 + r \cdot (e^{**}(T_1, \pi|_{T_1}) + e^{**}(T_2, \pi|_{T_2}))
\]

where \( T_1 \) and \( T_2 \) are the two smaller trees obtained from \( T \) by removing the edge \( e_1 \). For any ordered forest \( (F, \pi) \), we set

\[
e^{**}(F, \pi) := \max \{ e^{**}(T, \pi|_T) \mid T \text{ is a component of } F \}.
\]

A straightforward inductive argument (exploiting the fact that \( e^{**}(T_1, \pi|_{T_1}) \) and \( e^{**}(T_2, \pi|_{T_2}) \) can be minimized independently in (75)) yields that \( e^{**}(T) = \min_{\pi \in \Pi(E(T))} e^{**}(T, \pi) \) for any tree \( T \), and hence also

\[
e^{**}(F) = \min_{\pi \in \Pi(E(F))} e^{**}(F, \pi)
\]

for any forest \( F \). It is also easy to see that for any ordered forest \( (F, \pi) \) and any subforest \( H \subseteq F \) we have

\[
e^{**}(H, \pi|_H) \leq e^{**}(F, \pi).
\]
In the following we denote by $T(F)$ the set of connected components of a forest $F$ (these components are trees). For convenience we use the parameter $\alpha = \theta - 1$.

**Lemma 17.** For any ordered forest $(F, \pi)$, $\pi = (e_1, \ldots, e_f)$, and any $0 \leq \alpha \leq 1$ we have

$$\lambda_{1+\alpha,r}(F, \pi) \leq \sum_{T \in T(F)} \left(1 - \alpha e^*(T, \pi|T)\right),$$

where $\lambda_{1+\alpha,r}(\cdot)$ and $e^*(\cdot)$ are defined in (3) and (75), (76), respectively. Moreover, if $1 - \alpha e^*(F, \pi) \geq 0$, then (79) holds with equality.

**Proof.** We proceed by induction on $e(F)$. If $e(F) = 0$ the claim trivially holds as

$$\lambda_{1+\alpha,r}(F, \pi_\emptyset) \leq v(F) = \sum_{T \in T(F)} 1 \sum_{T \in T(F)} (1 - \alpha e^*(T, \pi|T)) .$$

For the induction step we denote by $T(e_1) \in T(F)$ the component that contains the edge $e_1$ and by $T_1$ and $T_2$ the two smaller trees obtained from $T(e_1)$ by removing the edge $e_1$. We have

$$\lambda_{1+\alpha,r}(F, \pi) = \lambda_{1+\alpha,r}(F \setminus e_1, \pi \setminus e_1) - (1 + \alpha) + (r - 1) \cdot \min_{J \subseteq F: e_1 \in J} \left(\lambda_{1+\alpha,r}(J \setminus e_1, \pi|J \setminus e_1) - 2\right)$$

$$\leq \sum_{T \in T(F \setminus e_1)} (1 - \alpha e^*(T, \pi|T)) - (1 + \alpha)$$

$$+ (r - 1) \cdot \min_{J \subseteq F: e_1 \in J} \left(\sum_{T \in T(J \setminus e_1)} (1 - \alpha e^*(T, \pi|T)) - 2\right) ,$$

which proves the first part of the lemma.

For the second part we show that if $1 - \alpha e^*(F, \pi) \geq 0$, then both inequalities in the above calculation are in fact equalities. Note that due to (78) we also have $1 - \alpha e^*(F \setminus e_1, \pi \setminus e_1) \geq 0$, which by induction implies equality in (80). To show equality in (81), we show that the minimum in (80) is attained for $J = T(e_1)$. Similarly to above it follows from (78) that for any subtree $T \subseteq F$ we have $1 - \alpha e^*(T, \pi|T) \geq 0$. Hence there is a forest $J \subseteq F$ minimizing

$$\min_{J \subseteq F: e_1 \in J} \sum_{T \in T(J \setminus e_1)} (1 - \alpha e^*(T, \pi|T))$$

for which $J \setminus e_1$ contains exactly two components, one for each endvertex of $e_1$. Again by (78) we may assume that these are chosen as large as possible, which implies that indeed the minimum is attained for $J = T(e_1)$, yielding equality in (81).

**Proof of Lemma 16.** Setting $\alpha \coloneqq 1/e^*(F)$, we need to prove that

$$\Lambda_{1+\alpha,r}(F) \max_{\pi \in \Pi(E(F))} \min_{H \subseteq F} \lambda_{1+\alpha,r}(H, \pi|H) = 0 .$$

(82)
For any $\pi \in \Pi(E(F))$ we have

$$
\min_{H \subseteq F} \lambda_{1+\alpha,r}(H, \pi|_H) \overset{\text{Lemma } \ref{lemma:lambda}}{\leq} \min_{H \subseteq F} \sum_{T \in T(H)} (1 - \alpha e^r(T, \pi|_T)) .
$$

(83)

By definition (76), any ordered forest $(F, \pi)$ has a component $T \in T(F)$ with

$$
e^r(T, \pi|_T) = e^r(F, \pi) \overset{\text{Lemma } \ref{lemma:exponent}}{=} e^r(F) = 1/\alpha .
$$

Choosing any such component as $H$ yields that the right hand side of (83) is nonpositive, implying that $\min_{H \subseteq F} \lambda_{1+\alpha,r}(H, \pi|_H) \leq 0$ for all $\pi \in \Pi(E(F))$. Furthermore, for $\widehat{\pi} \in \arg \min_{\pi \in \Pi(E(F))} e^r(F, \pi)$ and any $H \subseteq F$ we have

$$
1 - \alpha e^r(H, \widehat{\pi}|_H) \overset{\ref{lemma:lambda}}{=} 1 - \alpha e^r(F, \widehat{\pi}) \overset{\ref{lemma:exponent}}{=} 1 - \alpha e^r(F) = 0 .
$$

(84)

Consequently, Lemma \ref{lemma:lambda} guarantees equality in (83) for $\pi = \widehat{\pi}$. Moreover, (84) also shows that for $\pi = \widehat{\pi}$ and any choice of $H \subseteq F$ all terms of the sum in (83) are nonnegative. Since we already know that the right hand side of (83) is nonpositive, it must be equal to zero and we have $\min_{H \subseteq F} \lambda_{1+\alpha,r}(H, \widehat{\pi}|_H) = 0$, which proves (82).

6. Concluding Remarks

- We outline how Theorem 3 can be proved if the sampling model of the Achlioptas process is changed. In our definition of the process, sampling in each step is only from edges that have never been drawn before. In the literature one also finds the variant where sampling is from all $(n^2)$ edges in each step. Our approach can be adapted to this setting by considering $r$-matched multigraphs, in which the $r$-sets of edges are not necessarily disjoint, but are allowed to overlap. This relies on the fact that Theorem 8 and Lemma 9 continue to hold for the random $r$-matched multigraph resulting from this modified process.

In \cite{10}, an intermediate model is used, where sampling is from from all edges that have not been selected before. Since in this setting the $r$-sets presented during the process slightly depend on the edge selection strategy being used, there is no convenient ‘static’ random graph structure for which we can formulate variants of Theorem 8 and Lemma 9. However, we can ‘sandwich’ this model between the other two models as follows: For the upper bound proof we simply ignore steps in which a previously seen edge is drawn. The $r$-sets from all other steps are then distributed as in our model. For the lower bound, we sample from all $(n^2)$ edges in each step, but ignore steps in which a previously selected edge is drawn (assuming w.l.o.g. that in this case such an edge is selected by default). Standard calculations show that in both cases, the number of ignored steps is $o(N)$ a.a.s., implying that Theorem 3 also holds for this intermediate model.

- The convergence towards the offline exponent stated in \cite{6} is best seen in the formulation of Theorem 4. Definition 8 implies that for any $\pi \in \Pi(E(F))$ we have for $H \subseteq F$ with $m_2(F) = (e(H) - 1)/(v(H) - 2)$ that

$$
d^r(H, \pi|_H) \geq \frac{1 + r(e(H) - 1) + v(H) - 2}{2 + r(v(H) - 2)} ,
$$

where the right hand side tends to $m_2(F)$ for $r \to \infty$. Moreover, for any $\pi \in \Pi(E(F))$ and any $H \subseteq F$ we have

$$
d^r(H, \pi|_H) \leq \max \left\{ \frac{1}{2}, \max_{H' \subseteq F} \frac{e(H') - 1}{v(H') - 2} \right\} = m_2(F) .
$$
Together it follows with (9) that
\[
\lim_{r \to \infty} m^{r^*}(F) = m_2(F) ,
\]
which is equivalent to (6).

- If \( F \) is a forest, the threshold given by Theorem 5 is also achieved by the following simpler strategy: In step \( N \) of the Achlioptas process, calculate for each edge \( f \in E_N \) the value
\[
d'(f) := \max \{ e^{r^*}(T) \mid T \text{ is a tree } \land f \text{ closes a copy of } T \} ,
\]
and select \( f^N \) as the edge for which this value is minimal (compare this to (18)). Ties can be broken arbitrarily, and edge orderings can be ignored. Thus for the case of forests there is indeed a simple memoryless strategy, cf. the remarks in Section 1.2. Moreover, this strategy is universal in the sense that it does not depend on the specific forest that should be avoided. These simplifications rely on the facts that as long as \( N \ll n^{1-\alpha} \) for some \( \alpha > 0 \), a.a.s. the connected components of \( G'(n, N) \) are ‘tree-like’, and their size is bounded by a constant \( C = C(\alpha) \).

- It is not hard to see that the lower bound proof given in Section 3 yields the following slight strengthening of our main result: If \( N \ll N_0(F, r, n) \), then \( F \) can be avoided a.a.s. even if the \( r \)-sets of a random \( r \)-matched graph \( G'_{n, r, N} \) are presented in an order chosen by an adversary (instead of in a random order).

- Our results extend to the case where an entire family of forbidden graphs should be avoided simultaneously: The threshold for avoiding a nonempty finite family \( \mathcal{F} \) of fixed nonempty graphs in the generalized Achlioptas process with parameter \( r \) is \( N_0(\mathcal{F}, r, n) = n^{2-\theta^*} \), where \( \theta^* = \theta^*(\mathcal{F}, r) \) is the unique solution of
\[
\max_{F \in \mathcal{F}} \Lambda_{r, \theta}(F) = 0 .
\]
Equivalently, this can be written as \( N_0(\mathcal{F}, r, n) = n^{2-1/m^{r^*}(\mathcal{F})} \), where
\[
m^{r^*}(\mathcal{F}) := \min_{F \in \mathcal{F}} m^{r^*}(F) .
\]
A strategy achieving this threshold is obtained by replacing the set \( \mathcal{S}(F) \) with the set
\[
\mathcal{S}(\mathcal{F}) := \bigcup_{F \in \mathcal{F}} \mathcal{S}(F)
\]
in the strategy described in Section 3.

- In [13] and [15], a related ‘balanced Ramsey game’ was studied, where the player has \( r \) colors at her disposal and the \( r \) edges presented in each step have to be colored immediately subject to the restriction that each of the \( r \) colors is used for exactly one edge. In this setting, the goal is to avoid creating a monochromatic copy of some fixed graph \( F \). Note that any upper bound for the Achlioptas problem immediately carries over to the balanced Ramsey setting. It follows from lower bounds proved in [15] that for ‘easy’ cases (those for which the naive strategy as presented in Section 2 is best possible) the balanced Ramsey game and the Achlioptas problem have the same threshold. It is an open question whether this is true for all non-forests \( F \) (one easily sees that the two problems have different thresholds if \( F \) is e.g. a star).

References


