THE KŐNIG GRAPH PROCESS

NINA KAMČEV, MICHAEL KRIVELEVICH, NATASHA MORRISON, AND BENNY SUDAKOV

ABSTRACT. Say that a graph G has property $\mathcal{K}$ if the size of its maximum matching is equal to the order of a minimal vertex cover. We study the following process. Set $N := \binom{n}{2}$ and let $e_1, e_2, \ldots e_N$ be a uniformly random ordering of the edges of K_n, with n an even integer. Let G_0 be the empty graph on n vertices. For $m \geq 0$, G_{m+1} is obtained from G_m by adding the edge e_{m+1} exactly if $G_m \cup \{e_{m+1}\}$ has property $\mathcal{K}$. We analyse the behaviour of this process, focusing mainly on two questions: What can be said about the structure of G_N and for which m will G_m contain a perfect matching?

1. INTRODUCTION

The modern study of random graph processes began in 1959 with the inaugural papers of Erdős and Rényi [10, 11]. Given a uniformly random permutation $e_1, \ldots, e_N$ of $E(K_n)$, they studied the evolution and properties of the graph $G_{n,m}$ with edge set $\{e_1, \ldots, e_m\}$, which is now known as the *Erdős-Rényi random graph*. This work has since grown into a well-established research area with many important applications in theoretical computer science, statistical physics, and other branches of mathematics [5, 20, 17].

An important variant of the standard Erdős-Rényi process, often referred to as the *random greedy process*, is the following. Given a graph property $\mathcal{P}$, preserved by the removal of edges, begin with an empty n-vertex graph and at each step add an edge chosen uniformly at random from those that do not violate property $\mathcal{P}$. The random greedy process was first considered by Ruciński and Wormald [27] (in the case of bounded degree) and, following discussions of Bollobás and Erdős, by Erdős, Suen and Winkler in 1995 [13] (in the case of triangle-freeness). Their motivation was defining and analysing a natural probability measure on the set of $\mathcal{P}$-maximal graphs.

A particularly well studied property is that of being H-free for a general graph H. In many cases, the final graph obtained at the end of the H-free process has been used to give constructions of interest in extremal combinatorics. In particular, such constructions have been found to improve lower bounds on Turán numbers (see [29, 3]) and on off-diagonal Ramsey numbers (for example [1, 3, 4, 16]).

In addition to looking at the structure and properties of the final graph, one often asks questions about the evolution of the process itself (see, e.g., [3, 26]). The properties mentioned so far are *decreasing* (closed under removal of edges) and *local*. Monotonicity of $\mathcal{P}$ guarantees that the final graph G_N is maximal in $\mathcal{P}$ and facilitates the use of some common techniques such as coupling with a modified process. So far, global properties are far less well understood

and there is no standard approach to analysing these processes (see for instance, the properties of being planar [19], r-colourable [26] and k-matching-free [23]).

In this paper we consider a global non-monotone property of a graph G, that the size of a maximum matching $\nu(G)$ is equal to the cardinality of an optimal vertex cover $\tau(G)$. A *vertex cover* in G is a set of vertices incident to any edge of G. Equivalently, the complement of a vertex cover contains no edges of G and is thus an *independent set*. We say that the vertex cover C is *optimal* if there is no vertex cover of cardinality less than C. It is easy to see that in general $\nu(G) \le \tau(G) \le 2\nu(G)$. We say that G is a *Kőnig graph* (or has property $\mathcal{K}$) if $\nu(G) = \tau(G)$.

The properties $\nu(G)$ and $\tau(G)$ and the relationship between them have been studied in many contexts. A foundational theorem of Kőnig and independently Egerváry [22, 9] says that bipartite graphs have the Kőnig property. The problem of finding an optimal vertex cover NP-hard but it can be solved in polynomial time in Kőnig graphs via the maximum matching. However, most graphs are closer to the other end of the spectrum, where $\tau(G) \sim 2\nu(G)$. As, with high probability, $G_{n,m}$ with $m = \frac{1}{2}n(\log n + \omega(1))$ has a perfect matching [12], whereas $\tau(G_{n,m}) \sim n$ for $m \gg n$ [20]. [1]

In light of this, we are interested in the evolution of a random graph process constrained by the Kőnig property, defined as follows. Let G_0 be the empty graph on vertex set V, where $|V| = n$ and $N = \binom{n}{2}$. Let $e_1, e_2, \ldots e_N$ be a uniformly random ordering of the edges of the complete graph on K_n on V. At each step $m \ge 0$, the edge e_{m+1} is *offered* to G_m. Say that a vertex pair f is *acceptable* for G_m if $f \notin \{e_1, \ldots, e_m\}$ and $G_m + f$ has property $\mathcal{K}$. If e_{m+1} is acceptable for G_m, we set $G_{m+1} := G_m + e_{m+1}$ and say that the edge e_{m+1} is *accepted*. Otherwise we say that e_{m+1} is *rejected* and set $G_{m+1} := G_m$. An important observation is that an edge e is acceptable for G_m if and only if e is incident to an optimal cover in G_m or extends a maximum matching in G_m. In the remainder of the paper, we assume that the number of vertices n is even. Let G_m^{all} be the graph whose edge set is $\{e_1, \ldots, e_m\}$. Note that G_m^{all} is distributed as $G_{n,m}$. All the proofs translate to odd n if we define a perfect matching to be a matching of order $\lfloor \frac{n}{2} \rfloor$.

We remark that it is also natural to consider an alternative process in which $\tau(G) = \nu(G)$ is maintained not just for G_m, but for every subgraph of it. However, this condition is equivalent to bipartiteness and yields precisely the bipartite graph process considered by Erdős, Suen and Winkler [13]. Our process and the resulting graph are rather different.

Let us return to the Kőnig process and consider what can be said about the structure of G_N. We start with a simple proposition which is proved in Section 4.

Proposition 1.1. *For all $m \ge 1$, the graph G_m has a maximum matching that intersects every edge of G_m^{all}.*

At the end of the process, since all the edges of K_n have been offered, deterministically the final graph G_N has a perfect matching. This settles the value of $\nu(G_N)$ and raises a number of further questions about the typical structure of G_N and the evolution of the process (in particular, appearance of a perfect matching). As G_N has a vertex cover C of order $\frac{n}{2}$, it is a

---

[1]As usual, we say an event occurs *with high probability* if it occurs with probability tending to 1 as $n \to \infty$. We write $f \sim g$ if $f/g \to 1$, $f = o(g)$ or $f \ll g$ if $f/g \to 0$, $f = O(g)$ if $|f| \le Cg$ for some constant C. We write $\omega(1)$ to denote a function tending to infinity with underlying parameter n; we also write $\omega(n)$ for $\omega(n) = \omega(1) \cdot n$.

subgraph of the classical Erdős-Gallai graph G_n^*, which consists of a clique on $\frac{n}{2}$ vertices that is completely joined to an independent set on the other $\frac{n}{2}$ vertices. Given this, we wonder how close is the graph G_N to G_n^*, or in other words, how many vertex pairs incident to C are missing? How 'volatile' is the optimal cover typically in the initial stages of the process, and at which point does it become 'rigid' or unique? This question is also important for understanding the evolution of the process, in particular the proportion of acceptable edges. Our first main result is the following.

Theorem 1.2. *Let $\epsilon > 0$. With high probability, the Kőnig process satisfies the following properties.*

(i) $G_{(1+\epsilon)n\log n}$ contains a perfect matching.
(ii) $G_{4n\log n}$ has a unique optimal vertex cover C.
(iii) There are $O(n)$ vertex pairs incident to C that are not present in G_N.

Furthermore, we analyse in more detail the appearance of a perfect matching in G_m. In the Erdős-Rényi process $(G_{n,m})$, Bollóbas and Thomason [7] proved that the very edge that links the last isolated vertex to another vertex makes the graph connected and completes a perfect matching with high probability if n is even. This happens when $m = \frac{1}{2}n(\log n + \omega(1))$. In fact, they showed that if $m \geq \frac{1}{4}n(\log n + \omega(1))$, then $G_{n,m}$ contains a perfect matching on all but at most one non-isolated vertex with high probability. For $m < \frac{1}{4}n\log n$, there are other structural obstructions to containing a perfect matching – in this regime, $G_{n,m}$ is likely to contain vertices with two neighbours of degree one, only one of which can be contained in a perfect matching. However, *quantitatively*, the isolated vertices are the main obstruction throughout the evolution of the process, as shown by Frieze [18].

Our second main result is an analogue of [7] for our process. We show that, with high probability, a perfect matching in G_m occurs significantly later than in $G_{n,m}$.

Theorem 1.3. *Let $m := \left(\frac{1}{2} + \frac{1}{70}\right)n\log n$. With high probability, G_m contains isolated vertices.*

This delay in G_m is surprising, as one might guess that the number of isolated vertices decays roughly at the same rate as in $G_{n,m}$. This would indeed be the case if most pairs containing isolated vertices in G_m were to extend the current maximum matching.

The delay is, in spirit, similar to *Achlioptas processes*, which were conceived for the sake of influencing the typical appearance of graph theoretic properties. Initiating a fruitful line of research, Bohman and Frieze exhibited an Achlioptas process with a delayed phase transition [2]. Besides the phase transition, several other graph properties were considered. For instance, it is known that connectivity and occurrence of a Hamilton cycle (and hence a perfect matching) can be accelerated [21, 24].

As property $\mathcal{K}$ is non-monotone and global, many of our arguments involve novel ideas (to our knowledge) and could be of their own interest or be adapted to study other global properties. Our approach combines probabilistic arguments with combinatorial maximum-matching methods, resulting in intuitive and conceptually simple proofs.

The paper is structured as follows. In the next section, we introduce standard notation that will be used throughout and outline in detail how the proof will proceed. Then in Section 3 we recall some standard probabilistic tools and prove some preliminary results. In Section 4 we use standard techniques to find bounds on $\nu(G_m)$ for various time steps m

and prove Theorem 1.2(i). Theorem 1.2(iii) is proved in Section 5. In Section 6 we prove Theorem 1.2(ii). We also show that, at almost all time steps from $c_1 n \log n$ to $2n \log n$, the number of vertices contained in an optimal cover is close to $\frac{n}{2}$. This will be used in Section 7 to prove Theorem 1.3. We conclude in Section 8 by mentioning some related open problems.

2. Overview of Proof

In this section we give a brief overview of the arguments to come, introduce our main lemmas and define some notation that will be used throughout. Throughout the proof n will be taken to be sufficiently large when needed. Any statements made during the discussion in this section about G_m hold with high probability (but we may not write this every time). Of course, any formal statements are made explicit.

Throughout the paper, the probability measure conditional on $(e_1, \ldots, e_m)$ is $\mathbb{P}_m$, and the corresponding expectation is $\mathbb{E}_m$. In a slight abuse of notation, we use (G_m) for the probability space as well as the sampled process. The logarithm to base e is denoted by $\log$.

Our first goal in Section 4 is to show that, with high probability, when m is not too small, G_m contains a matching of size $\frac{n}{2}(1 - o(1))$. More precisely, we will prove the following.

Lemma 2.1. *Let $\gamma > 0$ be a sufficiently large constant and let $\gamma n \leq m \leq 2n \log n$. With probability $1 - O(n^{-2})$,*

$$\nu(G_m) \geq \frac{n}{2}\left(1 - e^{-3m/(10n)}\right).$$

This lemma will follow from the fact that each vertex is typically in linearly many acceptable pairs and fairly standard graph-theoretic arguments (see, for example, [17, Section 6.1]) relying on expansion properties of the graph. The same reasoning yields that, with high probability, after $m := (1 + \epsilon)n \log n$ steps, each vertex has degree at least $\Omega(\log n)$ and G_m has a perfect matching. This proves statement (i) of Theorem 1.2.

In Section 5 we prove Theorem 1.2(iii). Let C_N be an optimal cover in G_N and let $D_m \subseteq V(G_m)$ be the set of vertices contained in some optimal cover of G_m. How can an edge e incident to C_N be rejected during the process? This is only possible if e is not incident to a current optimal cover at the time step m when it is offered (i.e. $e \cap D_m = \emptyset$). The key idea in controlling those rejected edges is to show that for $m' = 1.1n \log n$, most vertices in an optimal vertex cover of $G_{m'}$ have been in an optimal cover for most $m < m'$. Therefore we rarely rejected an edge touching such a vertex. We also know that with high probability, $G_{m'}$ has a perfect matching. Then deterministically, no edges incident to C_N will be rejected after the time step m', since C_N is also an optimal cover in $G_{m'}$.

The question considered in Section 6 is how 'rigid' an optimal cover of G_m is during the earlier evolution of the process. We will see that if $|D_m|$ is significantly larger than $\frac{n}{2}$ in a positive proportion of steps m, we are accepting *too many* edges into our graph to maintain an independent set of order $\frac{n}{2}$. Let us state the lemma formally. For $m_0, t \geq 0$, let $T(m_0, t)$ be the time period of length t beginning at m_0, i.e. $T(m_0, t) := \{m_0, \ldots, m_0 + t - 1\}$.

Lemma 2.2. *Let $m_0 := n\sqrt{\log n}$. For $\gamma > 0$, let $T := T(m_0, \gamma n \log n)$. With high probability, there are at most $\gamma n(\log n)^{9/10}$ values of $m \in T$ such that $|D_m| > \left(\frac{1}{2} + (\log n)^{-1/10}\right)n$.*

This lemma will be useful in proving Theorem 1.3, as it gives us an upper bound on the number of acceptable pairs for G_m. We conclude Section 6 by showing that after $4n \log n$ steps, the optimal cover in G_m is unique, which proves Theorem 1.2(ii).

The proof of Theorem 1.3 is given in Section 7. Let us discuss roughly how the number of isolated vertices decays in the Kőnig process (G_m) compared to a typical Erdős-Rényi process $(G_{n,m})$. In a single time step of $(G_{n,m})$, each isolated vertex is connected with probability asymptotically $2/n$. This dynamic allows one to relate the number of isolated vertices in $G_{n,m}$ to the classical coupon-collector problem and implies that with high probability, the isolated vertices disappear when $m = \frac{1}{2}(n \log n + \omega(1))$.

Now consider the Kőnig process (G_m) after $\nu(G_m) \geq (\frac{1}{2} + o(1))n$. Let I_m be the number of isolated vertices in G_m. We will show that there are time steps (called *unhelpful*) in which each isolated vertex is connected with probability only $\frac{1}{n}(1 + o(1))$. The reason is that in an unhelpful step m, an isolated vertex v cannot be used to extend a maximum matching in G_m, so now consider the Kőnig process (roughly half of the vertex pairs containing v (those not incident to a current optimal cover) are unacceptable. This is shown in Subsection 7.1 using a careful analysis of a maximum matching in G_m. In Subsection 7.2 we show that the unhelpful states typically constitute a constant proportion of the time steps. Since in an unhelpful state m, I_m has a smaller decay rate than in $G_{n,m}$, it follows (via martingale arguments in Subsection 7.3) that $G_{(1/2+C)n \log n}$ has isolated vertices for some $C > 0$. Our argument does not give a sharp constant C, so we do not attempt to make marginal improvements.

3. Preliminaries and Probabilistic Tools

In this section we gather together some basic probabilistic tools and standard results that will be used throughout the paper.

3.1. The relationship between $G_{n,m}$ and $G(n,p)$.

Let $G(n,p)$ denote the n-vertex random graph in which every possible edge is present independently with probability p. The following lemma allows us to prove that G_m^{all} has certain properties, by considering the properties of $G(n, m/N)$. A graph property $\mathcal{P}$ is said to be *monotone increasing* if $G \in \mathcal{P}$ implies that $G + e \in \mathcal{P}$.

Lemma 3.1 ([17], Lemma 1.3). *Let $\mathcal{P}$ be any graph property and let $p = p(n)$ satisfy $n^2 p, n(1 - p)p^{-1/2} \to \infty$. If $m = p\binom{n}{2}$ and n is sufficiently large, then*

$$\mathbb{P}[G_{n,m} \in \mathcal{P}] \leq 10 m^{1/2} \mathbb{P}[G(n,p) \in \mathcal{P}].$$

Moreover, if $\mathcal{P}$ is monotone increasing, then $\mathbb{P}(G_{n,m} \in \mathcal{P}) \leq 3\mathbb{P}(G(n,p) \in \mathcal{P})$.

As a consequence, we immediately get a bound on the maximum degree in $G_{10n \log n}^{\text{all}}$.

Claim 3.2. *For $m = 10n \log n$, the graph G_m^{all} has maximum degree at most $200 \log n$ with probability $1 - O(n^{-2})$.*

Proof. Let $p := \frac{10n \log n}{\binom{n}{2}}$. Using the Chernoff bound (the third formulation in Theorem 3.3 stated below), the probability that a single vertex has degree $200 \log n$ in $G(n,p)$ is at most $2^{-200 \log n}$. Taking the union bound over all n vertices and applying Lemma 3.1 gives the required result. $\square$

3.2. Standard Estimates and Probabilistic Tools.

Here we collect together the standard probabilistic tools we will use during the proof. The first is a version of the Chernoff Bound taken from [8].

Theorem 3.3 (The Chernoff Bound). *Let $X_1, \ldots, X_n$ be a sequence of independent $[0,1]$-valued random variables and let $X = \sum_{i=1}^{n} X_i$. Then, for $0 < \varepsilon < 1$,*

$$\mathbb{P}\left(X < (1-\varepsilon)\mathbb{E}(X)\right) \leq e^{-\frac{\varepsilon^2 \mathbb{E}(X)}{2}},$$

$$\mathbb{P}\left(X > (1+\varepsilon)\mathbb{E}(X)\right) \leq e^{-\frac{\varepsilon^2 \mathbb{E}(X)}{3}}.$$

Moreover, if $t > 2e\mathbb{E}[X]$, then

$$\mathbb{P}[X > t] \leq 2^{-t}.$$

We will use a well-known result about the edge distribution in the random graph. As usual, $E_G(U, W)$ denotes the set of edges of a graph G with one endpoint in U and one in W and $E_G(U)$ the set of edges with both endpoints in U. We omit the index G when the graph is clear from the context. The number of edges in G is denoted by $e(G)$. This particular form is stated in [25] for $G(n, p)$, but Lemma 3.1 implies that it also holds for $G_{n,m}$.

Theorem 3.4. *Let $m = m(n) \leq 0.49n^2$. There exists a constant $\lambda > 0$ such that, with high probability, in $G_{n,m}$ every two disjoint sets $U, W \subset V$ of cardinality $|U| = u$, $|W| = w$ satisfy*

$$|E(U, W)| \in \frac{2muw}{n^2} \pm \lambda\sqrt{\frac{uwm}{n}} \quad and \quad |E(U)| \in \frac{mu^2}{n^2} \pm O\left(u\sqrt{\frac{m}{n}}\right).$$

We need another claim on the edge distribution in $G_{n,m}$. Although similar results are available in the literature, we will need an explicit bound on the probability so we include the proof here.

Lemma 3.5. *Let $\beta \leq 0.1$ and $m \geq 10\beta^{-2}n$. For $G := G_{n,m}$ and any set U of cardinality at most $\frac{\beta n}{2e}$, we have*

$$\mathbb{P}\left[|E_G(U)| \leq \frac{\beta m |U|}{n}\right] = 1 - O(n^{-2}).$$

Proof. We say that $G \in \mathscr{D}_u$ if some set $U \subset V(G)$ of cardinality u spans more than $\frac{\beta m u}{n}$ edges in G, and set $\mathscr{D} = \bigcup_{u=1}^{\beta n/(2e)} \mathscr{D}_u$. We will show that $\mathbb{P}[G(n,p) \in \mathscr{D}] = O\left(n^{-5}\right)$ in $G(n,p)$ with $p = \frac{2m}{n^2}$.

We estimate the probability of $\mathscr{D}_u$ by taking the union bound over all the vertex sets of cardinality u.

$$\mathbb{P}[G(n,p) \in \mathscr{D}_u] \leq \binom{n}{u}\binom{\frac{u(u-1)}{2}}{\frac{\beta n p u}{2}}p^{\frac{1}{2}\beta n p u} \leq \left(\frac{en}{u}\left(\frac{eu}{\beta n}\right)^{\frac{1}{2}\beta n p}\right)^u.$$

We proceed by splitting the range for u. First note for $u < \beta n p$, $\mathscr{D}_u$ is empty as a set of cardinality u cannot span more than $\frac{\beta n p u}{2} = \frac{\beta m u}{n}$ edges. If $u \leq \sqrt{n}$ and n is sufficiently large, then $eu/(\beta n) \leq n^{-1/3}$, so

$$\mathbb{P}[G(n,p) \in \mathscr{D}_u] \leq \left(en^{\frac{1}{2}} \cdot n^{-\frac{1}{3}\cdot\frac{1}{2}\beta n p}\right)^u \leq \left(en^{\frac{1}{2}-\frac{1}{6}\beta n p}\right)^u < n^{-2u}.$$

For $\sqrt{n} < u \leq \frac{\beta n}{2e}$, we recall the hypothesis that $\frac{1}{2}\beta n p \geq 10\beta^{-1}$ and deduce

$$\mathbb{P}[G(n,p) \in \mathscr{D}_u] \leq \left(\frac{en}{u}\left(\frac{eu}{\beta n}\right)^{\frac{1}{2}\beta n p}\right)^u \leq \left(2e^2\beta^{-1} \cdot 2^{-10\beta^{-1}}\right)^u < 2^{-u}.$$

Combining these estimates, we get

$$\mathbb{P}\left[G(n,p) \in \mathscr{D}\right] \leq \sum_{u=\beta np}^{\sqrt{n}} n^{-2u} + \sum_{u=\sqrt{n}}^{\frac{\beta n}{2e}} 2^{-u} = O\left(n^{-\beta np}\right) + O\left(n \cdot 2^{-\sqrt{n}}\right) = O\left(n^{-5}\right).$$

Lemma 3.1 gives $\mathbb{P}\left[G_{n,m} \in \mathscr{D}\right] = O\left(n^{-2}\right).$ $\qquad\square$

The following lemma is a bound for the lower tail of the binomial distribution. If we are looking to control very large deviations, elementary computations give a stronger estimate than the usual Chernoff bounds.

Lemma 3.6. *Let $X \sim \mathrm{Bin}(k,p)$ with $\mu = kp \to \infty$ as $k \to \infty$. Given $\eta > 0$ and a constant $0 < \delta \leq 1$ satisfying $\delta(2 + \log(1/\delta)) + (\log(\delta\mu))/\mu < \eta$ for sufficiently large k, we have*

$$\mathbb{P}\left[X \leq \delta\mu\right] \leq e^{-(1-\eta)\mu}.$$

Proof. By definition of $\mathrm{Bin}(k,p)$, $\mathbb{P}\left[X \leq \delta\mu\right] \leq \sum_{s=0}^{\delta\mu} \binom{k}{s} p^s (1-p)^{k-s}$. Standard inequalities give

$$\mathbb{P}\left[X \leq \delta\mu\right] \leq \sum_{s=0}^{\delta\mu} \left(\frac{ekp}{s}\right)^s e^{-(k-s)p} = \sum_{s=0}^{\delta\mu} \left(\frac{e^{1+p}\mu}{s}\right)^s e^{-kp},$$

As $\delta \leq 1$ and the summand is increasing in s for $s \leq \mu$, we can bound each summand by the final term, $s = \delta\mu$. It follows that

$$\mathbb{P}\left[X \leq \delta\mu\right] \leq \delta\mu \left(e^2 \delta^{-1}\right)^{\delta\mu} e^{-\mu}.$$

The hypothesis on δ is equivalent to $\log(\delta\mu) + \delta\mu(2 + \log(1/\delta)) \leq \eta\mu$, and hence

$$\mathbb{P}\left[X \leq \delta\mu\right] \leq e^{\eta\mu} e^{-\mu}.$$

$\qquad\square$

3.3. **Martingale concentration inequalities.** Recall that a sequence of random variables $X_0, X_1, \ldots$ is called a *martingale* if for each $i \geq 1$, we have $\mathbb{E}[X_i \mid X_0, \ldots, X_i] = X_{i-1}$. It is called a *submartingale* if $\mathbb{E}[X_i \mid X_0, \ldots, X_i] \geq X_{i-1}$. We will use two standard martingale concentration results in our proof. The first is Azuma's Inequality. The version we present here was taken from [8].

Theorem 3.7 (Azuma's Inequality). *Let $X_0, X_1, \ldots$ be a martingale such that for each $i \geq 0$ there exists a constant $c \geq 0$ such that $|X_m - X_{m-1}| \leq c$. Then,*

$$\mathbb{P}\left[|X_m - X_0| > \lambda\right] \leq 2\exp\left(-\frac{\lambda^2}{2c^2 m}\right).$$

The second is Freedman's Inequality. It gives a stronger concentration result than Azuma's inequality when the *average* differences $X_{m+1} - X_m$ are much smaller than the worst-case. To avoid working with filtrations and the corresponding notation, we state it in our specific context. The general statement can be found for instance in [28]. Moreover, in Section 7, we will apply a version of the Inequality for submartingales, where the bound only holds for the lower tail of Y_i. This formulation (with slightly stronger constants) can be found in [14].

Theorem 3.8 (Freedman's inequality). *Let $\sum_{i=1}^{m} Y_i$ be a real-valued submartingale, where $\mathbb{E}_{i-1}[Y_i] \geq 0$ and $|Y_i| \leq 1$ for $i \geq 0$. Let $W_m = \sum_{i=1}^{m} \mathbb{E}_{i-1}[Y_i^2]$. Then, for all $\ell > 0$ and $\sigma^2 > 0$,*

$$\mathbb{P}\left[\exists m : \sum_{i=1}^{m} Y_i \leq -\ell \text{ and } W_m \leq \sigma^2\right] \leq \exp\left[-\frac{\ell^2}{2\sigma^2 + 4\ell/3}\right].$$

If, in addition, $\mathbb{E}_{i-1}[Y_i] = 0$ for $i \in \mathbb{N}$,

$$\mathbb{P}\left[\exists m : \left|\sum_{i=1}^{m} Y_i\right| \geq \ell \text{ and } W_m \leq \sigma^2\right] \leq 2\exp\left[-\frac{\ell^2}{2\sigma^2 + 4\ell/3}\right].$$

Moreover, the following special case of Theorem 3.8 for indicator random variables will be useful in our applications.

Corollary 3.9. *Let X_i be an indicator random variable depending only on G_i and let $q_i := \mathbb{P}_{i-1}[X_i = 1]$. For any $m \in \mathbb{N}$ and $L > 0$,*

$$\mathbb{P}\left[\left(\left|\sum_{i=1}^{m}(X_i - q_i)\right| \geq Ln\right) \wedge \left(\sum_{i=1}^{m} q_i \leq 4Ln\right)\right] \leq e^{-Ln/10}.$$

Proof. Fix m, and define $T := \{1, \ldots m\}$. Let $Y_i := X_i - q_i$ for $i \geq 1$. By definition, $\mathbb{E}_{i-1}[Y_i] = 0$, so $\left(\sum_{i=1}^{m'} Y_i\right)_{m'}$ is a martingale with $|Y_i| \leq 1$. Moreover, $\sum_{i \in T} \mathbb{E}_i[Y_{i+1}^2] = \sum_{i \in T} q_i(1 - q_i)$, so our event implies the event stated in Freedman's inequality with $\sigma^2 = 4Ln$. Applying Freedman's inequality (with $\ell = Ln$) gives precisely

$$\mathbb{P}\left[\left(\left|\sum_{i \in T}(X_i - q_i)\right| \geq Ln\right) \wedge \left(\sum_{i \in T} q_i \leq 4Ln\right)\right] \leq 2\exp\left[-\frac{Ln}{8 + 4/3}\right] \leq e^{-Ln/10}.$$

□

4. Forming a large matching

Our goal in this section is to prove Lemma 2.1. It can be viewed as an analogue of Frieze's result [18] on $G_{n,m}$ where he showed that for $m = \Theta(n)$, $G_{n,m}$ contains a matching covering almost all non-isolated vertices with high probability.

We start by proving Proposition 1.1, a deterministic property of the Kőnig process. It will easily follow that after $O(n)$ steps we have a matching of linear size (see Corollary 4.1). A little more work (in Lemma 4.2) shows that we have a linear sized subgraph with 'large' minimum degree satisfying certain expansion properties. We are then able to conclude the proof of Lemma 2.1 using a standard expansion argument (see, for instance, Section 6.1 of [17]).

Proof of Proposition 1.1. By induction on m. The statement is trivial for $m = 0$. Suppose that it holds for the first $m - 1$ steps and consider the edge e_m. Let M be a matching where the set of vertices covered by M, denoted $V(M)$, is incident to $\{e_1, e_2, \ldots, e_{m-1}\}$. If e_m is disjoint from $V(M)$, we can take $M + e_m$ to be the required optimal matching. If e_m is incident to $V(M)$ and $\nu(G_m) = \nu(G_{m-1})$, then M still satisfies the condition.

Finally, if e_m is incident to $V(M)$ and $\nu(G_m) = \nu(G_{m-1}) + 1$, let M' be a matching of size $|M| + 1$ in G_m. The union of M and M' consists of cycles and paths alternating between

M and M', where one path P containing e_m has odd length, contains one more edge of M' than of M, and two endpoints outside M; all the remaining paths have even length. Then the matching M_1 created by replacing edges of M in P by those of M', is a maximum matching in G_m covering all vertices of M. Hence M' intersects every edge in $\{e_1, \ldots, e_m\}$, as required. $\qquad\square$

Corollary 4.1. *Let $\epsilon > 0$ and $m_0 := \epsilon^{-2}n$. With probability $1 - O\left(n^{-2}\right)$*

    (i) $\nu\left(G_{m_0}\right) \geq \frac{1-\epsilon}{2} \cdot n$, and

    (ii) whenever $m_0 \leq m \leq 2n\log n$, each vertex is contained in at least $\left(\frac{1}{2} - \epsilon\right)n$ pairs which are acceptable for G_m.

Proof. For every m, G_m^{all} has an independent set of order $n - 2\nu(G_m)$, namely the complement of $V(M)$, where M is the maximum matching granted by Proposition 1.1. Denoting the order of a maximum independent set in G by $\alpha(G)$, standard first-moment computations (see, e.g., [20, Section 7]) yield

$$\mathbb{P}\left[\alpha\left(G_{m_0}^{\mathrm{all}}\right) \geq \epsilon n\right] \leq e^{-\epsilon n} = O\left(n^{-2}\right).$$

It follows that with probability $1 - O\left(n^{-2}\right)$, $n - 2\nu(G_{m_0}) < \epsilon n$, as required for (i).

This tells us that with high probability $\tau(G_m) \geq \left(\frac{1}{2} - \frac{\epsilon}{2}\right)n$ for $m \geq m_0$. Any pair containing v and a vertex in an optimal vertex cover C_m of G_m is acceptable for G_m. By Corollary 3.2, with probability $1 - O(n^{-2})$, G_m^{all} has maximum degree $O(\log n)$ which implies that $o(n)$ pairs incident to v have been offered so far. Thus the number of acceptable pairs incident to v is at least $\left(\frac{1}{2} - \epsilon\right)n$, as required for (ii). $\qquad\square$

To avoid confusion, we remark that the bound in (i) is rather crude since the independence number of $G_{n,m}$ with $m = \epsilon^{-2}n$ is actually $\Theta\left(\epsilon^2 n \log\left(\epsilon^{-1}\right)\right)$. The probability 'benchmark' $O\left(n^{-2}\right)$ across this section is also arbitrarily chosen – all the probability bounds are significantly stronger.

Even though we need a stronger bound on $\nu(G_m)$, Corollary 4.1 is a very useful tool, providing a lower bound on the probability that e_{m+1} is acceptable for G_m. As usual, we let $\delta(G)$ denote the minimum degree of G. The neighbourhood of a vertex set S in a graph G, excluding S, is denoted by $N_G(S)$. We may omit the subscript when it is clear which graph plays the role of G. The following facts about the edge distribution of G_m will be used for our expansion arguments.

Lemma 4.2. *Let $0 < \beta \leq 10^{-3}$. There exists $\gamma > 0$ such that for all $\gamma n \leq m \leq 2n\log n$, with probability at least $1 - O\left(n^{-2}\right)$, G_m has the following properties.*

  (i) G_m has a subgraph H such that $|V(H)| \geq n\left(1 - e^{-m/(3n)}\right)$ and $\delta(H) \geq \frac{10\beta m}{n}$.

  (ii) For any set $S \subseteq V(H)$ with $|S| \leq \frac{\beta n}{16}$, $|N_H(S) \cup S| > 2|S|$.

Proof. First consider (i). We will use the following claim.

Claim 4.3. *Let $\alpha := e^{-m/(3n)}$ and let $\mathcal{A}$ be the event that there exists a set S of order αn such that $|E_{G_m}\left(S, V \setminus S\right)| \leq 10\alpha\beta m$. Then $\mathbb{P}[\mathcal{A}] = O\left(n^{-2}\right)$.*

Proof. Let $\mathcal{B}$ be the event that G_m^{all} has maximum degree at most $200\log n$ and the statements (i) and (ii) of Corollary 4.1 hold with $\epsilon = \frac{1}{20}$. By Claim 3.2 and Corollary 4.1, we can pick γ

large enough to ensure that $\mathcal{B}$ occurs with probability at least $1 - O\left(n^{-2}\right)$. It is now sufficient to show that

$$\mathbb{P}\left[\mathcal{A} \mid \mathcal{B}\right] \leq n^{-2}. \tag{4.1}$$

Let S be a set of order αn. As we condition on $\mathcal{B}$ occurring, we can choose $\gamma/10$ to be sufficiently large such that for all $m/10 \leq m' \leq 2n \log n$, each vertex in $G_{m'}$ is contained in at least $\frac{9}{20}n$ pairs acceptable for $G_{m'}$. Also, by choosing γ to be sufficiently large, we can ensure that $|S| \leq \frac{n}{100}$.

Combining these two facts with the maximum-degree bound gives that for all $m/10 \leq m' \leq 2n \log n$, each $v \in S$ is in at least $\frac{2n}{5}$ acceptable pairs (v, w) for $G_{m'}$ such that $w \notin S$. So the probability that $e_{m'+1}$ is acceptable and has exactly one endpoint in S is at least

$$\frac{\frac{2n}{5}|S|}{\binom{n}{2}} \geq \frac{4\alpha}{5}.$$

Hence

$$\mathbb{P}\left[\left|E_{G_m}\left(S, V \setminus S\right)\right| \leq 10\alpha\beta m \mid \mathcal{B}\right] \leq \mathbb{P}\left[\mathrm{Bin}\left(\frac{9m}{10}, \frac{4\alpha}{5}\right) \leq 10\alpha\beta m\right] \leq e^{-\frac{1}{2}\alpha m}.$$

The last inequality follows from Lemma 3.6 (applied with $\mu = 36\alpha m/50$, $\eta = 11/36$ and $\delta = 1/72$). Taking the union bound over all sets of order αn

$$\mathbb{P}\left[\mathcal{A} \mid \mathcal{B}\right] \leq \binom{n}{\alpha n} \cdot e^{-\frac{1}{2}\alpha m} \leq \left(e\alpha^{-1}e^{-\frac{m}{2n}}\right)^{\alpha n} = \left(e^{-\frac{m}{6n}+1}\right)^{\alpha n} \leq e^{-\frac{\alpha m}{10}}.$$

Since $n \leq m \leq 2n \log n$, a very crude computation gives $\alpha m = e^{-m/(3n)}m \geq e^{-2 \log n/3}m \geq n^{1/3}$, so $\mathbb{P}\left[\mathcal{A}\right] \leq e^{-\alpha m/10} \leq n^{-2}$, completing the proof of (4.1) and hence the proof of the claim. $\qquad\square$

Now consider applying the following algorithm to G_m. Let $H_0 := G_m$ and $R_0 := \emptyset$. Now for each $i \geq 0$, if there exists some $v \in H_i$ such that $\deg_{H_i}(v) < \frac{10\beta m}{n}$, then define $H_{i+1} := H_i \setminus \{v\}$ and $R_{i+1} := R_i \cup \{v\}$. We terminate the algorithm when no such v exists, and denote the final step by j. Claim 4.3 implies that $\mathbb{P}\left[j \geq \alpha n\right] = O\left(n^{-2}\right)$. For, if $R_{\alpha n}$ is defined, then $|E_{G_m}(R_{\alpha n}, V \setminus R_{\alpha n})| \leq 10\alpha\beta m$ by construction. This completes the proof of (i).

To show (ii), assume that part (i), as well as the conclusion of Lemma 3.5 hold. Let H be the subgraph given by (i), and let $S \subseteq V(H)$ be a set of cardinality $|S| \leq \frac{\beta}{16}n$. Let $T = N_H(S) \cup S$ and suppose that $|T| \leq 2|S| \leq \frac{\beta}{8}n$. By Lemma 3.5 and the minimum-degree assumption about H, we have

$$\frac{\beta m}{n}|T| \geq |E_{G_m}(T)| \geq \frac{10\beta m}{2n}|S|,$$

and so $|T| \geq 5|S|$, a contradiction. So $|T| > 2|S|$, as required for (ii). $\qquad\square$

We are now ready to present the proof of Lemma 2.1.

Proof of Lemma 2.1. Let $0 < \beta \leq 10^{-3}$ and $\zeta := 2^{10}\beta^{-2}$. Given $m \geq \gamma n$ and γ sufficiently large, define $m' = m - \zeta n \geq \frac{9m}{10}$. Let $\mathcal{B}$ be the event that statements (i) and (ii) of Lemma 4.2 hold for m' and let γ be large enough to ensure $\mathbb{P}(\mathcal{B}) = 1 - O\left(n^{-2}\right)$ (possible by Lemma 4.2). We show that

$$\mathbb{P}\left[\nu(G_m) < \frac{n}{2}\left(1 - e^{-3m/(10n)}\right)\ \middle|\ \mathcal{B}\right] = O\left(n^{-2}\right), \tag{4.2}$$

which will imply the lemma. In what follows, we condition on $\mathcal{B}$ occurring.

Hence there exists a set L of vertices which span a subgraph of minimum degree at least $\frac{10\beta m'}{n}$ in $G_{m'}$, such that $n - |L| \leq n e^{-m'/3n} \leq n e^{-3m/10n}$. We may ensure L has an even number of vertices by removing an arbitrary vertex if $|L|$ is odd. Define $H_i := G_i[L]$ for each $i \geq m'$. We use an expansion argument due to Bollobás and Frieze [6] to show that H_m contains a perfect matching. We will include the proof since the claim is not stated explicitly in [17].

Claim 4.4 ([17]). *Let H be an n-vertex graph in which every $S \subseteq V(H)$ with $|S| \leq k$ satisfies $|N_H(S)| > |S|$. If H does not have a perfect matching, then there are at least $\binom{k}{2}$ vertex pairs f such that $\nu(H + f) > \nu(H)$.*

Proof. Let M be a maximum matching in H and v a vertex not contained in $V(M)$. In this case, we say that M *isolates* v. Using a sequence of *flips*, we will find many vertex pairs uv which extend M. For $e = xx' \in M$ and $f = xu$, where $u \neq v$ is isolated by M, a *flip* from e to f is the operation of replacing M by $M' = M + f - e$. Let $A(v)$ be the set of vertices y such that v and y are isolated by some matching M' obtained from M by a sequence of flips.

We will show that any vertex in $N_H(A(v))$ is matched in M to some vertex of $A(v)$, implying that $|N_H(A(v))| \leq |A(v)|$. Let $x \in N_H(A(v))$. In particular $x \notin A(v)$ and so $x \in M$. Let x' denote the unique neighbour of x in M. We will show the vertex x' is in $A(v)$. This will imply that $|N_H(A(v))| \leq |A(v)|$. Since $x \in N_H(A(v))$ there is a vertex $u \in A(v)$ adjacent to x. Let M' be a maximum matching obtained from M by a sequence of flips that isolates u. First, suppose that $xx' \in M'$. Then we can flip the edge xx' with the edge xu, isolating x'. Thus $x' \in A(v)$. If xx' is not in M', then at some point in the sequence of flips from M to M' a flip from xx' to another edge has occurred. If this happens, then either x or x' is in $A(v)$. By assumption, $x \notin A(v)$, so in fact $x' \in A(v)$. Hence $|N_H(A(v))| \leq |A(v)|$, as required.

Our assumption on H in the claim hypothesis implies that $|A(v)| \geq k$. Applying the same argument to all vertices $x \in A(v)$ (which are isolated by some maximum matching M') gives us at least $\frac{k^2}{2}$ distinct pairs xy such that $\nu(H + xy) > \nu(H)$. $\qquad\square$

Consider any step i with $m' < i \leq m$ and $\nu(G_i) < \frac{n}{2}$. As we condition on $\mathcal{B}$, we may apply Claim 4.4 with $k = \frac{\beta}{16}n$ to get that there are at least $\binom{\beta n/16}{2}$ vertex pairs f such that $\nu(G_i + f) > \nu(G_i)$. As $i \leq 2n \log n$, only $o(n^2)$ vertex pairs have been offered so far. Therefore the probability that $\nu(G_{i+1}) > \nu(G_i)$ is at least the probability that such an f is offered, which is

$$(4.3) \qquad \frac{\binom{\beta n/16}{2} - i}{\binom{n}{2} - i} \geq \frac{\beta^2}{2^9}.$$

Let $Y \sim \mathrm{Bin}\left(\zeta n, \frac{\beta^2}{2^9}\right)$. We have

$$\mathbb{P}\left[H_m \text{ has no perfect matching} \mid \mathcal{B}\right] \leq \mathbb{P}\left[Y < \frac{n}{2}\right] = e^{-\Omega(n)} \leq n^{-2}.$$

This completes the proof of the lemma.

$\qquad\square$

Analogous arguments actually give a stronger result, which will be used to prove Theorem 1.2(i).

Corollary 4.5. *Let $\delta > 0$, $m \gg n$. Let L be a fixed vertex subset of even order. If $G_{m-\omega(n)}[L]$ has minimum degree at least $\frac{\delta m}{n}$, then $G_m[L]$ has a perfect matching with high probability.*

Proof. Let L be a vertex subset of even order such that $H := G_{m-\omega(n)}[L]$ has minimum degree at least $\frac{\delta m}{n}$. Then, analogously to the argument in 4.2(ii), we obtain that, with high probability, any subset $S \subseteq V(H)$ with $|S| \leq \frac{\delta n}{160}$ satisfies $|N_H(S) \cup S| > 2|S|$. Now, as in the proof of Lemma 2.1, using Claim 4.4 we see that $G_m[L]$ has a perfect matching with high probability. $\qquad\square$

To conclude the section, we show that G_m has a perfect matching for $m = (1 + \epsilon)n \log n$ to prove Theorem 1.2(i). It is not difficult to show the claim for $m = (1 + o(1))n \log n$ by controlling distances between 'low-degree' vertices in G_m (see, e.g., [17]), but we chose to include the slightly weaker statement, which is restated here for the benefit of the reader.

Theorem 1.2(i). *Let $\epsilon > 0$ and $m = (1 + \epsilon)n \log n$. With high probability, G_m has a perfect matching.*

Proof. Let $t := \frac{\epsilon}{4}n \log n$. We will first show that with high probability, G_{m-t} has minimum degree at least $\delta \log n$.

By Corollary 4.1, if $t \leq m' \leq 2n \log n$, each vertex is in at least $\left(\frac{1}{2} - o(1)\right) n > \left(\frac{1}{2} - \frac{\epsilon}{40}\right) n$ vertex pairs which are acceptable for $G_{m'}$. Therefore, for all $k \geq 0$, we have

$$\mathbb{P}\left[\deg_{G_{m-t}}(v) \leq k\right] \leq \mathbb{P}\left[X \leq k\right],$$

where $X \sim \text{Bin}\left(m - 2t, \left(\frac{1}{n} - \frac{\epsilon}{20n}\right)\right)$. Let $\delta \leq 1$ be small enough for Lemma 3.6 to hold with $\eta = \frac{\epsilon}{20}$. By the choice of δ and applying Lemma 3.6, we have

$$\mathbb{P}\left[\deg_{G_{m-t}}(v) < \delta \log n\right] \leq \exp\left(-\left(1 - \frac{\epsilon}{20}\right)(m - 2t)\left(\frac{1}{n} - \frac{\epsilon}{20n}\right)\right) \leq n^{-1-\frac{\epsilon}{10}}.$$

Taking the union bound over all n vertices gives that with high probability, G_{m-t} has minimum degree at least $\delta \log n$.

As $t = \omega(n)$, we can apply Corollary 4.5 to G_{m-t} with $L = [n]$ and deduce that G_m has a perfect matching. $\qquad\square$

5. The structure of G_N

The main purpose of this section is to prove Theorem 1.2(iii). That is, we bound the number of rejected vertex pairs incident to an optimal cover of G_N. Let C_N denote an optimal cover of G_N. Let $m = 1.1n \log n$. When G_m has a perfect matching (which it does with high probability by Theorem 1.2(i)), C_N will also be an optimal cover in G_m and any edge incident to it will be accepted. So it suffices to control the edges of G_m^{all} rejected from an optimal cover C_m of G_m when $m = 1.1n \log n$. This is done in Lemma 5.3. This lemma and the previous observation immediately imply Theorem 1.2(iii). We note that, in particular, our results in this section do not rely on uniqueness of C_N (Theorem 1.2(ii)).

We start by introducing some concepts that will be used throughout the section. Recall that $D_m \subseteq V(G_m)$ is the set of vertices contained in some optimal cover of G_m. For a time period $T := T(m_1, t)$ and vertex $v \in V$, define the *weight* $W_T(v)$ of a vertex as

$$W_T(v) := |\{m \in T : v \in D_m\}|.$$

Note that $W_T(v)$ is a function of our random process. For a set $S \subseteq V$, define the *average weight* of S in T as

$$W_T(S) := \frac{1}{|T|} \sum_{v \in S} W_T(v).$$

The main ingredient in the proof of Lemma 5.3 is the following lemma (Lemma 5.1). It is proved using a martingale trick similar to one that will be used in Lemma 6.1. The main difference is that here we apply Freedman's inequality (Corollary 3.9), whereas there we apply Azuma's inequality. This is because Azuma's inequality considers the worst-case change of a martingale (X_m). In our case the *typical* changes are much smaller. Therefore Freedman's inequality gives a stronger bound, which is also necessary for the computations.

Lemma 5.1. *For $\gamma > 0$, let $m_1 \ll t := \gamma n \log n$ and let $T := T(m_1, t)$. The following holds with high probability.*

(i) *No set S of order at least $\frac{n}{3}$ with $W_T(S) \geq \frac{150n^2}{t}$ is independent in G_{m_1+t}.*

(ii) *Let $a \geq 10$. For any set U of order $\frac{n}{2}$ with $W_T(U) \geq \frac{n}{2} - \frac{an^2}{t}$, the number of edges incident to U which were rejected during $T + 1 := T(m_1 + 1, t)$ is at most $5an$.*

Proof of Lemma 5.1. Let $m \geq m_1$ and let $S \subseteq V$ be a set of order at least $\frac{n}{3}$. For each m, define Q'_{m+1} to be the set of vertex pairs in $S^{(2)}$ which are acceptable for G_m. Let s be the maximum integer such that

$$(5.1) \qquad \sum_{i=m_1+1}^{s} \frac{|Q'_i|}{\binom{n}{2} - i + 1} \leq 79n.$$

For $m_1 + 1 \leq j \leq s$, define $Q_j := Q'_j$, and for $j > s$ define $Q_j := \emptyset$. The reason we truncate the sequence $(Q'_j)_{j \geq m_1+1}$ in this manner is to deal with a technicality in our application of Freedman's inequality.

Let X_m be the indicator random variable of the event that $e_m \in Q_m$. Moreover, define

$$q_{m+1} := \mathbb{P}_m[X_{m+1} = 1] = \frac{|Q_{m+1}|}{\binom{n}{2} - m},$$

so that by definition we have

$$\sum_{m \in T} q_{m+1} \leq 80n.$$

Given this, by applying Corollary 3.9 with $L = 20$, we see that

$$(5.2) \qquad \mathbb{P}\left[\left| \sum_{m \in T} (X_{m+1} - q_{m+1}) \right| \geq 20n \right] \leq e^{-2n}.$$

Let $\mathcal{A}$ be the event that $W_T(S) \geq \frac{150n^2}{t}$. Let $\mathcal{B}$ be the event that $X_{m+1} = 0$ for all $m \in T$. We will show that $\mathbb{P}[\mathcal{A} \wedge \mathcal{B}] \leq e^{-2n}$.

To use the condition on $W_T(S)$, we will need a simple relation between q_m and $W_T(S)$.

Claim 5.2. *If $\mathcal{A}$ occurs, then $\sum_{m \in T} q_{m+1} \geq 40n$.*

Proof. Any vertex pair not contained in G_m^{all} but intersecting $S \cap D_m$ is acceptable for G_m. Therefore, we have $|Q_{m+1}| \geq \frac{1}{2}|S \cap D_m|(|S| - 1) - m$. Summing over m and using $\sum_{m \in T}|S \cap D_m| = tW_T(S)$, which is a restatement of the definition of $W_T(S)$, we get

$$\sum_{m \in T} q_{m+1} \geq \sum_{m \in T} \frac{2|Q_{m+1}|}{n^2} \geq \frac{tW_T(S)|S|}{n^2} - O\left(\frac{t^2}{n^2}\right).$$

By the claim assumption and the definition of t, this is at least $50n - O(\log^2 n) \geq 40n$, and the claim follows. $\square$

By Claim 5.2, if both $\mathcal{A}$ and $\mathcal{B}$ occur, then

$$\left|\sum_{m \in T}(X_{m+1} - q_{m+1})\right| = \sum_{m \in T} q_{m+1} \geq 40n.$$

So by (5.2), $\mathbb{P}[\mathcal{A} \wedge \mathcal{B}] \leq e^{-2n}$ as required. Observing that $\mathcal{B}$ holds if S is independent in G_{m_1+t} and taking the union bound over S, the probability that (i) does not hold is at most $2^n e^{-n}$. This completes the proof of (i).

The proof of (ii) follows very similarly. Let U be a vertex set of order $\frac{n}{2}$. Now, for $m \geq m_1$, let R_{m+1} be the set of vertex pairs intersecting $U \setminus D_m$ and let $r_{m+1} := \frac{|R_{m+1}|}{\binom{n}{2} - m}$. If $W_T(U) \geq \frac{n}{2} - \frac{an^2}{t}$, then

$$\sum_{m \in T} r_{m+1} \leq \sum_{m \in T} \frac{4|R_{m+1}|}{n^2} \leq \sum_{m \in T} \frac{4n(|U| - |U \cap D_m|)}{n^2} = \frac{4t}{n}\left(\frac{n}{2} - W_T(U)\right) \leq 4an.$$

Let Z_m be the indicator random variable of the event $e_m \in R_m$. Then $\mathbb{P}_m[Z_{m+1} = 1] = r_{m+1}$. Note that the edge e_m, which is incident to U, can be rejected only if $e_m \in R_m$. Therefore $\sum_{m \in T} Z_{m+1}$ counts the number of rejected edges incident to U. Analogously to part (i), applying Corollary 3.9 to r_m with the constant $L = a$ gives

$$\mathbb{P}\left[\left(\sum_{m \in T} Z_{m+1} \geq 5an\right) \wedge \left(W_T(U) \geq \frac{n}{2} - \frac{an^2}{t}\right)\right] \leq e^{-an/10}.$$

Taking the union bound over all possible U and recalling the hypothesis $a \geq 10$, we get that (ii) holds with probability at least $1 - 2^n e^{-n} = 1 - o(1)$.

$\square$

We now apply Lemma 5.1 to control rejected edges adjacent to an optimal cover and use this to prove Lemma 5.3. As discussed above, this immediately implies Theorem 1.2(iii).

Lemma 5.3. *Let $m := 1.1n\log n$. With high probability, if C is any optimal cover in G_m, then $O(n)$ edges of G_m^{all} incident to C have been rejected.*

Proof. Set $t = m - \gamma_0 n$ and $T := T(\gamma_0 n, t)$, where γ_0 is a large constant chosen so that the conclusion of Lemma 2.1 holds. Namely, with probability $1 - O(n^{-2})$,

$$(5.3) \qquad\qquad \nu(G_\ell) \geq \frac{n}{2}\left(1 - e^{3\ell/(10n)}\right) \quad \text{for} \quad \ell \in T.$$

We have $|D_\ell| \geq \nu(G_\ell)$. So (5.3) implies

$$(5.4) \qquad W_T(V) \geq \frac{n}{2} - \frac{n}{2t} \sum_{\ell \in T} e^{-3\ell/(10n)} \geq \frac{n}{2} - \frac{ne^{-3\gamma_0 n/(10n)}}{2t\left(1 - e^{-3/(10n)}\right)} \geq \frac{n}{2} - \frac{20n^2}{t},$$

where we used that $1 - e^{-x} \geq x/2$ for $x \leq 1/2$. As $V \setminus C$ is an independent set in G_m, Lemma 5.1 (i) implies that with high probability $W_T(V \setminus C) \leq \frac{150n}{\log n}$. Combining this with (5.4) gives that $W_T(C) \geq \frac{n}{2} - \frac{200n}{\log n}$ with high probability. Now applying Lemma 5.1 (ii) to C gives that, with high probability, the number of edges incident to C rejected during T is $O(n)$.

Clearly at most $\gamma_0 n$ vertex pairs are rejected before time $\gamma_0 n$. This completes the proof of the lemma and of Theorem 1.2(iii). □

6. Rigidity and uniqueness of an optimal cover

We now turn our attention to analysing optimal covers throughout our process. In this section we will first prove Lemma 2.2, which concerns the 'rigidity' of an optimal cover, and Theorem 1.2(ii), which tells us that with high probability, $G_{4n \log n}$ has a unique optimal cover.

We start with an elementary observation. For $m \geq 0$, conditioned on $(e_i)_{i \leq m}$, let p_{m+1} be the probability that e_{m+1} is acceptable for G_m. Recall that e_{m+1} is acceptable whenever it is incident to an optimal vertex cover. So for $0 \leq m \leq 2n \log n$, if $\nu(G_m) = \tau(G_m)$ is of order at least $n\left(\frac{1}{2} - r(n)\right)$, then

$$(6.1) \quad p_{m+1} \geq \frac{\binom{n}{2} - \binom{(\frac{1}{2}+r)n}{2} - m}{\binom{n}{2} - m} \geq \frac{n^2 - n - \left(\frac{1}{2} + r\right)^2 n^2 - 4n \log n}{n^2} \geq \frac{3}{4} - 2r - \frac{5 \log n}{n}.$$

A key ingredient for the proof of Lemma 2.2 is the following lower bound on the number of edges accepted into our graph during a certain time period.

Lemma 6.1. *For $\gamma > 0$, let $T := T(m_0, \gamma n \log n)$. Let G be the graph consisting of all edges accepted into (G_m) during the period T. With high probability,*

$$(6.2) \qquad |E(G)| \geq \sum_{m \in T} p_m - n\sqrt{\log n}.$$

Proof. For $m \geq 0$, define X_m to be the indicator random variable of the event that e_m is accepted and $Y_m := X_m - p_m$. By definition of p_m, we have $\mathbb{E}_{m-1}[Y_m] = 0$, so $\left(\sum_{i=m_0}^{m'} Y_i\right)$ is a martingale. Set $Y := \sum_{m \in T} Y_m$. Moreover, $|Y_m| \leq 1$ for each m, so we can apply Azuma's inequality (Theorem 3.7) with $\lambda = n\sqrt{\log n}$. We get

$$\mathbb{P}\left[Y < -n\sqrt{\log n}\right] \leq \exp\left(-\frac{n^2 \log n}{2\gamma n \log n}\right).$$

It follows that, with probability $1 - e^{-\Omega(n)}$,

$$|E(G)| = \sum_{m \in T} X_m \geq \sum_{m \in T} p_m - n\sqrt{\log n}.$$

□

Call a time step m φ-*flexible* if $|D_m| \geq \left(\frac{1}{2} + \varphi\right) n$. In all our arguments, this notion will be used with the same function $\varphi = \varphi(n) = (\log n)^{-1/10}$. We now prove Lemma 2.2. Our proof relies on the statistics of the process – informally, if many time steps were φ-flexible, we would be accepting too many edges.

Proof of Lemma 2.2. Recall that $m_0 = n\sqrt{\log n}$, $T = T(m_0, \gamma n \log n)$ and $\varphi = (\log n)^{-1/10}$. For ease of notation, let $t := |T| = \gamma n \log n$. Let us assume, in order to obtain a contradiction, that the number of φ-flexible steps is greater than φt. Let G be the graph consisting of edges accepted during the interval T, and $H := G_{m_0+t}^{\text{all}}$. Note that H is distributed as G_{n,m_0+t}.

By Lemma 2.1, we know that with high probability $\nu(G_{m_0})$ contains a matching of size $\frac{n}{2}(1 - r)$, where $r := r(n) = e^{-3m_0/(10n)} \ll \log^{-1} n \ll \varphi$. As this property is increasing, we have that G_m with $m \geq m_0$ also contains a matching of at least this size. Thus, using (6.1), with high probability for all $m \geq m_0$, we have $p_m \geq \frac{3}{4} - 3r$. Moreover, if the step m is flexible, we have a stronger bound $p_m \geq \frac{3}{4} + \frac{\varphi}{2}$.

By applying Lemma 6.1 and the above analysis, with high probability we have

$$(6.3) \qquad |E(G)| + n\sqrt{\log n} \geq \sum_{m \in T} p_m \geq \frac{3t}{4} + \frac{\varphi}{2} \cdot \varphi t - 3r(1 - \varphi)t \geq \left(\frac{3}{4} + \frac{\varphi^2}{4}\right) t.$$

Let $m_1 := m_0 + t = t(1 + O((\log n)^{-1/2}))$. Let C be an optimal vertex cover in G_{m_1}, which has size at least $\frac{n}{2}(1 - \log^{-1} n)$. As $V \setminus C$ is an independent set of size at most $\frac{n}{2}(1 + \log^{-1} n)$, by applying Theorem 3.4 to graph $H \sim G_{n,m_1}$ we obtain

$$|E(G_{m_1})| \leq |E_H(C)| + |E_H(V \setminus C, C)| \leq \left(\frac{3}{4} + O\left((\log n)^{-1/2}\right)\right) m_1$$

$$= \left(\frac{3}{4} + O\left((\log n)^{-1/2}\right)\right) t.$$

As $G_{m_1} \supseteq G$, this contradicts (6.3). This completes the proof of Lemma 2.2. $\qquad \square$

We finish the section by proving Theorem 1.2(ii), which will be restated here to aid the reader. The key idea is to analyse the set of rejected pairs and to use these to obtain information about an optimal cover.

Theorem 1.2(ii). With high probability $G_{4n \log n}$ has a unique optimal vertex cover.

Proof. Set $m_0 := \frac{5}{4} n \log n$, $m_1 := \frac{3}{2} n \log n$ and $m_2 := 4n \log n$. By Theorem 1.2(i), with high probability G_{m_0} has a perfect matching and hence an optimal cover of cardinality $n/2$. For each $i \geq 0$, let $\mathcal{C}_i$ be the set of optimal covers of G_i. Observe that for all $i \geq m_0$, we have $\mathcal{C}_{i+1} \subseteq \mathcal{C}_i$, as adding edges can only eliminate optimal covers.

Let $E_0 := \{e_1, \ldots, e_{m_0}\}$ and $E_1 := \{e_{m_0+1}, \ldots, e_{m_1}\}$. Note that $G' := G_{m_1}^{\text{all}} \setminus E_0 \sim G_{n,m'}$, where $m' = \frac{n}{4} \log n$. So by Theorem 3.4, with high probability, in $G_{n,m'}$ every set of cardinality $n/2$ contains at least $(1 - O((\log n)^{-1/2}))\frac{n}{16} \log n$ edges. However, G_{m_1} contains an independent set of cardinality at least $n/2$. So, with high probability, at least $(1 - O((\log n)^{-1/2}))\frac{n}{16} \log n$ pairs of E_1 are rejected. Let U be the set of vertices spanning these rejected pairs.

We will next show that $|U| \geq (1 - (\log n)^{-1/4})\frac{n}{2}$. Suppose, for a contradiction, that $|U| < (1 - (\log n)^{-1/4})\frac{n}{2}$. Then applying Theorem 3.4 shows that in $G' \sim G_{n,m'}, m' = \frac{n}{4} \log n$,

with high probability we have

$$|E(U)| \leq \frac{1}{16} n \log n - \Omega(n(\log n)^{3/4}) < (1 - O((\log n)^{-1/2}))\frac{1}{16} n \log n,$$

a contradiction for n sufficiently large.

Now observe that if $(u,v) \in E_1$ is rejected, then there exists no $C \in \mathcal{C}_i$ such that $u \in C$ or $v \in C$. This together with our observation from the first paragraph shows that, in fact, every cover in $\mathcal{C}_{m_1}$ contains neither u nor v. So in particular, no vertex of U is contained in a cover in $\mathcal{C}_{m_1}$ (or $\mathcal{C}_{m_2}$).

Now let $E_2 := \{e_{m_1+1}, \ldots, e_{m_2}\}$ and consider $e := (u,v) \in E_2$ such that $u \in U$ and $v \notin U$. As in the previous paragraph, if e is rejected then no cover in $\mathcal{C}_{m_2}$ contains u or v. However, if e is accepted, since (by the previous paragraph) u is in no cover of $\mathcal{C}_{m_2}$, then every cover in $\mathcal{C}_{m_2}$ contains v. So for each $v \in V \setminus U$, let $\mathcal{E}_v$ be the event that E_2 contains a pair (u,v) for some $u \in U$. If $\mathcal{E}_v$ occurs, then we know that either v is contained in every cover or no cover in $\mathcal{C}_{m_2}$.

So if, $\mathcal{A} := \bigcup_{v \in V \setminus U} \mathcal{E}_v$ occurs, then $\mathcal{C}_{m_2}$ contains a unique optimal cover. It remains to show that $\mathcal{A}$ occurs with high probability. For a particular $v \in V$, say that e_i is *good* for v if $e_i = (u,v)$ for some $u \in U$. Let $d_i(v)$ be the degree of v in G_i^{all}. For $i \leq m_2$, the probability that e_i is good for v is at least

$$\frac{|U| - d_i(v)}{\binom{n}{2} - i + 1} \geq \frac{(1 - o(1))}{n},$$

as $|U| \geq (1 - (\log n)^{-1/4})n/2$ and, by Claim 3.2 we have $d_i(v) = O(\log n)$, as $G_i^{\mathrm{all}} \subseteq G_{10n \log n}^{\mathrm{all}}$. So the number of pairs in E_2 that are good for v is at least X, where $X \sim \mathrm{Bin}\left(\frac{5}{2}n \log n, \frac{1-o(1)}{n}\right)$. We have

$$\mathbb{P}[X = 0] = \left(1 - \frac{1 - o(1)}{n}\right)^{\frac{5}{2}n \log n} \leq e^{-2 \log n}.$$

Hence the probability that $\mathcal{E}_v$ does not occur is at most $\frac{1}{n^2}$. Applying the union bound over the vertices in $V \setminus U$ gives that $\mathcal{A}$ occurs with high probability, as required. $\square$

7. Delayed perfect matching threshold

The focus of this section is to prove Theorem 1.3, which says that for $m = \left(\frac{1}{2} + \frac{1}{70}\right) n \log n$, G_m typically has isolated vertices. Throughout this section, we set $m_2 := \frac{3}{8}n \log n$ and $m_3 := 2n \log n$. We assume that G_{m_2} has a subgraph H which has a perfect matching and satisfies

$$(7.1) \qquad |V(H)| \geq n - ne^{-\frac{3m_2}{10n}} \geq n\left(1 - n^{-\frac{1}{20}}\right) \quad \text{and} \quad \delta(H) \geq 10^{-3} \log n,$$

as by Lemma 4.2 and Corollary 4.5 this event occurs with high probability.

In order to show that our graph does not contain a perfect matching, we will carefully track the number of vertices with at most one neighbour in G_m. In order to understand how these vertices are used at a particular time step to extend a current maximum matching, we require some information on separation of small-degree vertices in G_m (see Lemma 7.1 below). One important consequence will be that, typically in G_m, no vertex has three neighbours of degree one. For $u, v \in V$, let $\mathrm{dist}_m(u,v)$ be the distance from u to v in G_m and let $\mathrm{dist}_m^{\mathrm{all}}(u,v)$ be

the distance from u to v in G_m^{all}. We take the convention $\text{dist}(u,u) = 0$ for any underlying graph and any vertex u. The following lemma will mostly be applied with $a = 3$.

Lemma 7.1. *For sufficiently small $\epsilon > 0$, there exists $\delta > 0$ such that with high probability the following statement holds. For a positive integer $a \leq 5$ and all $(1+\epsilon)n \log n/a \leq m \leq m_3$, there are no distinct vertices $u_1, \ldots, u_a$ such that, for all $i, j \in [a]$, we have $\deg_{G_m}(u_i) < \delta \log n$ and $\text{dist}_m(u_i, u_j) \leq 10$.*

Proof. Let $\ell = (1+\epsilon)n \log n/a$, and let $\mathcal{A}$ be the event that there exist $u_1, \ldots, u_a$ such that for all $i, j \in [a]$, we have $\deg_{G_\ell}(u_i) \leq \delta \log n$ and $\text{dist}_{m_3}^{\text{all}}(u_i, u_j) \leq 10$. Since for any $\ell \leq m \leq m_3$ and any $i, j \in [a]$, we have $\deg_{G_m}(u_i) \geq \deg_{G_\ell}(u_i)$ and $\text{dist}_m(u_i, u_j) \leq \text{dist}_{m_3}(u_i, u_j) \leq \text{dist}_{m_3}^{\text{all}}(u_i, u_j)$, it suffices to show that $\mathbb{P}[\mathcal{A}] = o(1)$.

Let $U := \{u_1, \ldots, u_a\} \subseteq V$ and define

$$D := \sum_{i=1}^{a} \deg_{G_\ell}(u_i).$$

If $\text{dist}_{m_3}^{\text{all}}(u_i, u_j) \leq 10$ for all $i, j \in [a]$, then there is a connected subgraph of $G_{m_3}^{\text{all}}$ on at most $11 \cdot (a-1) < 50$ vertices containing U. In particular, this subgraph has a spanning tree. Let $\mathcal{T}_F$ be the event that $G_{m_3}^{\text{all}}$ contains a fixed labelled tree F with $U \subseteq V(F)$ and $|V(F)| := f < 50$.

First we prove an upper bound on $\mathbb{P}[D < a\delta \log n \mid \mathcal{T}_F]$. Let $T' := \{i : e_i \in E(F)\}$ and $T := T(\gamma n, \ell - \gamma n)$. Let Y_m be the indicator random variable of the event that e_m is an acceptable pair with one endpoint in U and the other in $V \setminus F$. Using Corollary 4.1 (ii), for γ sufficiently large and any $\gamma n + 1 \leq m \leq \ell$ such that $m \notin T'$ we have that there are at least $a(1/2 - \epsilon/4)n$ acceptable pairs touching one of the vertices in U. Therefore

$$\mathbb{P}_{m-1}[Y_m = 1] \geq \frac{a(1/2 - \epsilon/4)n}{(1 + o(1))\binom{n}{2}} \geq a \cdot \frac{1 - \epsilon/2}{n}.$$

Thus, letting $X \sim \text{Bin}\left(\ell(1 - o(1)), \frac{a(1-\epsilon/2)}{n}\right)$, we have

$$\mathbb{P}\left[\sum_{m \in T \setminus T'} Y_m < k \mid \mathcal{T}_F\right] \leq \mathbb{P}[X < k].$$

Noting that $\mathbb{E}[X] = \frac{a(1-\epsilon/2)}{n} \cdot \ell(1 - o(1)) > \left(1 + \frac{\epsilon}{4}\right) \cdot \log n$ and applying Lemma 3.6 with δ sufficiently small and $\eta = \frac{\epsilon}{8}$, we get

$$\mathbb{P}[D < a\delta \log n \mid \mathcal{T}_F] \leq \mathbb{P}\left[\sum_{m \leq \ell} Y_m \leq \delta \cdot a \log n \mid \mathcal{T}_F\right] \leq e^{-\left(1 + \frac{\epsilon}{10}\right) \log n}.$$

For a fixed tree F we now show that $\mathbb{P}[\mathcal{T}_F] = O\left(\left(\frac{\log n}{n}\right)^{f-1}\right)$. To see this, note that the probability that the random graph $G(n, p)$ with $p = m_3 \binom{n}{2}^{-1}$ contains F is precisely p^{f-1}. Using Lemma 3.1, the fact that $\mathcal{T}_F$ is a monotone increasing event and the crude estimate $p < \frac{5 \log n}{n}$, we get

$$\mathbb{P}[\mathcal{T}_F] \leq 3\left(\frac{5 \log n}{n}\right)^{f-1} = O\left(\left(\frac{\log n}{n}\right)^{f-1}\right).$$

Putting this together with the previous bound gives

$$\mathbb{P}\left[(D < a\delta \log n) \wedge \mathcal{T}_F\right] = n^{-1-\frac{\epsilon}{10}} \cdot O\left(\left(\frac{\log n}{n}\right)^{f-1}\right) = o\left(n^{-f}\right).$$

As for fixed f there are $O\left(n^f\right)$ choices for F and $U \subseteq V(F)$, we can take the union bound over U and F to get $\mathbb{P}\left[\mathcal{A}\right] \leq \mathbb{P}\left[D < a\delta \log n\right] = o(1)$. □

Given this lemma, and noting that $m_2 = \frac{3}{8}n \log n > (1+\epsilon)n \log n/3$ for a positive constant ϵ, we may assume for the rest of the proof that whenever $m \geq m_2$, no vertex in G_m has three neighbours of degree one. We are now able to introduce the specific definitions we need for the main part of our proof.

Say that a vertex is an *isolate* or *is isolated* in G_m if it has no neighbours in G_m. If there exist two vertices v_1, v_2 of degree one in G_m that share a neighbour, then we arbitrarily choose one to be a *quasi-isolate* and the other to be its *partner* (by the previous paragraph there cannot typically be a third vertex of degree one sharing a neighbour with v_1 and v_2). The partner is *not* a quasi-isolate. Let J_m denote the set of vertices that are either isolates or quasi-isolates in G_m.

Define M_m to be a maximum matching in G_m with $V(M_m) := C_m \cup B_m$, where C_m is an optimal vertex cover in G_m, chosen subject to the following:

(M1) If $v \in C_m$ is matched to $u \in B_m$ by M_m and has a neighbour $w \notin V(M_m)$, then $\deg_{G_m}(w) \leq \deg_{G_m}(u)$.

(M2) M_m contains no quasi-isolate.

Achieving (M1) is possible by swapping u for w if this is not the case (this will give another matching of the same size). Similarly, by swapping a quasi-isolate with its partner, (M2) is achievable. So such a matching M_m always exists. Note that there may be many choices for such an M_m.

By definition of M_m, the set J_m is disjoint from $V(M_m)$. Define the set of *helpers* to be $H_m := V \setminus (V(M_m) \cup J_m)$ and observe that the number of helpers $|H_m|$ is independent of the particular choice of M_m. Intuitively, the helpers are vertices attached to M_m so as to create many vertex pairs which would extend M_m.

Our aim is to show that typically G_m contains no helpers for a constant proportion of time steps m (Lemma 7.4). In other words, there is a matching covering all vertices apart from isolates and quasi-isolates. At such steps, the rate of losing isolated vertices is slower than in G_m^{all}, so Theorem 1.3 will follow. The argument consists of two main lemmas. The first one is an analogue of the hitting time result of Bollobás and Thomason [7] for our setting.

Lemma 7.2. *With high probability, the number of steps $m \geq m_2$ at which $|H_m| \geq 2$ is $O\left(n^{1-1/20}\right)$.*

The second lemma tells us that typically at many time steps we have a non-zero even number of isolates or quasi-isolates.

Lemma 7.3. *With high probability, for at least $\frac{n \log n}{33}$ steps $m \geq m_2$ we have $|J_m| > 0$ and $|J_m|$ is even.*
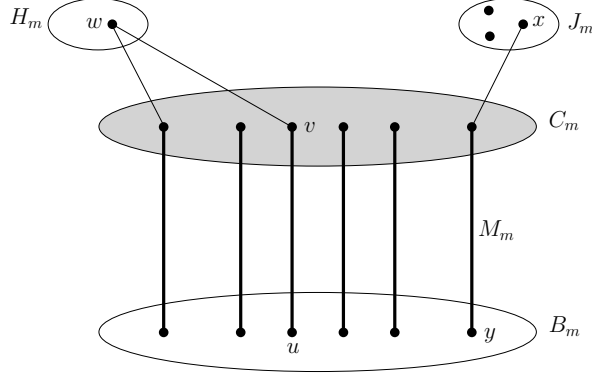
FIGURE 1. An illustration of helpers, isolates, quasi-isolates and the matching M_m. The vertex x is a quasi-isolate and y is its partner, so both have degree 1 in G_m. By (M1), $\deg(w) \leq \deg(u)$. Recall that all additional edges are incident to C_m.

The main work in this section is devoted to proving these two lemmas. Lemma 7.2 will be proved in Section 7.1, and Lemma 7.3 will be proved in Subsection 7.2. Before doing this, let us first show how the two lemmas imply the desired result.

Lemma 7.4. *With high probability, for at least $\frac{n \log n}{34}$ time steps $m \geq m_2$, we have $H_m = \emptyset$ and $J_m \neq \emptyset$.*

Proof. Since the number of vertices n is even, $|J_m| + |H_m| = n - |V(M_m)|$ is always even. In other words, $|J_m|$ and $|H_m|$ have the same parity, so Lemma 7.3 gives that with high probability, $|H_m|$ is even and $J_m \neq \emptyset$ for at least $\frac{n \log n}{33}$ steps m. However, Lemma 7.2 shows that, with high probability, $|H_m| \geq 2$ for $o(n \log n)$ steps. The result follows. $\qquad\square$

7.1. **Extending maximum matchings in G_m.** Our goal in this subsection is to prove Lemma 7.2. That is, we wish to show that typically, *most* of the time G_m contains a matching which is as large as possible given the restrictions posed by isolates and quasi-isolates. Therefore it should not come as a surprise that we need to take a careful look at the structure of G_m.

The aim is to show that, for $m_2 \leq m \leq m_3$, if we have at least two helpers in G_m, then there are many choices of e_{m+1} whose addition would increase the size of a maximum matching (see Lemma 7.6 for the precise statement). Firstly, we need an expansion property of our graph, stronger than that given by Lemma 4.2 (ii). For the rest of the section, fix δ_0 to be the constant from Lemma 7.1 applied with $\epsilon = 0.01$ (so that $m_2 > (1 + \epsilon)n \log n/3$).

Lemma 7.5. *With high probability the following statement holds. There exists $\zeta > 0$ such that for all $m_2 \leq m \leq m_3$ and for every set $S \subseteq V$ of cardinality at most $\frac{n}{4 \log n}$ where every vertex in S has at least $\delta_0 \log n$ neighbours outside of S, we have $|N_{G_m}(S)| \geq \zeta \log n|S|$.*

Proof. By Theorem 3.4, with high probability every set S with $|S| \leq \frac{n}{4 \log n}$ satisfies

$$E_{G_m}(|S|, |N_{G_m}(S)|) \leq E_{G_{m_3}}(|S|, |N_{G_{m_3}}(S)|) \leq \frac{2m_3|S||N_{G_{m_3}}(S)|}{n^2} + \lambda \sqrt{\frac{m_3|S||N_{G_{m_3}}(S)|}{n}},$$

where λ is a positive constant granted by the Theorem. If for some S we have $|N_{G_{m_3}}(S)| < \zeta \log n |S|$ with a positive constant ζ, then the right-hand side is at most

$$\frac{4\zeta \log^2 n |S|^2}{n} + \lambda |S| \log n \sqrt{2\zeta} \le |S| \log n (\zeta + \lambda \sqrt{2\zeta}) < \frac{\delta_0}{2} |S| \log n,$$

when ζ is chosen to be sufficiently small. But by hypothesis, $E_{G_m}(|S|, |N_{G_m}(S)|) \ge \delta_0 |S| \log n$, a contradiction. $\qquad\square$

For the rest of the section we will assume that our process throughout steps $m_2 \le m \le m_3$ satisfies the properties granted by Lemma 7.1 and Lemma 7.5, without explicitly referring to the probabilistic statements.

Say that a vertex is *large* if it has degree at least $\delta_0 \log n$ in G_m and *small* otherwise. For a vertex $u \in V(M_m)$, let $g(u)$ denote the vertex that is matched to u in M_m. Similarly for a set $S \subseteq V(M_m)$, let $g(S) := \{g(u) := u \in S\}$. For $m \ge m_2$ we obtain the following properties as an easy consequence of Lemma 7.5 and the definition of M_m.

(P1) For any v, $N_{G_m}^5[v]$ contains at most two vertices of degree at most $\delta_0 \log n$ in G_m, where $N^i[v]$ denotes the set of vertices within distance at most i from v.

(P2) If a helper has degree one, then its neighbour w in C_m satisfies $\deg_{G_m}(g(w)) \ge 2$.

We will now introduce some more definitions. Let M be a matching in G_m. Say that a path $P := u_1 \ldots u_k \in G_m$ is *M-alternating* if the edges of P alternate between being in and out of M. Say that an M-alternating path P *augments* M if $u_1, u_k \notin V(M)$. Observe that if P augments M, then $M' := (M \setminus P) \cup (P \setminus M)$ is a larger matching in G_m than M. Our next lemma will show that if $|H_m| \ge 2$, then there are $\Omega(n^2)$ vertex pairs that have not yet been offered that will create a path that augments a maximum matching in G_m.

Lemma 7.6. *Let $m_2 \le m < m_3$. If $|H_m| \ge 2$, then for some absolute constant $\eta > 0$, the probability that $\nu(G_{m+1}) > \nu(G_m)$ is at least η.*

Proof. Let u_1, v_1 be distinct vertices in H_m. For ease of notation, let $C = C_m$, $B = B_m$ and $M := M_m$. We will find ηn^2 distinct vertex pairs (x, y) such that $x, y \in B$ (and hence $xy \notin E(G_m)$), and such that $G_m \cup \{xy\}$ contains an M-alternating path from u_1 to v_1 that augments M. As $m = o(n^2)$, at least $\frac{\eta n^2}{2}$ of these pairs are not contained in $e_1, \ldots, e_m$ and hence $\mathbb{P}[\nu(G_{m+1}) > \nu(G_m)] \ge \eta$.

In order to find the pairs (x, y), whose addition creates a path in G_m that augments M, we will first find disjoint paths P_1 from u_1 to u^* and Q_1 from v_1 to v^*, where u^* and v^* are large vertices in B_m. We will then be able to use the expansion properties of the graph induced by the large vertices to extend P_1 and Q_1 to many disjoint M-alternating paths terminating at distinct pairs of vertices in B. The main difficulty in the proof is finding the large family of extensions.

We begin by finding P_1 and Q_1.

Claim 7.7. *There exist disjoint M-alternating paths $P_1 = u_1 \ldots u^*$ and $Q_1 = v_1 \ldots v^*$, where u^* and v^* are large vertices and $|P_1|, |Q_1| \le 5$.*

Proof. If u_1 and v_1 share a neighbour then they do not both have degree one, otherwise u_1 or v_1 would be a quasi-isolate. So we are able to pick $u_2 \ne v_2 \in C$ such that $u_1 u_2$ and $v_1 v_2$ are edges of G_m. If $u_3 := g(u_2)$ and $v_3 := g(u_3)$ are both large, set $P_1 := u_1 u_2 u_3$ and $Q_1 := v_1 v_2 v_3$.

Therefore we may assume that u_3 is small. It follows from (M2) that u_1 is also small as $\deg_{G_m}(u_1) \leq \deg_{G_m}(u_3)$. By (P2), $\deg(u_3) \geq 2$ and so pick $u_4 \in C$ to be a neighbour of u_3 distinct from u_2. Define $u_5 := g(u_4)$. By Lemma 7.1, $N^2_{G_m}[u_2]$ contains at most two small vertices, so u_5 is large. Set $P_1 := u_1 \ldots u_5$.

We are left with two cases for v_3. If v_3 is small, then $N^3_{G_m}[u_2] \cap N^3_{G_m}[v_2] = \emptyset$ since otherwise we could find small vertices u_1, u_3, v_3 (and even v_1) at mutual distances at most 10. Therefore we can find v_5 analogously to u_5 and set $Q_1 := v_1 \ldots v_5$. If v_3 is large and $u_4 \neq v_2$, we may set $Q_1 = v_1 v_2 v_3$. Finally, let v_3 be large and $u_4 = v_2$. Then v_1 is large since otherwise $N^3_{G_m}[u_4]$ contains three small vertices. Hence v_1 has another neighbour $v_2' \notin \{u_2, v_2\}$, so set $v_3' := g(v_2')$. The vertex v_3' is also contained in $N^3_{G_m}[u_4]$ (recalling that $u_4 = v_2$) and therefore it is a large vertex (otherwise we again will have three small vertices in $N^3_{G_m}[u_4]$). Thus we may set $Q_1 = v_1 v_2' v_3'$. $\square$

The following claim gives a large family $\mathcal{P}$ of pairs of M-alternating paths extending P_1 and Q_1, such that each pair $(P, Q) \in \mathcal{P}$ yields a unique pair of endpoints (u_P, v_Q). Let $c := \frac{1}{4} \min\{\delta_0, \zeta\}$, where ζ is the constant from Lemma 7.5.

Claim 7.8. *There exists a family $\mathcal{P}$ of pairs of paths (P, Q) with the following properties.*

(i) *$P = P_1 P_2$ and $Q = Q_1 Q_2$ are M-alternating disjoint paths, where $P_2 = u^* \ldots u_P$ and $Q_2 = v^* \ldots v_Q$ with $u_P, u_Q \in B$.*

(ii) *$|\{(u_P, v_Q) : (P, Q) \in \mathcal{P}\}| \geq \frac{c^2 n^2}{16}$.*

Let us first explain why the lemma follows from this claim, before proving it. By (i), for each $(P, Q) \in \mathcal{P}$, the addition of the edge $u_P v_Q$ creates an M-alternating path that augments M. By (ii), we have at least $\frac{c^2 n^2}{16}$ such vertex pairs. At most $m = O(n \log n)$ of these pairs appear in $e_1, \ldots, e_m$, which implies the statement of the lemma with $\eta := \frac{c^2}{16}$.

Proof of Claim 7.8. We will construct disjoint sets $U_1, \ldots, U_k, V_1, \ldots, V_k \subseteq B \setminus (P_1 \cup Q_1)$. These sets will have the *path property* that, for every j and every pair $(x_j, y_j) \in U_j \times V_j$, there exist sequences $x_1, \ldots, x_{j-1}$ and $y_1, \ldots, y_{j-1}$, such that for all $i < j$ we have $x_i \in U_i$, $y_i \in V_i$ and $P_1 g(x_1) x_1 g(x_2) x_2 \ldots g(x_j) x_j$ and $Q g(y_1) y_1 g(y_2) y_2 \ldots g(y_j) y_j$ are M-alternating vertex-disjoint paths in G_m. See Figure 2 for an example of the paths we will create.

In particular, we will construct these sets such that, for each i, we have $|U_i| \geq c \log n |U_{i-1}|$ and $|V_i| \geq c \log n |V_{i-1}|$. We will show that, given U_i and V_i, we can create appropriate U_{i+1} and V_{i+1} until we reach some k where $|U_k|, |V_k| \geq \frac{c}{4} n$. Such sets U_k, V_k, with the property described in the above paragraph, will give the family of paths required for the claim.

These sets are constructed iteratively as follows. Set $U_0 := \{u^*\}$ and $V_0 := \{v^*\}$. Having defined U_i and V_i, we will find disjoint sets $U_{i+1} \subseteq B$ and $V_{i+1} \subseteq B \setminus (P_1 \cup Q_1)$ such that:

(i) U_{i+1} and V_{i+1} are disjoint from $\bigcup_{j=1}^{i} U_j \cup \bigcup_{j=1}^{i} V_j$;

(ii) $U_{i+1} \subseteq g(N(U_i))$ and $V_{i+1} \subseteq g(N(V_i))$.

(iii) Every vertex in $g(U_{i+1}) \cup g(V_{i+1})$ is large.

(iv) If $|U_i| \leq \frac{n}{4 \log n}$, then $|U_{i+1}| = |V_{i+1}| \geq c \log n |U_i|$,

We will keep constructing until we reach U_k and V_k such that $|U_k| = |V_k| \geq \frac{c}{4} n$. So the process runs for $O\left(\frac{\log n}{\log \log n}\right)$ steps. Observe that such a family of sets has the path property, and so constructing such a family will immediately give the claim.
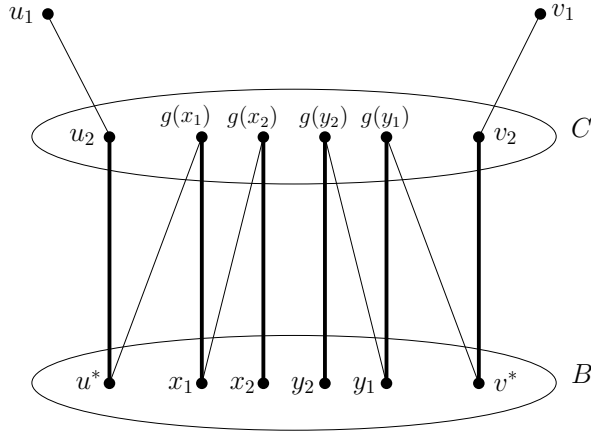
FIGURE 2. An example of the way the M-alternating paths will be constructed. The edges of M are bold.

Suppose that U_i and V_i have been defined for $0 \leq i < k$. If $i = 0$, set $X = N_{G_m}(u^*)$ and $Y = N_{G_m}(v^*)$. For $i > 0$, such that $|U_i| = |V_i| \leq \frac{n}{4 \log n}$, let $X := N_{G_m}(U_i)$ and $Y := N_{G_m}(V_i)$. For i such that $|U_i| = |V_i| > \frac{n}{4 \log n}$, pick $U' \subseteq U_i$ and $V' \subseteq V_i$ each of cardinality $\frac{n}{4 \log n}$ and set $X := N_{G_m}(U')$ and $Y := N_{G_m}(V')$.

As $U_i \cup V_i \subseteq B$ and B is an independent set, using (c) we may apply Lemma 7.5 to each of U_i and V_i. This gives that, for all i, we have $|X|, |Y| \geq 4c \log n |U_i|$. As $U_i, V_i \subseteq B$, we have $X \cup Y \subseteq C$. Therefore we can choose disjoint sets X' and Y', each of cardinality $2c \log n |U_i|$ such that $X' \subseteq X$ and $Y' \subseteq Y$. Observe that $|P|, |Q| \leq 5$ and

$$\sum_{j=1}^{i} |U_j| \leq |U_i| \left(1 + (c \log n)^{-1} + (c \log n)^{-2} + \dots \right) = O(|U_i|).$$

Also note that, by Lemma 7.1 and construction, $g(X')$ and $g(Y')$ contain $O(|U_i|)$ small vertices. So it is possible to pick $U_{i+1} \subseteq g(X')$ and $V_{i+1} \subseteq g(Y')$ satisfying (a)–(d), as required. This completes the proof of the claim. $\square$

As discussed above, the lemma follows directly from this claim. $\square$

Therefore, whenever $|H_m| \geq 2$, the probability that e_{m+1} extends a current maximum matching is bounded away from zero. We will use this fact to deduce that there cannot typically be a large number of steps where $|H_m| \geq 2$.

Proof of Lemma 7.2. Let η be the constant from the statement of Lemma 7.6 and $f(n) := \frac{n^{1-1/20}}{2}$. By our assumptions on the process, G_{m_2} contains a matching of size $\frac{n}{2} - f(n)$. By Lemma 7.6, in a step m with $|H_m| \geq 2$ the probability that the matching number increases is at least η. Note that from step m_2, the matching number cannot increase more than $f(n)$ times during the process.

Let $X \sim \text{Bin}(10 f(n)/\eta, \eta)$. Then $\mathbb{E}(X) = 10 f(n)$ and, by the Chernoff bound (Theorem 3.3) with $\epsilon = 9/10$, we have

$$\mathbb{P}\left[X < f(n) \right] \leq e^{-\frac{81}{100} \cdot 10 f(n)} = o(1).$$

Thus with high probability, after $10f(n)/\eta$ steps with $|H_m| \geq 2$ a perfect matching is reached. As once a perfect matching is reached, there are no helpers, the result follows.   □

7.2. **States with zero or one helpers.** We now show that typically, in a constant proportion of time steps after m_2, we have that $|J_m| > 0$ and $|J_m|$ is even, thus proving Lemma 7.3. The lemma and its proof formalise the intuition that the number of steps where $|J_m|$ is even should be comparable to the number of steps where it is odd.

We start the analysis from m_2 and, as before, we assume that G_{m_2} satisfies the properties of Lemma 4.2, (7.1) and Lemma 7.1. We first prove the following lower bound on the number of isolated vertices in G_{m_2}.

Lemma 7.9. *With high probability, G_{m_2} has at least $\frac{1}{2}n^{1/4}$ isolated vertices.*

Proof. Let $p = \frac{m_2}{\binom{n}{2}}$ and let Y_{m_2} be the number of isolated vertices in $G(n, p)$. One can check that $\mathbb{E}\left[Y_{m_2}\right] \sim ne^{-2m_2/n} \sim n^{1/4}$ and $\mathrm{Var}\left[Y_{m_2}\right] \sim \mathbb{E}\left[Y_{m_2}\right]$. So by Chebyshev's inequality, we have $\mathbb{P}\left[Y_{m_2} \leq \frac{1}{2}n^{1/4}\right] = o(1)$. Since this is a decreasing event, the same bound holds in $G_{m_2}^{\mathrm{all}}$ by Lemma 3.1. Any isolated vertex in $G_{m_2}^{\mathrm{all}}$ is also isolated in G_{m_2}, so the statement follows.   □

Proof of Lemma 7.3. We modify the process (G_m) to avoid a minor technical issue. The process (G'_m) is defined identically to (G_m) with one change: for $m > 2n \log n$, the offered pair e_m is drawn uniformly at random from the set $F := V^{(2)} \setminus \{e_1, e_2, \ldots, e_{2n \log n}\}$. The number of isolates and quasi-isolates in G'_m is denoted by J'_m.

The purpose of the modification is so that in each step, the probability of being offered a particular edge is at most $|F|^{-1}$. Note that, by definition, G'_m coincides with G_m in the first $2n \log n$ steps. Therefore, we may also assume that G'_m satisfies the properties given by Lemma 4.2, (7.1) and Lemma 7.1. It suffices to prove the lemma for the modified process G'_m as Theorem 1.2(i) implies that, with high probability, $G_{(2n \log n)}$ has a perfect matching and hence no quasi-isolates, and consequentially the same holds for $G'_{2n \log n}$. Hence we proceed with the proof for G'_m.

Let $\ell := \lfloor \frac{1}{4}n^{\frac{1}{4}} \rfloor$. For $k \in [2\ell]$, define

$$T_k := |\{m \in \mathbb{N} : |J'_m| = k\}| \text{ and } M(k) := \min\{m : |J'_m| \leq k\}.$$

As J_{m_2} contains the isolates in G_{m_2}, Lemma 7.9 implies that $M(2\ell) \geq m_2$ with high probability. We remark that $\mathbb{P}_{M(k+1)}[\cdot]$ in the following claim is a slight abuse of notation because $M(k+1)$ is a random variable, but its definition is clear: it is the probability of an event conditioned on the pairs $e_1, e_2, \ldots e_{M(k+1)}$.

For each $1 \leq k \leq 2\ell$, we will define a new random variable and show that it stochastically dominates T_k. Let X_k be mutually independent geometrically distributed random variables defined by

$$\mathbb{P}\left[X_k = r\right] = \left(1 - \frac{4.1k}{n}\right)^r \frac{4.1k}{n} \quad \text{for } r \in \mathbb{N} \cup \{0\}.$$

Please note that this form of the geometric distribution is shifted by one compared to the more usual formulation.

Claim 7.10. *For $\log^2 n \leq k \leq 2\ell$ and $r \geq 0$,*

$$\mathbb{P}_{M(k+1)}\left[T_k > r\right] \geq \mathbb{P}\left[X_k > r\right].$$

Proof. We start by proving the statement for $r = 0$, that is,

$$(7.2) \qquad \mathbb{P}_{M(k+1)}[T_k = 0] \leq \frac{4.1k}{n} = \mathbb{P}[X_k = 0].$$

If $T_k = 0$, then certainly $T_{k+1} > 0$ since we can lose at most two vertices of J'_m at any step m. Recall from above that $M(k+1)$ is the first time step such that $|J'_{M(k+1)}| = k + 1$. Let $U \subseteq V$ be the set of isolates, quasi-isolates and their partners in $G'_{M(k+1)}$, and note that $k + 1 \leq |U| \leq 2k + 2$. We will have $T_k = 0$ only if, from $M(k+1)$ onwards, an edge between two vertices of U (which will always be acceptable) is offered before an acceptable edge between a vertex of U and a vertex of degree at least two in $G'_{M(k+1)}$. We emphasise that this is only a necessary condition.

In each step $m \geq M(k+1)$, the probability that an edge between two vertices of U is offered is (using the definition of G'_m) at most

$$\frac{\binom{2k+2}{2}}{\binom{n}{2} - 2n \log n} = (1 + o(1))\frac{4k^2}{n^2}.$$

We have assumed (recall (7.1)) that $G'_{M(k+1)}$ contains a subgraph H which has a perfect matching and satisfies $|V(H)| \geq (1 - o(1))n$ and $\delta(H) = \Omega(\log n)$. This implies that $G'_{M(k+1)}$ has at least $(1 - o(1))\frac{n}{2}$ vertices v of degree at least 2 such that uv is acceptable for any $u \in U$. So the probability that an acceptable edge between a vertex of U and a vertex of degree at least 2 is offered is at least

$$\frac{(1 - o(1))\frac{n}{2} \cdot (k+1) - m}{\binom{n}{2} - m} \geq (1 - o(1))\frac{k}{n},$$

as $\log^2 n \leq k \leq 2\ell$ and thus $m = o(nk)$. Therefore, for $\log^2 n \leq k \leq 2\ell$, we have $\mathbb{P}_{M(k+1)}[T_k = 0] \leq (1 + o(1))\frac{4k}{n} \leq \frac{4.1k}{n}$, as required.

Secondly, we show that for $r > 0$

$$(7.3) \qquad \mathbb{P}_{M(k+1)}[T_k > r \mid T_k > 0] \geq \left(1 - \frac{4.1k}{n}\right)^r.$$

In other words, we are bounding the probability that the number of isolates and quasi-isolates does not decrease in r consecutive steps. Assuming the event $\{T_k > 0\}$, consider the event that for all $m \in \{M(k) + 1, M(k) + 2, \ldots, M(k) + r\}$, e_m is not incident to an isolate, a quasi-isolate or their partner in G_{m-1}. Note that this event implies $T_k > r$, so we will derive a lower bound on its probability.

For each $m \geq M(k) + 1$, the number of vertex pairs incident to an isolate, quasi-isolate or its partner in G_{m-1} is at most $2kn$. The probability that e_m is one of those pairs is at most

$$\frac{2kn}{\binom{n}{2} - 2n \log n} \leq \frac{4k}{n}\left(1 + \frac{8 \log n}{n}\right).$$

Using this bound for each of the r time steps, we get that

$$\mathbb{P}_{M(k+1)}[T_k > r \mid T_k > 0] \geq \left(1 - \frac{4.1k}{n}\right)^r.$$

Using (7.2), we conclude that

$$\mathbb{P}_{M(k+1)}\left[T_k > r\right] = \mathbb{P}_{M(k+1)}\left[T_k > r \mid T_k > 0\right]\mathbb{P}_{M(k+1)}\left[T_k > 0\right] \geq \left(1 - \frac{4.1k}{n}\right)^{r+1}.$$

On the other hand, a simple calculation yields $\mathbb{P}\left[X_k > r\right] = \left(1 - \frac{4.1k}{n}\right)^{r+1}$, so the statement follows. $\square$

Recall that $\ell = \lfloor \frac{1}{4}n^{\frac{1}{4}}\rfloor$ and let $T_{\text{even}} := \sum_{i=\log^2 n}^{\ell} T_{2i}$. To prove Lemma 7.3, we will show that $T_{\text{even}} = \Omega(n \log n)$ with high probability. It suffices to prove a lower bound on $X_{\text{even}} := \sum_{i=\log^2 n}^{\ell} X_{2i}$.

Claim 7.11. *For any $t \geq 0$,*

$$\mathbb{P}\left[T_{\text{even}} > t\right] \geq \mathbb{P}\left[X_{\text{even}} > t\right].$$

Proof. Since $T_{2\ell}, T_{2\ell-2}, \ldots, T_{2k+2}$ are determined by the first $M(2k+1)$ steps of the process, Claim 7.10 implies that for any $t \geq 0$,

$$\mathbb{P}\left[T_{2k} > t \mid T_{2\ell}, T_{2\ell-2}, \ldots, T_{2k+2}\right] \geq \mathbb{P}\left[X_{2k} > t\right].$$

This implies the required statement (or in other words, that T_{even} stochastically dominates X_{even}). For a proof, see, e.g., [17, Lemma 21.22]. We remark that in [17] they consider random variables taking values in $[0, 1]$, but this assumption is never used in their proof. $\square$

To conclude the proof of the lemma, we will show that $X_{\text{even}} \geq \frac{n \log n}{33}$ with high probability. The analysis is essentially that of the classical coupon collector problem, which is described for instance in [15].

By definition $\mathbb{E}\left[X_{2i}\right] = \frac{n}{8.2i} - 1$ and $\text{Var}\left[X_{2i}\right] = \left(1 - \frac{8.2i}{n}\right)\left(\frac{n}{8.2i}\right)^2$. Also recall that $\ell = \lfloor \frac{1}{4}n^{\frac{1}{4}}\rfloor$. By linearity of expectation we have

$$\mathbb{E}\left[X_{\text{even}}\right] = \sum_{i=\log^2 n}^{\ell}\left(\frac{n}{8.2i} - 1\right) = \frac{n}{8.2}\left(\frac{1}{5\log n} + \cdots + \frac{1}{\ell}\right) - O(\ell) \sim \frac{n}{8.2}\cdot \log \ell > \frac{n \log n}{32.9}.$$

We now calculate the variance of X_{even} in order to use Chebyshev's inequality to show that X_{even} does not deviate too far from its expectation. As the variables X_{2i} are independent, we have

$$\text{Var}\left[X_{\text{even}}\right] = \sum_{i=\log^2 n}^{\ell} \text{Var}\left[X_{2i}\right] = \sum_{i=\log^2 n}^{\ell} O\left(\frac{n^2}{i^2}\right) \leq O\left(n^2\right)\sum_{i=1}^{\infty}\frac{1}{i^2} = O\left(n^2\right).$$

So by Chebyshev's inequality, we have

$$\mathbb{P}\left[|X_{\text{even}} - \mathbb{E}(X_{\text{even}})| \geq \sqrt{n \log n}\right] = O\left(\frac{1}{\log n}\right) = o(1).$$

It follows that, with high probability, $X_{\text{even}} \geq \frac{n \log n}{33}$.

So, by Claim 7.11, with high probability $T_{\text{even}} \geq \frac{n \log n}{33}$. Assuming the statement of Lemma 7.9, all the time steps counted in T_{even} occur after m_2, as required. $\square$

We remark that we are not able to give an optimal constant for Lemma 7.3 because we do not establish full control of the transition probabilities (i.e., X_k is just a convenient lower bound for T_k), and we are ignoring any delay that arises before the time step m_2. Therefore, our computations are fairly crude.

7.3. **Decay of the number of isolated vertices.**

We complete the proof of Theorem 1.3 by tracking the number of isolated vertices (not quasi-isolates). Our goal is to show that, with high probability, at time $m = (\frac{1}{2} + \frac{1}{70})n \log n$, the graph G_m contains isolated vertices. Denote the number of isolated vertices in G_m by I_m. From the previous section, we know that G_m contains no helpers in a constant proportion of steps and we wish to show that this 'slows down' the decay of the number of isolates compared to G_m^{all}.

Before presenting the techinical details of the proof, let us explain heuristically what we will prove. As before, throughout the section we continue to condition on the events given by Lemmas 7.1, 7.9 and (7.1). Fix $\varphi = \varphi(n) = (\log n)^{-1/10}$. Say that a time step m is *unhelpful* if $|H_m| = 0$ and is not φ-flexible, that is, $|D_m| < \left(\frac{1}{2} + \varphi\right) n$ (recalling that D_m is the set of vertices contained in some optimal cover at time step m). The expected change in the number of isolates at a time step depends on whether the step is unhelpful or not. Informally, we have that 'on average'

$$(7.4) \qquad I_i - I_{i+1} \leq I_i Z_i,$$

where

$$(7.5) \qquad Z_i = \begin{cases} \frac{1+4\varphi}{n}, & \text{if } i \text{ is unhelpful,} \\ \frac{2+\varphi}{n}, & \text{otherwise.} \end{cases}$$

The formal statements and justification follow in the proof. Solving (7.4), we might expect that

$$(7.6) \qquad I_m \geq I_0 \exp\left(-\sum_{i=1}^{m} Z_i\right),$$

and we will use martingales and Freedman's Inequality (Theorem 3.8) to track the isolates and show that they behave essentially as in (7.6). The reader may find it helpful to compare this decay rate with $G_{n,m}$, where 'on average', the number of isolated vertices decreases by a factor of $\left(1 - \frac{2}{n}(1 + o(1))\right)$ per time step. Martingale concentration inequalities can be used to show that with high probability, the number of isolated vertices in the graph $G_{n,m}$ is asymptotically $ne^{-2m/n}$ when $m \leq \frac{n \log n}{2}$.

There are two technicalities we would like highlight. Firstly, define

$$(7.7) \qquad r := \min\{i : I_i \leq \lfloor \log n \rfloor\}.$$

We will work with the truncated variable $\hat{I}_m = \max\{I_m, I_r\}$ as the difference equations stop holding if I_m is too small. Secondly, we apply Freedman's inequality (as apposed to, e.g. Azuma's) in order to work with larger differences as long as they only occur rarely.

Let us now proceed with the details of the proof.

Proof of Theorem 1.3. As before, let $\varphi := (\log n)^{-1/10}$. Let

$$A_m := |\{m_2 \leq i \leq m_2 + m - 1 : \text{ step } i \text{ is unhelpful}\}|.$$

The following claim says that the number of isolated vertices in our process decays slower than in the corresponding G_m^{all}.

Claim 7.12. *Let $m_2 \leq m \leq 2n \log n$. With high probability,*

(7.8)
$$\log \left(\frac{\hat{I}_{m_2+m}}{\hat{I}_{m_2}} \right) \geq -\frac{2m}{n} + \frac{A_m}{n} - \frac{6\varphi m}{n}.$$

Proof. For $m_2 \leq i \leq m_2 + m$, define random variables (Y_i) by $Y_{i+1} := \log \left(\frac{\hat{I}_{i+1}}{\hat{I}_i} \right) + Z_i$, where Z_i is defined in (7.5). By definition,

$$\sum_{i=m_2}^{m_2+m-1} Z_i = \frac{1+4\varphi}{n} \cdot A_m + \frac{(2+\varphi)}{n} \cdot (m - A_m).$$

The sequence (Y_i) is set up in order to track the logarithm of $\hat{I}_i$, since we expect $\hat{I}_i$ to decrease exponentially. Moreover, the definition of Z_i has been chosen to ensure that $\mathbb{E}_i [Y_{i+1}] \geq 0$, i.e., $\sum_{i=m_2}^{m} Y_i$ is a submartingale. To prove this, we first estimate the probability that $\hat{I}_{i+1} = \hat{I}_i - 1$.

First consider the case where $i < r$, where r is defined in (7.7), and $Z_i = \frac{1+4\varphi}{n}$, so i is unhelpful. As $i < r$, $\hat{I}_i = I_i$. In this case, a pair uv with v isolated is acceptable for G_i only if $u \in D_i \cup J_i$, or if u is the partner of a quasi-isolate. Since $|D_i| < \left(\frac{1}{2} + \varphi \right) n$ and $|J_{m_2}| < n^{0.95}$ by (7.1), we have

$$\mathbb{P}_i [I_{i+1} = I_i - 1] \leq \frac{I_i(|D_i| + 2|J_i|)}{\binom{n}{2} - i} \leq \frac{(1+3\varphi)I_i}{n}.$$

Now we compute $\mathbb{E}_i [Y_{i+1}]$. We have

$$\mathbb{E}_i [Y_{i+1}] = \mathbb{P}_i [I_{i+1} = I_i - 2] \log \left(1 - \frac{2}{I_i} \right) + \mathbb{P}_i [I_{i+1} = I_i - 1] \log \left(1 - \frac{1}{I_i} \right) + Z_i$$

$$\geq \frac{I_i^2}{2} \cdot \frac{2}{n(n - 5 \log n)} \cdot \log \left(1 - \frac{2}{I_i} \right) + \frac{(1+3\varphi)I_i}{n} \log \left(1 - \frac{1}{I_i} \right) + Z_i.$$

The first summand, corresponding to $\{I_{i+1} = I_i - 2\}$, is of a lower order. This event happens when we offered one of $\binom{|I_i|}{2}$ acceptable pairs with both vertices in I_i. Its denominator is $n - 5 \log n$ to account for $i \leq 2n \log n$ edges which have been offered so far. We also use the inequality $\log(1 - x) \geq -x - x^2$ for $x < 0.1$ to obtain the bound

$$\mathbb{E}_i [Y_{i+1}] \geq O \left(\frac{I_i}{n^2} \right) + \frac{(1+3\varphi)I_i}{n} \left(-\frac{1}{I_i} - \frac{1}{I_i^2} \right) + \frac{1+4\varphi}{n} \geq 0.$$

In the case $i < r$ and $Z_i = \frac{2+\varphi}{n}$, we do a similar computation to see that $\mathbb{E}_i [Y_{i+1}] \geq 0$. The only change is in the second summand, since now any pair containing an isolated vertex v can be acceptable and therefore the probability that v is not isolated in G_{i+1} is at most $\frac{2}{n - 5 \log n}$. This gives

$$\mathbb{E}_i [Y_{i+1}] \geq \frac{I_i^2}{2} \cdot \frac{2}{n(n - 5 \log n)} \log \left(1 - \frac{2}{I_i} \right) + \frac{2I_i}{n - 5 \log n} \log \left(1 - \frac{1}{I_i} \right) + Z_i$$

$$\geq O \left(\frac{I_i}{n^2} \right) + \frac{2I_i}{n - 5 \log n} \left(-\frac{1}{I_i} - \frac{1}{I_i^2} \right) + \frac{2+\varphi}{n} \geq 0.$$

In the case $i \geq r$, we have $\hat{I}_{i+1} = \hat{I}_i = \hat{I}_r$ by (7.7), so $\mathbb{E}_i[Y_{i+1}] = 0 + Z_i > 0$.

To control $\sum_{i=m_2}^{m} Y_i$ we will use Freedman's Inequality (Theorem 3.8), so let us bound the predictable quadratic variation of the process. Since $\log\left(\frac{\hat{I}_{i+1}}{\hat{I}_i}\right) \leq 0$,

$$\mathbb{E}_i\left[Y_{i+1}^2\right] \leq \mathbb{E}_i\left[\log^2\left(\frac{\hat{I}_{i+1}}{\hat{I}_i}\right)\right] + \mathbb{E}_i\left[Z_i^2\right]$$

$$\leq \mathbb{P}_i\left[\hat{I}_{i+1} = \hat{I}_i - 2\right]\log^2\left(1 - \frac{2}{\hat{I}_i}\right) + \mathbb{P}_i\left[\hat{I}_{i+1} = \hat{I}_i - 1\right]\log^2\left(1 - \frac{1}{\hat{I}_i}\right) + Z_i^2.$$

The first two summands in the following inequality correspond to the event where $\hat{I}_{i+1} < \hat{I}_i$, which only occurs for $i > r$. The third term corresponds to the event where $\hat{I}_{i+1} = \hat{I}_i$. We use the notation $\log^2 x = (\log x)^2$. Once again, we use the inequality $-\log(1 - x) \leq 2x$ (for $x < 0.1$) to conclude that

$$\mathbb{E}_i\left[Y_{i+1}^2\right] \leq \frac{2\hat{I}_i^2}{n^2} \cdot \frac{16}{\hat{I}_i^2} + \frac{4\hat{I}_i}{n} \cdot \frac{4}{\hat{I}_i^2} + \frac{5}{n^2} \leq \frac{20}{n\hat{I}_i}.$$

Summing over i and using $\frac{1}{\hat{I}_i} \leq \frac{1}{\log n}$, we get

$$W_m = \sum_{i=m_2}^{m_2+m} \mathbb{E}_i\left[Y_{i+1}^2\right] \leq \frac{20m}{n\log n} \leq 40.$$

Applying Theorem 3.8 with $\ell = \sqrt{\log n}$, we get that with high probability $\sum_{i=m_2+1}^{m_2+m} Y_i \geq -\sqrt{\log n}$. Therefore

$$\log\left(\frac{\hat{I}_{m_2+m}}{\hat{I}_{m_2}}\right) = \sum_{i=m_2+1}^{m_2+m}(Y_i - Z_{i-1}) \geq -\sqrt{\log n} - \sum_{i=m_2}^{m_2+m-1} Z_i$$

$$= -\sqrt{\log n} - \frac{2+\varphi}{n}\cdot(m - A_m) - \frac{1+4\varphi}{n}\cdot A_m \geq -\frac{2m}{n} + \frac{A_m}{n} - \frac{6\varphi m}{n}.$$

$\square$

Let $C := \frac{1}{34}$ and $m := \left(\frac{1}{8} + \frac{34C}{70}\right)n\log n$. Assume that $|H_i| = 0$ and $|J_i| > 0$ for at least $Cn\log n$ steps $i \geq m_2$. By Lemma 7.4, this holds with high probability. If one of those steps satisfies $i \geq m + m_2$, then $|J_{m_2+m}| > 0$. Lemma 7.1 with $a = 2$ implies that with high probability there are no quasi-isolates at time $m_2 + m$, so $I_{m_2+m} > 0$ as required. Therefore we may assume that all the steps i with $|J_i| > 0$ occur before $m_2 + m$. This, along with the fact that (by Lemma 2.2) the number of φ-flexible steps is typically $o(n\log n)$, implies that $A_m \geq Cn\log n(1 - o(1))$. By substituting in our value of m, we have

$$\frac{2m}{n} - \frac{A_m}{n} + \frac{6\varphi m}{n} < \left(\frac{1}{4} + \frac{34C}{35} - C + o(1)\right)\log n = \left(\frac{1}{4} - \frac{C}{35} + o(1)\right)\log n.$$

This along with Claim 7.12 implies that with high probability,

$$\log\left(\frac{\hat{I}_{m_2+m}}{\hat{I}_{m_2}}\right) \geq -\frac{2m}{n} + \frac{A_m}{n} - \frac{6\varphi m}{n} > -\left(\frac{1}{4} - \frac{C}{40}\right)\log n.$$

Since $\hat{I}_{m_2} = I_{m_2} > \frac{1}{2}n^{1/4}$ by Lemma 7.9, we get that $\hat{I}_{m_2+m} \geq \frac{1}{4}n^{\frac{1}{4}-\frac{1}{4}+\frac{C}{40}} > n^{C/50} > \lfloor \log n \rfloor$, and therefore $I_{m_2+m} > 0$. It remains to verify that the constant is as stated in Theorem 1.3, that is, $m_2 + m = \left(\frac{3}{8} + \frac{1}{8} + \frac{34C}{70}\right) n \log n = \left(\frac{1}{2} + \frac{1}{70}\right) n \log n$.

$\square$

8. Concluding Remarks and Open Problems

In this paper we studied the random graph process which at every step maintains the property that the matching and cover number of the current graph coincide. We have made progress in understanding the properties of the graph obtained at various stages of this process. Yet several interesting problems remain open. Theorems 1.2 and 1.3 give us bounds on the typical appearance of a perfect matching in G_m. We have a heuristic argument which suggests the threshold is $(1 + o(1))\frac{3}{4}n \log n$ and we wonder if this can be made rigorous.

Question 8.1. *What is the threshold for appearance of a perfect matching in G_m?*

Lemma 2.2 says that with high probability, most of the time when $m \geq cn \log n$ for a small constant $c > 0$, the union of optimal covers D_m has $\frac{n}{2}(1 + o(1))$ vertices. We wonder if this holds *all the time*.

Question 8.2. *Is it true that, with high probability, the union of optimal covers D_m have $\frac{n}{2}(1 + o(1))$ vertices for all $m \geq cn \log n$ for a small positive constant c?*

In light of Theorem 1.2, it would also be interesting to determine precisely when G_m has a unique optimal vertex cover and to determine the correct order of the number of vertex pairs that are not present in G_N (in particular, whether it is sublinear in n). In order to get precise bounds for any of the properties discussed above, we believe it is necessary to analyse more carefully the earlier stages of the process (before $m = \frac{3}{8}n \log n$). We have made essentially no attempt to do this.

Another direction of research could be to study a corresponding random Kőnig hypergraph process. Here we fix $r \geq 3$ and n such that $r|n$. Let $N := \binom{n}{r}$. $e_1, \ldots, e_N$ be a uniformly random ordering of the edges of the complete r-uniform hypergraph on n vertices. Let G_0 be an independent set of n vertices. For $i = 0, \ldots, N - 1$ we define $G_{i+1} := G_i + e_i$ if $\tau(G_i + e_i) = \nu(G_i + e_i)$, and $G_{i+1} := G_i$ otherwise. As in the graph process, we expect to reach a hypergraph G, where $\tau(G) = \nu(G) = \frac{n}{r}$. We also expect $V(G) = T \cup I$, where $|T| = \frac{n}{r}$, nearly all edges touching T are present and I is independent. It seems that the hypergraph process will require a different set of tools, as there is no concept analogous to alternating paths.

Finally, it is also interesting to study random graph processes preserving other global properties, for example the property of being a perfect graph.

References

[1] T. Bohman. The triangle-free process. *Adv. Math.*, 221(5):1653–1677, 2009.

[2] T. Bohman and A. Frieze. Avoiding a giant component. *Random Structures Algorithms*, 19(1):75–85, 2001.

[3] T. Bohman and P. Keevash. The early evolution of the H-free process. *Invent. Math.*, 181(2):291–336, 2010.

[4] T. Bohman and P. Keevash. Dynamic concentration of the triangle-free process. In *The Seventh European Conference on Combinatorics, Graph Theory and Applications*, volume 16 of *CRM Series*, pages 489–495. Ed. Norm., Pisa, 2013.

[5] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[6] B. Bollobás and A. M. Frieze. On matchings and Hamiltonian cycles in random graphs. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 23–46. North-Holland, Amsterdam, 1985.

[7] B. Bollobás and A. Thomason. Random graphs of small order. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 47–97. North-Holland, Amsterdam, 1985.

[8] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, Cambridge, 2009.

[9] J. Egerváry. Matrixok kombinatorius tulajdonságairól. *Matematikai és Fizikai Lapok*, 38(1931):16–28, 1931.

[10] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[12] P. Erdős and A. Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar.*, 17:359–368, 1966.

[13] P. Erdős, S. Suen, and P. Winkler. On the size of a random maximal graph. In *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993)*, volume 6, pages 309–318, 1995.

[14] X. Fan, I. Grama, and Q. Liu. Hoeffding's inequality for supermartingales. *Stochastic Process. Appl.*, 122(10):3545–3559, 2012.

[15] M. Ferrante and M. Saltalamacchia. Lecture Notes: The Coupon Collectors Problem, 2014. URL: `http://mat.uab.cat/matmat/PDFv2014/v2014n02.pdf`.

[16] G. Fiz Pontiveros, S. Griffiths, and R. Morris. The triangle-free process and the Ramsey number $R(3, k)$. *Mem. Amer. Math. Soc.*, 263(1274):v+125, 2020.

[17] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016.

[18] A. M. Frieze. On large matchings and cycles in sparse random graphs. *Discrete Math.*, 59(3):243–256, 1986.

[19] S. Gerke, D. Schlatter, A. Steger, and A. Taraz. The random planar graph process. *Random Structures Algorithms*, 32(2):236–261, 2008.

[20] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*, volume 45. John Wiley & Sons, 2011.

[21] M. Kang and K. Panagiotou. On the connectivity threshold of Achlioptas processes. *J. Comb.*, 5(3):291–304, 2014.

[22] D. Kőnig. Gráfok és mátrixok. *Matematikai és Fizikai Lapok*, 38(1931):116–119, 1931.

[23] M. Krivelevich, M. Kwan, P.-S. Loh, and B. Sudakov. The random k-matching-free process. *Random Structures Algorithms*, 53(4):692–716, 2018.

[24] M. Krivelevich, E. Lubetzky, and B. Sudakov. Hamiltonicity thresholds in Achlioptas processes. *Random Structures Algorithms*, 37(1):1–24, 2010.

[25] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Soc. Math. Stud.*, pages 199–262. Springer, Berlin, 2006.

[26] M. Krivelevich, B. Sudakov, and D. Vilenchik. On the random satisfiable process. *Combin. Probab. Comput.*, 18(5):775–801, 2009.

[27] A. Ruciński and N. C. Wormald. Random graph processes with degree restrictions. *Combin. Probab. Comput.*, 1(2):169–180, 1992.

[28] J. A. Tropp. Freedman's inequality for matrix martingales. *Electron. Commun. Probab.*, 16:262–270, 2011.

[29] G. Wolfovitz. Lower bounds for the size of random maximal H-free graphs. *Electron. J. Combin.*, 16(1):Research Paper 4, 26, 2009.

School of Mathematics, Monash University, VIC 3800, Australia
E-mail address: `nina.kamcev@monash.edu`

Department of Mathematics, ETH, Zurich, Switzerland
E-mail address: `benjamin.sudakov@math.ethz.ch`

School of Mathematical Sciences, Tel Aviv, Israel
E-mail address: `krivelev@tauex.tau.ac.il`

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, UK
E-mail address: `morrison@dpmms.cam.ac.uk`