WHAT IS THE PROBABILITY THAT A RANDOM GRAPH WITH A GIVEN DEGREE SEQUENCE IS CONNECTED?

LOUIGI ADDARIO-BERRY

Department of Mathematics and Statistics, McGill University, Montréal, Canada

BRUCE REED

Institute of Mathematics, Academia Sinica, Taiwan. Supported by NSTC Grant 112-2115-M-001 -013 -MY3

DAO CHEN YUAN

Department of Mathematics and Statistics, McGill University, Montréal, Canada

Abstract. An n-tuple $\mathcal{D} = (d(1), \ldots, d(n))$ is a *feasible degree sequence* if there is a graph on $\{1, \ldots, n\}$ such that i has degree $d(i)$. Any such graph will have $m = \sum_{i=1}^{n} d(i)/2$ edges. Letting $G(\mathcal{D})$ be a graph chosen uniformly from those with the given degree sequence, we upper-bound the probability that $G(\mathcal{D})$ is disconnected based on the number of vertices of degree d for small d, and develop a powerful tool for proving such bounds. If there are any vertices of degree zero the probability G is disconnected is 1, so we assume there are no such vertices. Our results then imply that if there are $o(\sqrt{m})$ vertices of degree 1 and $o(m)$ vertices of degree 2 then with high probability G is connected, while if there are no vertices of degree 1 or 2 then the probability G is disconnected is $O(\frac{n^4}{m^6})$.

1. Introduction

An n-tuple $\mathcal{D} = (d(1), \ldots, d(n))$ is a *feasible degree sequence* if there is a graph on $\{1, \ldots, n\}$ such that i has degree $d(i)$. We are interested in the probability that the random graph $G = G(\mathcal{D})$ on $[n] = \{1, \ldots, n\}$ chosen uniformly from those with the given degree sequence is connected.

The answer to this question does not depend on the order of the degrees in the sequence as any ordering gives the same probability of connectivity. Our analysis and discussion will assume that the degrees appear in non-decreasing order, although our theorems are stated without this condition.

For G to be connected there can be no vertex of degree 0, and we restrict our attention to such degree sequences from now on. If every vertex has degree 1 then G is not connected; indeed, in this case every component of G has 2 vertices. If every vertex of G has degree 2, then G is the disjoint union of cycles. It is well known since at least 1981 [13] that for such degree sequences, the probability that G is a single cycle is $o(1)$. On the other

E-mail addresses: louigi.addario@mcgill.ca, bruce.al.reed@gmail.com, dao.yuan@mcgill.ca.
Date: April 28, 2026.
2020 *Mathematics Subject Classification.* 05C80,60C05.

hand, as shown by Wormald in the same year [12], for every fixed $r \geq 3$, if every vertex has degree r then the probability that $G(\mathcal{D}_n)$ is connected goes to 1 as n goes to infinity.

The answer to the question that we tackle in this paper *does* depend on the number of vertices of low degree. For every $1 \leq i \leq n-1$, we use $n_i = n_i(\mathcal{D})$ to denote the number of vertices of degree i in $\mathcal{D}$. We note that every graph with degree sequence $\mathcal{D}$ has $m = m_{\mathcal{D}} = \frac{1}{2}\sum_{i=1}^{n} d(i)$ edges. If $n_1 > m$ then there must be a component which is an edge, so for such degree sequences the probability G is connected is 0. On the other hand, we prove that if $n_1 = o(\sqrt{m})$ and $n_2 = o(m)$ then with high probability $G(\mathcal{D})$ is connected.[1] More strongly, we have the following theorem, which is stated in terms of the invariants

$$u_{edge} = \frac{\max(n_1-1,0)^2}{m},\ u_{\Delta} = \frac{\max(n_2-2,0)^3}{m^3},\ u_{\Delta+1} = \frac{n_1\max(n_2-1,0)^2 n_3}{m^4},$$
$$u_{K_4-e} = \frac{\max(n_2-1,0)^2\max(n_3-1,0)^2}{m^5},\ u_{K_4} = \frac{\max(n_3-3,0)^4}{m^6},\ u_{K_5^+} = \frac{n}{m^6}.$$

These invariants bound, in order, the probability of components which are: an edge, a triangle, a triangle attached to a degree-1 vertex, a clique of size 4 with an edge removed, a clique of size 4, and a clique of size 5 and other small graphs.

Theorem 1.1. *For any feasible degree sequence $\mathcal{D} = (d(1), \dots, d(n))$, the probability that $G(\mathcal{D})$ is disconnected is*

$$O\left(u_{edge} + u_{\Delta} + u_{\Delta+1} + u_{K_4-e} + u_{K_4} + u_{K_5^+}\right).$$

Corollary 1.2. *If $n_1 = o(\sqrt{m})$ and $n_2 = o(m)$ then whp G is connected. If $n_1 < 2$ and $n_2 < 2$ then the probability G is disconnected is $O(\frac{n_3^4+n}{m^6})$.*

We note that the upper bounds on n_1 and n_2 ensuring G is connected whp, are not tight. Indeed m is the required lower bound on the number of vertices of degree 1 to ensure a random graph with a given degree sequence is disconnected with nonzero probability. For example, if $\mathcal{D} = (d(1), \dots, d(n))$ has $d(n) = n-1$ and $d(i) = 1$ for $i < n$, then $G(\mathcal{D})$ is always a star and connected. In the same vein, if $\mathcal{D}$ has $d(n) = d(n-1) = n-1$ and $d(i) = 2$ for $i < n-1$, then $G(\mathcal{D})$ is also always connected. However, as we discuss more fully in Section 1.3, for a large class of degree sequences it can be shown that if $n_1 = \Omega(\sqrt{m})$ or $n_2 = \Omega(m)$ then whp $G(\mathcal{D})$ is disconnected. For example, this is true if $d(n) = o(\sqrt{m})$.

A key tool in proving Theorem 1.1, is the following result which is of independent interest.

Theorem 1.3. *For any $\gamma > 0$, for any degree sequence satisfying $n_1 \leq m^{1-\gamma}, n_2 \leq 10^{-6}m$, the probability that there is more than one component with more than $4(\log m)^4$ edges is $o(m^{-6\log\log m})$.*

This theorem allows us to obtain tight bounds on the probability that $G(\mathcal{D})$ is connected simply by determining the probability that there is a component of size at most $4(\log m)^4$.

Theorem 1.1 is just the tip of the iceberg with respect to possible applications of Theorem 1.3 which will allow the computation of precise bounds on the probability that $G(\mathcal{D})$ has up to any fixed number c of components for, e.g., a degree sequence where every vertex has degree at most r.

[1]We use *with high probability*, or whp, to mean with probability tending to one as m tends to infinity (or equivalently as $n \to \infty$, since we considering simple graphs).

1.1. **Generating $G(\mathcal{D})$ and the Configuration Model.** We consider a set consisting of $d(i)$ labelled half-edges corresponding to each vertex i. Any matching on the half-edges corresponds to a unique, half-edge labelled multigraph (possibly with loops and parallel edges). Each simple graph with degree sequence $\mathcal{D}$ corresponds to $\prod_{i=1}^{n} d(i)!$ matchings. So, if we generate a random matching $\mathcal{R} = \mathcal{R}_{\mathcal{D}}$ conditional on it yielding a multigraph without loops and parallel edges, then the corresponding random graph is $G(\mathcal{D})$. Our analysis involves constructing such a matching and hence the graph $G(\mathcal{D})$.

In the configuration model on multigraphs, we simply generate an unconditioned random matching and take the corresponding multigraph. If we generate the matching one edge at a time then at each step, given the current partial matching, every unmatched half-edge is equally likely to be matched to every other unmatched half-edge, which makes the analysis of this model relatively easy. For example, the expected number of edges components in the configuration model is exactly $\frac{\binom{n_1}{2}}{m-1}$ and a straightforward analysis shows that if $n_1 = \omega(\sqrt{m})$ then with high probability there is an edge component and hence $G(\mathcal{D})$ is disconnected.

In proving our results we exploit the fact that if the degree sequence is "well-behaved", and the partial matching we have constructed is small, then even conditioned on generating a multigraph without loops and parallel edges, every unmatched half-edge is *almost* equally likely to be matched to every other unmatched half-edge which does not create a loop or a parallel edge.

This is not the case for all degree sequences. For example, if the degree sequence is $\{1, 1, .., 1, n-1\}$ and has n terms, then the probability that two vertices of degree 1 are joined is zero. The complication for this example comes from the existence of a vertex of very high degree. Similar problems may arise if the sum of the degrees of the neighbours of some vertex x, denoted $D(x)$, is large. We use $D^* = D^*(\mathcal{D})$ to denote $\sum_{i=n-d(n)+1}^{n} d(i)$ which is an upper bound on $D(x)$ for all x in every graph with degree sequence $\mathcal{D}$. We note that if $\Delta = o(\sqrt{m})$, then $D^* = o(m)$.

In order to handle the degree sequences in our analysis, we use a technique known as *switching*, which we now introduce.

1.2. **Switching.** Switching was introduced by Senior in 1951 [11], under the name *transfusions*, to study representative graphs of partitions (see also [2]). Havel showed in 1955 [6] that all graphs on the same degree sequence can be obtained from one another via switchings. It was first used by McKay in 1981 [9] to study random graphs of fixed degrees and was the key to Wormald's proof in the same year [12] that almost every k-regular graph is connected. We shall use it to both

(a) prove that, if the degree sequence is "well-behaved" and the partial matching we have constructed is small, then every unmatched half-edge is almost equally likely to be matched to every other unmatched half-edge which does not create a loop or a parallel edge, and
(b) handle badly behaved degree sequences.

A graph J is obtained from a graph H by a *switching* on a pair of oriented edges xy and uv of H if $V(J) = V(H)$ and $E(J) = E(H) - \{xy, uv\} \cup \{xv, uy\}$ (see Figure 1). We note that a switching is only possible if $x \neq v, u \neq y$ and $xv, uy \notin E(G)$. Furthermore, if J is obtained from H by switching on xy and uv then it is also obtained from H by switching on yx and vu. Thus there are either no or exactly two switchings from H to J. Finally, if J is obtained from H by switching on xy and uv then H is obtained from J by switching on xv and uy. So there are two switchings from H to J if and only if there are two switchings from J to H. We also note that if we have a matching on half-edges associated with H and J, the switching is also a switching on two oriented edges of that matching.

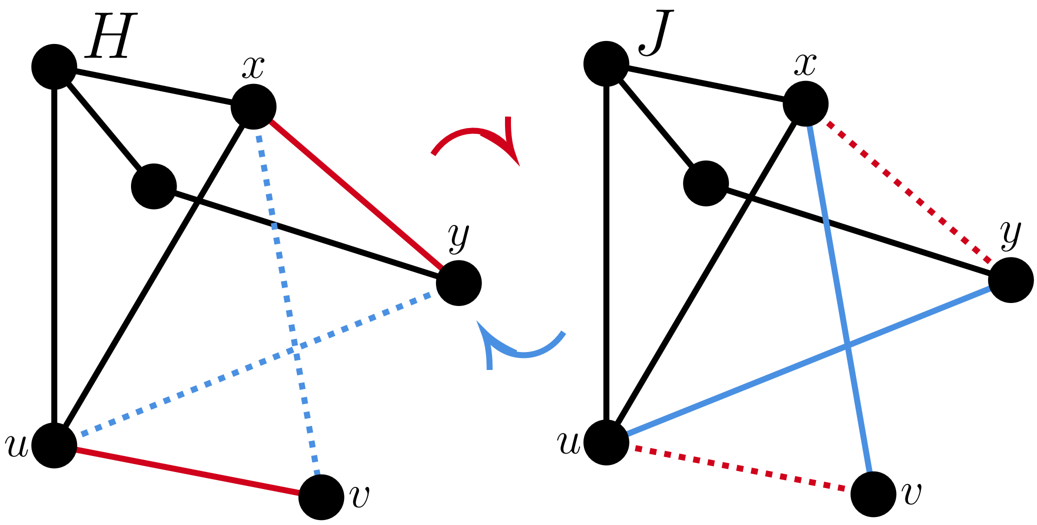

FIGURE 1. Example of a switching at $\{xy, uv\}$ to get edges $\{xv, uy\}$ from graphs H to J. In this and all subsequent figures of switchings, edges used in the switching are red and edges used to undo the switching are blue. Dashed lines represent the absence of edges, unless otherwise noted in the caption.

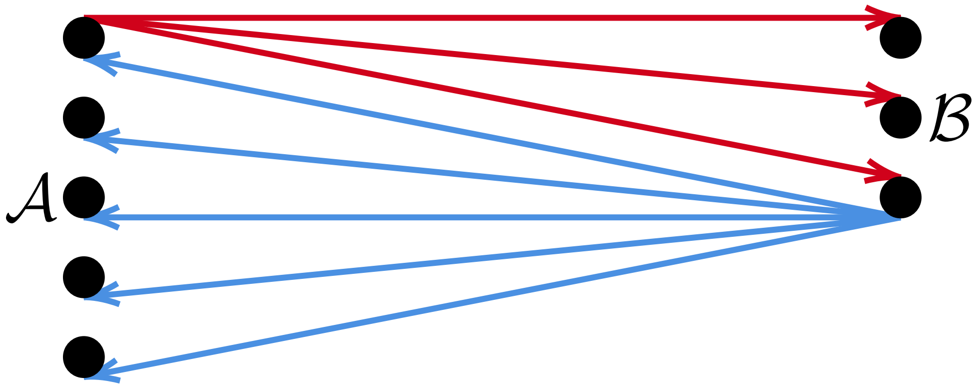

FIGURE 2. If every graph in $\mathcal{A}$ has at most $\Delta(\mathcal{A})$ switchings (in red) into $\mathcal{B}$, while every graph in $\mathcal{B}$ has at least $\delta(\mathcal{B})$ switchings (in blue) into $\mathcal{A}$, then $|\mathcal{B}| \le \frac{\Delta(\mathcal{A})}{\delta(\mathcal{B})}|\mathcal{A}|$.

Given two disjoint families $\mathcal{A}$ and $\mathcal{B}$ of graphs with degree sequence $\mathcal{D}$, we can compare their relative sizes by considering switchings between them. Specifically if for every graph H in $\mathcal{A}$ there are at most $\Delta(\mathcal{A})$ graphs in $\mathcal{B}$ which can be obtained from H by a switching and for every graph J in $\mathcal{B}$ there are at least $\delta(\mathcal{B})$ graphs in $\mathcal{A}$ which can be obtained from H by a switching then, as illustrated in Figure 2,

$$|\mathcal{B}| \le \frac{\Delta(\mathcal{A})}{\delta(\mathcal{B})}|\mathcal{A}|.$$

It is this observation which allows us to exploit an analysis of switching.

To illustrate its power, especially when degrees are bounded, we prove the following lemma.

Lemma 1.4. *Suppose that N is a matching on the half-edges of size at most $\frac{m}{8}$, which is extendable to a full matching giving a simple graph, and that h_v and h_w are half-edges not matched in N corresponding to distinct vertices v and w. Suppose either that w has* *degree at most $\frac{\sqrt{m}}{10}$ or that $D^* \leq \frac{m}{100}$. Then the conditional probability, given $N \subseteq \mathcal{R}$, that $h_v h_w$ is an edge of R and that $D(v) < \frac{m}{100}$, is at most $\frac{4}{3m}$.*

Proof. Let $\mathcal{F}$ denote the event that N is a submatching of $\mathcal{R}$, $D(v) < \frac{m}{100}$, and $h_v h_w \in R_{\mathcal{D}}$, Let $\mathcal{F}'$ denote the event that N is a submatching of $\mathcal{R}$, $h_v h_w \notin R_{\mathcal{D}}$. We count switchings between $\mathcal{F}$ and $\mathcal{F}'$. There are at most two switchings from a graph in $\mathcal{F}'$ to a graph in $\mathcal{F}$, using $\{h_v h, h_w h'\}$ (where h, h' are the half-edges matched to h_v, h_w respectively). To switch from a graph in $\mathcal{F}$ to one in $\mathcal{F}'$, it suffices to switch $h_v h_w$ with any edge not in N, such that the result is a simple graph. There are at most $\frac{m}{8}$ edges in N, at most $\frac{m}{100}$ edges incident to a neighbour of v, and at most $\min(\binom{d(w)}{2}, D^*) < \frac{m}{100}$ edges both of whose endpoints are neighbours of w. This leaves more than $\frac{3m}{4}$ edges which can be switched with $h_v h_w$ and hence at least $\frac{3m}{2}$ switchings. So $\mathbf{P}\{\mathcal{F}\} \leq \frac{4}{3m}$. □

1.3. **Previous Results.** The connectivity of $G(\mathcal{D})$ was first studied by Wormald in 1981 [12], who proved that if $\mathcal{D}$ has $d(1) = k \geq 3$ and $d(n) \leq C$ constant, then $G(\mathcal{D})$ is connected whp. Luczak[8] showed in 1992 that for any $\mathcal{D}$ where $d(1) \geq 3$ and $d(n) \leq n^{0.01}$, $G(\mathcal{D})$ is connected whp, and provided a characterization for connectivity whp when $d(1) = 2$ and $d(n) \leq n^{0.01}$. For d-regular degree sequences, Cooper, Frieze and Reed [3] showed in 2002 that they are connected whp for any $3 \leq d \leq \epsilon n$ for a small constant $\epsilon > 0$.

Federico and van der Hofstad [4] proved in 2017 that for sequences of degree sequences $\{\mathcal{D}_n\}$ such that for some constant C and every n, $\sum_{i=1}^{n} d(i)^2 < Cn$, we have the following: $\mathbf{P}\{G(\mathcal{D}_n) \text{ is connected}\} = 1 - o(1)$ precisely if the number of vertices of degree 1 is $o(\sqrt{m})$ and the number of vertices of degree 2 is $o(m)$. This latter result is a consequence of a more general result on the configuration model, which translates to the uniform simple graph model when the mean of the square of the degree of a uniformly random vertex is bounded.

Joos, Perarnau, Rautenbach, and Reed [7], building on earlier work of Molloy and Reed [10] from 1995, stated in 2018 precise conditions on a sequence of degree sequences which ensures that a graph chosen uniformly from those with the given degree sequence has a giant component[2].

Gao and Ohapkin [5] proved in 2023 that if either:

(a) $D^* = o(m)$, or

(b) the total degree of vertices of degree at least $\frac{\sqrt{2m}}{\log 2m}$ is no more than $(1 - \Omega(1))m$,

then if $n_1 = o(\sqrt{m})$ and $n_2 = o(m)$ then $G(\mathcal{D})$ is connected whp. They also proved a partial converse, namely that under assumption (a), for all $c > 0$ there is an $\epsilon > 0$, such that if $n_1 > c\sqrt{m}$ or $n_2 > cm$, then $G(\mathcal{D})$ is disconnected with probability greater than ϵ for all large m.

Our results strengthen the results of Gao and Ohapkin in that we make no assumptions and that we give upper bounds on the probability that $G(\mathcal{D})$ is disconnected rather than simply proving it is $o(1)$. Furthermore, under the following significant weakening of their assumption (a), namely under the assumption (a') $D^* \leq \frac{m}{3}$, we can show (see 7) that the bounds in Theorem 1.1 are tight up to a constant factor provided $n_1 > 1$, $n_2 > 2$ or $n_3 = \Omega(n^{1/4})$. We are able to prove our bounds are tight with a much weaker assumption because of Theorem 1.3, which allows us to restrict our analysis to determining the probability of the existence of very small components.

2. A Proof Outline

We explore the component containing each vertex v by iteratively growing a tree T rooted at v until we have constructed a spanning tree for the component. Along with T

[2]This means that there is some $\epsilon > 0$ such that the probability G has a component with at least ϵn vertices is $1 - o(1)$.

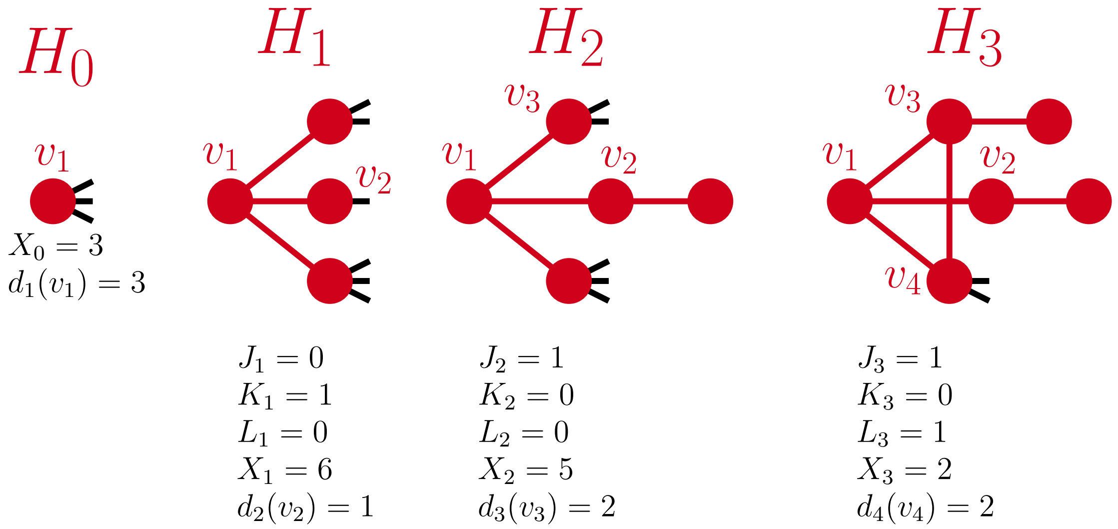

FIGURE 3. The starting vertex and the first 3 iterations of an exploration. The figures illustrate the start of each iteration. The H_i are depicted in red for $i = 0, \dots, 3$ and the open half-edges are black.

we build a subgraph H of $G[V(T)]$ and the submatching N of $\mathcal{R}_\mathcal{D}$ corresponding to H. In each iteration, for some vertex v' in the current tree we expose all the matching edges leaving half-edges incident to v' which have not already been exposed. We generate $G(\mathcal{D})$ and the tree at the same time, choosing the set of edges leaving a vertex so as to generate a uniformly chosen graph with the given degree sequence conditioned on the existence of the edges which have already been exposed.

At the start of iteration i we have exposed some subtree T_{i-1} of T and a subgraph H_{i-1} of the subgraph of G induced by T_i, which corresponds to a submatching N_{i-1} of N. We call a half-edge *open* when its endpoint has been revealed to belong to T, but the half-edge it is adjacent to in $R_\mathcal{D}$ has not yet been chosen. We let X_{i-1} be the number of open half-edges at the start of iteration i, and $d_i(v)$ is the number of open half-edges of v for every $v \in [n]$. Initially $T_0 = H_0$ is simply some fixed vertex $v = v_1$, and we have $d(v) = d_1(v_1)$ open half-edges corresponding to v, so $X_0 = d(v)$.

In iteration $1 \leq i \leq m$, provided $X_{i-1} > 0$, we choose a vertex v_i such that $d_i(v_i)$ is minimal among vertices $\{v \in V(H_{i-1}) : d_i(v_i) > 0\}$, breaking ties by prioritizing vertices with smaller labels, and reveal the edges of $\mathcal{R}_\mathcal{D}$ containing these half-edges. For each open half-edge h incident to v_i matched to a half-edge incident to a vertex w, we add $v_i w$ to H, and delete h from the set of open half-edges. If w was not in H we also add $v_i w$ to T and add the $d(w) - 1$ half-edges incident to w which were not matched to h to the set of open half-edges. If w is in T we delete the half-edge corresponding to w matched to h from the set of open half-edges.

Thus, if w is in T, the number of open half-edges decreases by 2, while if w is not in T, the number of open half-edges changes by $d(w) - 2$; so it decreases by 1 if $d(w) = 1$, stays the same if $d(w) = 2$, and increases otherwise.

We will have completely explored the component when X_i drops to 0. We treat X_i as the position of a random walk which begins at $X_0 = d(v)$ and where the i'th step has size $X_i - X_{i-1}$. Our focus is on the behaviour of this walk. As is typical, we first compute $\mathbf{E}\left[X_i - X_{i-1} \mid H_{i-1}\right]$ and then show X_i is concentrated around its conditional expected value.

To simplify the analysis of this walk, if there are no more open half-edges, we set $d_i(v_i) = X_i - X_{i-1} = 0$. This allows us to assume the walk takes m steps.

We let J_i, K_i, and L_i be, respectively, the number of half-edges incident to v_i which, during the i'th iteration, were exposed to be matched to a half-edge corresponding to a vertex of degree 1, a vertex of degree 2, or a vertex of H_{i-1}. Thus

$$X_i - X_{i-1} \geq d_i(v_i) - 2J_i - K_i - 3L_i \geq -2d_i(v_i).$$

The first step in our proof of Theorems 1.3 is to show in Section 3 that the probability there are two components of size $\Omega(m)$ is $O(m^{-8\log\log m})$. Equivalently we show that the probability that there are u and v in different components such that for both the exploration from u and v, at some iteration i, we have $X_i + |E(H_i)| = \Omega(m)$ is $O(m^{-8\log\log m})$. Our approach is guided by the intuition that if there are a large number of edges coming out of two disjoint sets, there will be at least one edge joining them with probability close to 1. This implies we need only show that if $X_i + |E(H_i)| = \Omega(m)$ for some iteration i, then there exists another iteration $1 \leq i' \leq i$ where $X_{i'} = \Omega(m)$.

If $n_1, n_2, X_{i-1} \leq \frac{m}{10^6}$ and $D^* \leq \frac{m}{100}$, then applying Lemma 1.4, we obtain that, conditioning on H_{i-1} throughout, each of J_i, K_i and L_i has expectation at most $\frac{4}{3(10^6)}$ and hence $\mathbf{E}[X_i - X_{i-1}] > (1 - 4(10^{-6}))d(v_i)$. Thus, the expected increase in the number of open half-edges during the iteration is nearly as large as the increase in the number of edges of H_i, and the expected increase in X_i is nearly half the expected increase in $X_i + |E(H_i)|$. We show that in this case, the walk's position is reasonably concentrated around its expectation. This allows us to show that if v is in a component of size $\Omega(m)$ then it is very likely that for some i, $X_i = \Omega(m)$.

If $n_1, n_2, X_{i-1} \leq \frac{m}{10^6}$ and $D^* \geq \frac{m}{100}$, then we are prevented from applying Lemma 1.4 and the above approach only because $D(v_i)$ may exceed $\frac{m}{100}$ and the degree $d(w)$ of the other vertex in Lemma 1.4 may exceed $\frac{\sqrt{m}}{10}$. We handle the second problem by accepting as a positive outcome the existence of a vertex of degree at least $\sqrt{m}(\log m)^{-2}$ in T, because we will show that all such vertices are in the same component whp when $D^* \geq \frac{m}{100}$. So, until this occurs, both v_i and all its neighbours in T have degree well below $\sqrt{m}$. Thus, if $D(v_i) > \frac{m}{100}$, then it is not hard to see v_i must have a neighbour of degree at least $\sqrt{m} > 5d(v_i)$ outside T. But this implies that $X_i - X_{i-1} > 5d_i(v_i) - 2 - 2J_i - K_i - 3L_i \geq 2d(v_i)$ which is an even better bound than what we needed. This allows us to show that if v is in a component of size $\Omega(m)$ then it is very likely that either for some i, $X_i = \Omega(m)$ or the component containing v contains a vertex of degree $\sqrt{m}(\log m)^{-2}$.

To complete this first step in our proof of Theorem 1.3, we show the following, which confirms the above intuitions:

(1) If $D^* \leq \frac{m}{100}$ then for any vertices u and v, the probability that u and v are in different components, but the explorations from both u, v reach $X_i = \Omega(m)$ at some point, is $O(m^{-8\log\log m})$.
(2) If $D^* > \frac{m}{100}$ then the probability that there exist two vertices of degree $\sqrt{m}(\log m)^{-2}$ in different components is $O(m^{-8\log\log m})$.
(3) Furthermore, if $D^* > \frac{m}{100}$ then for any vertices u, v where $d(u) \geq \sqrt{m}(\log m)^{-2}$, the probability that u, v are in different components, and the exploration from v reaches $X_i = \Omega(m)$ at some point, is $O(m^{-8\log\log m})$.

In Section 4, we provide a deeper analysis of the behaviour of one iteration, focusing on bounding the probability that J_i, K_i and L_i take specific values via switching arguments. In Section 5, we first use the results of Section 4 to bound the probability that there is a component of G whose size exceeds $4(\log m)^4$ but is $o(m)$. We then combine this result with the results of Section 3 to prove Theorem 1.3. Finally in Section 6, we use the results of Section 4 to analyze the probability of the existence of components of size less than $4(\log m)^4$ thereby completing the proof of Theorem 1.1.

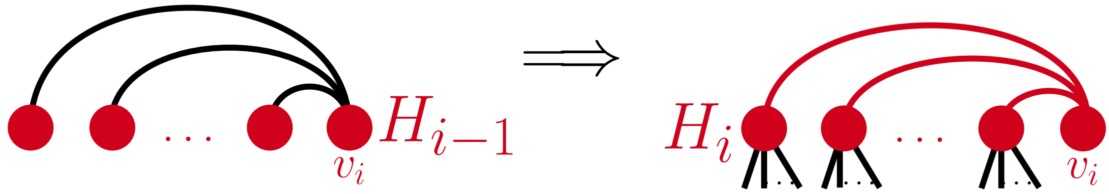

FIGURE 4. An illustration of $X_i \geq L_i(L_i - 1)$. The fact that v_i had the least amount of open half-edges forces each other vertex hit by a back edge to have at least $L_i - 1$ other open half-edges.

To conclude this section, we explain the reason for our choice of v_i as the vertex of T with open half-edges minimizing the number of such edges. This choice of v_i allows us to deterministically bound L_i which helps bound the maximum absolute value of a step. This is useful when proving the position of the walk is concentrated around its expected value. Specifically, as Figure 4 illustrates, by our choice of v_i, each of the L_i vertices of H_{i-1} to which v_i was directly matched during the i'th iteration was incident to at least $d_i(v_i) \geq L_i$ open half-edges at the start of the iteration, so $X_{i-1} \geq d_i(v_i)(L_i + 1)$. This implies that $X_i \geq L_i(L_i - 1)$ and that $L_i \leq \sqrt{X_{i-1}}$.

3. The Vertices in Large Components Are All in The Same Component

In this section we discuss one angle to analyze the exploration, which allows us to take our first step towards proving Theorem 1.3, by proving the following:

Lemma 3.1. *If m is large enough and $n_1, n_2 \leq \frac{m}{10^6}$, then for sufficiently small ϵ, the probability there are two components each of which has at least ϵm edges is at most $m^{-8 \log\log m}$.*

We start by proving a key lemma, Lemma 3.3, which shows that any two vertices u, v of degree at least $(\log m)^4$ and where $d(u)d(v) \geq m(\log m)^4$ are in the same component with high probability. Because the switching used in the proof, as depicted in Figure 6, requires u to have many non-leaf neighbours, we first prove Lemma 3.2, which states that no single vertex with $\Omega((\log m)^3)$ degree will have almost all neighbours be leaves.

Lemma 3.2. *Let $\mathcal{D}$ be such that m is large enough and $n_1 \leq n(1 - \log^{-1} m)$. Fix $u \in [n]$ with $d(u) \geq 100(\log m)^3$. Then the probability that u has no more than $\frac{d(u)}{2(\log m)^2}$ neighbours of degree at least 2 is at most $m^{-49 \log\log m}$.*

Proof. For every $0 \leq i \leq d(u)$ let $\mathcal{F}_i$ denote the event that u has i neighbours of degree at least 2. We consider switchings from $\mathcal{F}_i$ into $\mathcal{F}_{i+1}$ which swap $\{ua, bv\}$ with $\{uv, ba\}$, where $d(a) = 1$, v is a non-neighbour of u of degree at least 2, and b is a neighbour of v, as in Figure 5. Such a switching is always valid.

In $\mathcal{F}_i$, there are $d(u) - i$ choices of ua and $n - i - n_1$ choices of v (each giving at least one choice of bv), so there are at least $(d(u) - i)(n - i - n_1)$ switchings from $\mathcal{F}_i$ into $\mathcal{F}_{i+1}$. In $\mathcal{F}_{i+1}$, there are at most $i + 1$ choices of uv and at most n_1 choices of ba, so at most $n_1(i + 1)$ switchings from $\mathcal{F}_{i+1}$ back into $\mathcal{F}_i$. Thus for any $0 \leq i < \frac{d(u)}{(\log m)^2} - 1$, we have

$$\mathbf{P}\{\mathcal{F}_i\} \leq \mathbf{P}\{\mathcal{F}_{i+1}\} \frac{n_1(i+1)}{(d(u) - i)(n - i - n_1)} \leq \mathbf{P}\{\mathcal{F}_{i+1}\} \frac{n \frac{d(u)}{(\log m)^2}}{\frac{d(u)}{2} \cdot \frac{n}{2 \log m}} \leq \frac{4}{\log m}.$$

So for any fixed $0 \leq K \leq \frac{d(u)}{2(\log m)^2}$, $\mathbf{P}\{\mathcal{F}_K\} \leq (\frac{4}{\log m})^{50 \log m}$ and we union bound over the choices of K. □

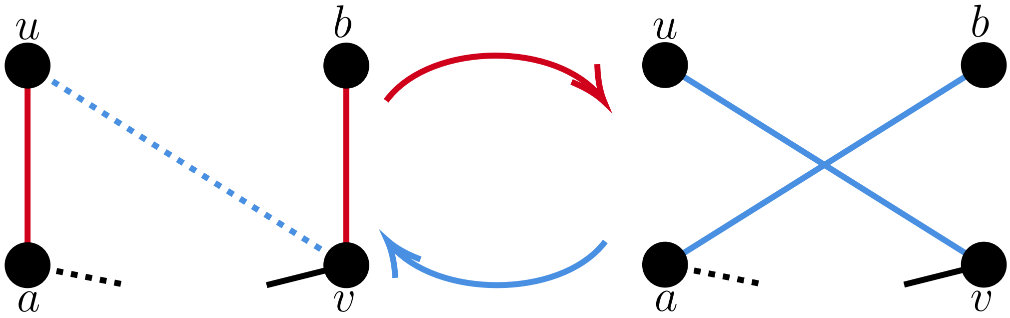

FIGURE 5. Switching used in the proof of Lemma 3.2: we switch away degree 1 neighbours of u (vertex a) one-by-one for non-neighbours of u of degree 2 or more (vertex v).

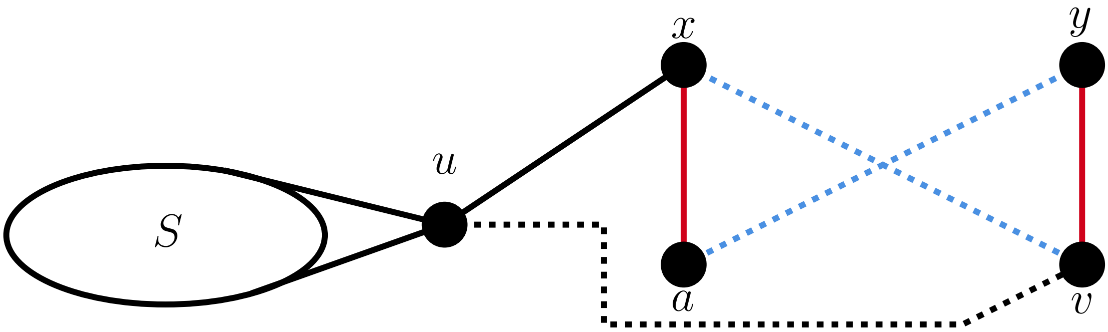

FIGURE 6. The switching used in the proof of Lemma 3.3. The vertex a is added into S after the switch. The dashed black lines indicate that no paths of length at most 4 exist between u and v outside of S, which implies the absence of the edges xv and ay.

Lemma 3.3. *Let $\mathcal{D}$ be such that m is large enough and $n_1 \leq n(1 - \log m)$. Fix $u, v \in [n]$ with $d(u), d(v) \geq (\log m)^4$ and such that $d(u)d(v) \geq m(\log m)^4$. Then the probability that u and v are not in the same component is at most $m^{-9 \log\log m}$.*

Proof. Wlog u has $d(u) \geq \sqrt{m}(\log m)^2$. Let $\mathcal{A}$ denote the event that u has at least $\frac{d(u)}{(\log m)^2}$ neighbours of degree at least 2. Let $\mathcal{B} = \mathcal{B}_0$ denote the event that u and v are not in the same component. For $i \geq 0$, let $\mathcal{B}_{i+1}$ denote the event where u and v have i common neighbours, and in the graph obtained by deleting these common neighbours and at most i other neighbours of u, there is no path of at most four edges from u to v. Let $\mathcal{F}_i = \mathcal{A} \cap \mathcal{B}_i$.

Fix $i \geq 0$. For every G in $\mathcal{F}_i$ we denote by $S = S(G)$ the set of at most i neighbours of u such that there is no path of length at most 4 between u and v in $G - S \cup (N(u) \cap N(v))$. We claim we can switch from a graph G in $\mathcal{F}_i$ to a graph G' by swapping $\{xa, yv\}$ with $\{xv, ya\}$ whenever $x \in N(u) \setminus (N(v) \cup S)$, $a \neq u$ and $y \in N(v) \setminus N(u)$, as in Figure 6. Note that since $x \notin S$ and there is no path of length at most 4 between u and v in $G - S \cup (N(u) \cap N(v))$, $ya \notin E(G)$ and our claim is true. In G' there are $i + 1$ common neighbours of u and v. Any induced path of length at most 4 between u and v either goes through $N(u) \cap N(v)$, or through both $N(u) \setminus N(v)$ and $N(v) \setminus N(u)$. Since no new vertices were added to $N(u) \setminus N(v)$ or $N(v) \setminus N(u)$ in G' from G, every path of at most 4 edges from u to v in $G - S \cup (N(u) \cap N(v))$ must use ya, and since none of yv, yu or av is an edge of G', the only possible such path is $uaybv$ for some common neighbour b of y and v. Adding a to S and noting that $N(u)$ does not change, we see that indeed $G' \in \mathcal{F}_{i+1}$.

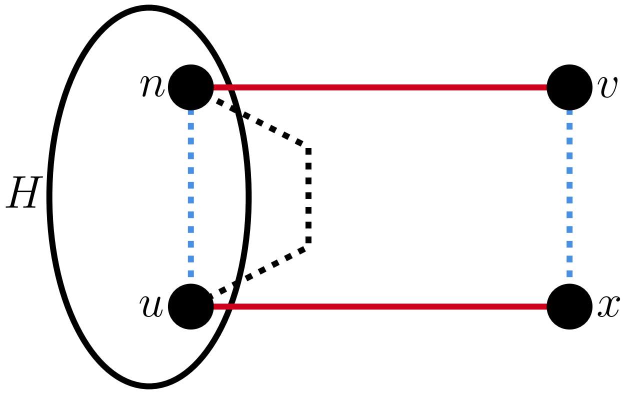

FIGURE 7. First switching in the proof of Lemma 3.5. The black lines indicate no paths of length at most 3 exist between n, u, implying the absence of the blue edges.

For any G in $\mathcal{F}_i$ for $i \leq 10 \log m$ there are at least $\frac{d(u)}{2(\log m)^2} - 2i > \frac{d(u)}{3(\log m)^2}$ choices of xa and at least $d(v) - 2i > \frac{d(v)}{2}$ choices of yv. On the other hand, for a G' in $\mathcal{F}_{i+1}$, there are at most $2(i+1)m$ choices of $\{xv, ya\}$. Thus for $0 \leq i < 10 \log m$,

$$\mathbf{P}\{\mathcal{F}_i\} \leq \mathbf{P}\{\mathcal{F}_{i+1}\} \frac{20m \log m}{\frac{d(u)d(v)}{6(\log m)^2}} \leq \mathbf{P}\{\mathcal{F}_{i+1}\} \frac{120}{\log m}$$

so $\mathbf{P}\{\mathcal{F}_0\} \leq (\frac{120}{\log m})^{10 \log m} < m^{-9.5 \log \log m}$. By Lemma 3.2, it follows that $\mathbf{P}\left\{\mathcal{A}^{\complement}\right\} < m^{-49 \log \log m}$ so $\mathbf{P}\{\mathcal{B}_0\} < m^{-9.5 \log \log m} + m^{-49 \log \log m}$. □

In particular, when $d(n)$ is high enough, all vertices of degree at least $\sqrt{m}(\log m)^{-2}$ are in the same component directly by Lemma 3.3.

Corollary 3.4. *Let $\mathcal{D}$ be such that m is large enough, $n_1 \leq n(1 - \log m)$, and $d(n) \geq \sqrt{m}(\log m)^6$. Then with probability $1 - o(m^{-8 \log \log m})$, all the vertices of degree at least $\sqrt{m}(\log m)^{-2}$ are in the same component.*

Proof. Applying Lemma 3.3 to $d(n)$ and any vertex v with $d(v) \geq \sqrt{m}(\log m)^{-2}$, the probability v is not in the same component as $d(n)$ is at most $m^{-9 \log \log m}$. Summing up over all the at most m choices for v yields the desired result. □

The next lemma states that when $D^* \geq \epsilon m$ but $d(n) \leq \sqrt{m}(\log m)^6$, the vertices of degree at least $\sqrt{m}(\log m)^{-2}$ are also in the same component. Lemma 3.5 and Corollary 3.4 together show that when $D^* \geq \epsilon m$ or $d(n) \geq \sqrt{m}(\log m)^6$, every exploration which reaches a vertex of degree at least $\sqrt{m}(\log m)^{-2}$ is exploring the same component.

Lemma 3.5. *For all sufficiently small $\epsilon > 0$, the following holds. Let $\mathcal{D}$ be such that m is large enough, $d(n) \leq \sqrt{m}(\log m)^6$, and $D^* \geq \epsilon m$. Then with probability $1 - o(m^{-8 \log \log m})$, all the vertices of degree at least $\sqrt{m}(\log m)^{-2}$ are in the same component.*

Proof. We let H be the set of vertices of degree at least $\frac{\sqrt{m}}{log^7 m}$. Our hypotheses imply that the sum of the degrees of the vertices in H is at least $\frac{\epsilon m}{2}$ and that $|H| \geq \frac{\sqrt{m}}{log^{20} m}$.

We claim that the probability that there is no vertex of H within distance three of at least $m^{1/8}$ vertices of H is at most $m^{-10 \log \log m}$.

To prove our claim, for $i \leq 2m^{1/8}$, we let $\mathcal{F}_i$ be the set of graphs for which n is adjacent to i vertices of H, no vertex except n is incident to more than $m^{1/8}$ vertices of $H - n$ and there are at most $2(i+1)m^{1/8}$ vertices of H at distance within 3 of any vertex of H

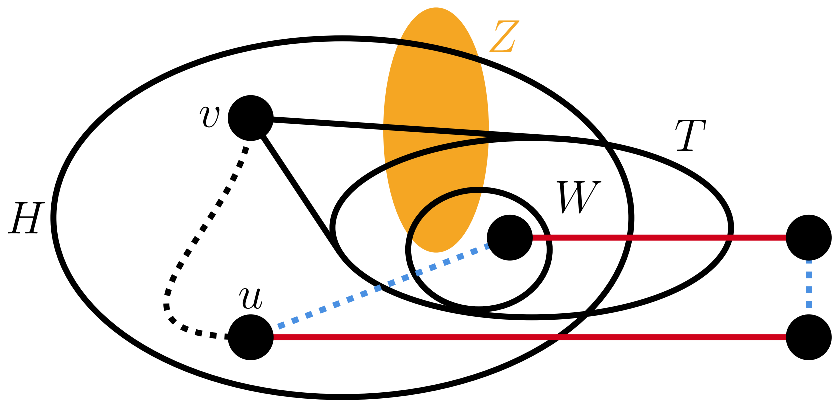

FIGURE 8. Second switching in the proof of Lemma 3.5. Every switching makes a member of W into a neighbour of u. The dotted black line indicates there are no paths between u, v outside of Z, which implies the absence of the blue edges.

in $G - n$ and at most $2(i+1)^2 m^{1/8}$ vertices of H at distance within 3 of n. We show $\mathbf{P}\left\{\cup_{i \le m^{1/8}} \mathcal{F}_i\right\} \le m^{-12 \log\log m}$.

To do so, for $i \le 2m^{1/8}$, we consider a swap from a graph in $\mathcal{F}_i$ using edges nv, ux where v is a neighbour of n not in H, u is a vertex of H at distance at least four from n and x is a neighbour of u not in H as depicted in Figure 7. We obtain a graph G' such that n is adjacent to $i+1$ vertices of H and no vertex except n is adjacent to more than $m^{1/8}$ vertices of $H - n$. For any w in $H - n$, any vertices of H which are now at distance 3 from w in $G - n$ but were not previously must be joined to w by a path with three edges whose internal vertices are v and x. Hence they must be one of the at most $2m^{1/8}$ vertices in $H - n$ which are neighbours of v or x. Any vertices of H which are now at distance 3 from n in G but were not previously must either have been at distance at most 3 from u in $G - n$ or be joined to w by a path with three edges whose internal vertices are v and x. But the latter is impossible as neither x nor v is a neighbour of n in G'. So $G' \in \mathcal{F}_{i+1}$.

Now, there are at least $\frac{\sqrt{m}}{\log^8 m} - i > m^{3/7}$ choices for v. There are at least $|H| - 2(i+1)^2 m^{1/8} > m^{3/7}$ choices for u. For each such choice there are at least $\frac{\sqrt{m}}{\log^8 m} - m^{1/8} - 1 > m^{3/7}$ choices for x. Hence there are at least $m^{9/7}$ choices for our swap. on the other hand, for any G' in $\mathcal{F}_{i+1}$ there are at most $2(i+1)m < m^{8/7}$ swaps from G' to graphs in $\mathcal{F}_i$.

So for each $i \le 2m^{1/8}$, we have $\frac{|\mathcal{F}_i|}{|\mathcal{F}_{i+1}|} \le m^{-1/7}$ and our claim is proved.

We next claim that for any v and $u \in H$, the probability that u and v are in different components and there are at least $m^{1/8}$ vertices of H within distance 3 of v is at most $m^{-12 \log\log m}$. The lemma follows from this, so it remains to prove the claim.

To this end we let $\mathcal{F}_i$ be the event that there are there are at least $m^{1/8}$ vertices of H within distance 3 of v, there are i edges between these vertices and u and there is a set Z of at most $2i$ vertices such that u and v are in different components of $G - Z$.

For $i \le \log^2 m$, and any graph in $\mathcal{F}_i$ we let T be a tree of height three obtained starting from the one-vertex tree containing only v by repeatedly adding a vertex of H at distance at most three from v and a subset of a shortest path from this vertex to v until the tree contains a set W of at least $m^{1/8}$ vertices of H at distance at most 3 from v, i of which are adjacent to u (these latter are added first). We add at most three vertices to T in each iteration so $|T| \le 3m^{1/8} + 1$. We can obtain a graph in $\mathcal{F}_{i+1}$ by swapping any pair of

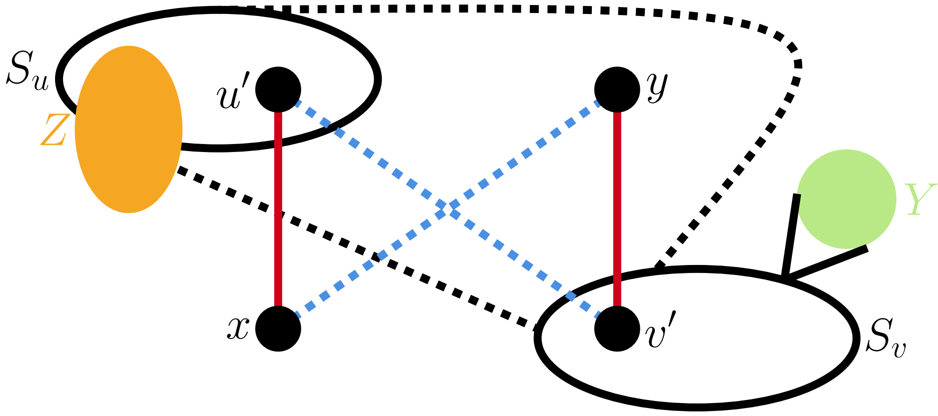

FIGURE 9. Switching used in the proof of Lemma 3.6. The black dotted lines indicate that there are no edge between Z and S_v outside of Y, and no paths between S_u and S_v outside of Z. The edge $yv' \notin Y$, so $y \notin Z$. The vertices u' and x are not in Z, so the fact that there are no paths between S_u and S_v outside of Z ensures the absence of $u'v'$ and xy.

edges consisting of (i) an edge from one of the at least $m^{1/9}$ vertices of W nonadjacent to u to one of its at least $\frac{\sqrt{m}}{\log^7 m} - 3m^{1/8} - 1 - 2i > m^{13/27}$ neighbours outside $Z \cup T$ and (ii) any of the at least $\frac{\sqrt{m}}{\log^7 m} - 3m^{1/8} - 1 - 2i > m^{13/27}$ edges from u to a neighbour outside $Z \cup T$, as depicted in Figure 8. So, there are at least $m^{29/27}$ such swaps from G. On the other hand there are at most $(i+1)m$ swaps to $\mathcal{F}_i$ from a G' in $\mathcal{F}_{i+1}$. Our claim then follows in the standard way. □

We now have a strategy to prove Lemma 3.1, by exploring out of every $v \in [n]$ as described in Section 2. We have two cases based on D^* and $d(n)$:

- If $D^* \geq \epsilon m$ or $d(n) \geq \sqrt{m}(\log m)^6$, explore until $X_i \geq \epsilon m$ or $X_i = 0$, or until a vertex of degree at least $\sqrt{m}(\log m)^{-2}$ is reached. This means $V(H_i)$ has maximum degree less than $\sqrt{m}(\log m)^{-2}$ until the iteration where a vertex of degree at least $\sqrt{m}(\log m)^{-2}$ is reached.
- If $D^* < \epsilon m$ and $d(n) < \sqrt{m}(\log m)^6$, explore until $X_i \geq \epsilon m$ or $X_i = 0$.

To apply the strategy, we prove another key lemma which shows that any two vertices whose explorations reach $\Omega(m)$ are with high probability in the same component, and so are any vertex whose exploration reaches $\Omega(m)$ and any vertex of degree exceeding $\sqrt{m}(\log m)^{-2}$ (we will refer to such vertices as *high degree vertices* in the remainder of this section). Together, this tells us that we may stop an exploration if it reaches ϵm or hits a high degree vertex.

Lemma 3.6. *Let* $0 < \epsilon < 10^{-6}$ *and let* $\mathcal{D}$ *be such that* m *is large enough and* $n_1 \leq (1 - \frac{1}{\log m})n$*. Fix* $u, v \in [n]$ *and let* $\mathcal{F}$ *denote the event that* u, v *are not in the same component, the exploration starting from* u *has boundary size reaching at least* ϵm *without reaching* 0 *and either (i) the same holds for* v *or (ii)* v *is a vertex of degree exceeding* $\sqrt{m}(\log m)^{-2}$*. Then* $\mathbf{P}\{\mathcal{F}\} < m^{-7\log\log m}$*.*

Proof. We let $\mathcal{A}$ denote the event that no vertex of degree at least $\sqrt{m}(\log m)^{-2}$ lies in a different component from any vertex of degree at least $\sqrt{m}(\log m)^6$. We let $\mathcal{B}$ be the event that (ii) holds and v has fewer than $\sqrt{m}(\log m)^{-5}$ non-leaf neighbours. By Corollary 3.4 and Lemma 3.2, $\mathbf{P}\left\{\mathcal{A}^{\complement} \cup \mathcal{B}\right\} = o(m^{-8\log\log m})$, so it is enough to show $\mathbf{P}\{(\mathcal{F} \cap A) \setminus \mathcal{B}\} < m^{-8\log\log m}$. We prove this conditioned on the choice of H_u, which is the exploration from u up to the time at which the boundary first reached size ϵm, and H_v, which is defined symmetrically if (i) holds and is simply $H_1(v)$ if (ii) holds. We use $\mathcal{F}' = \mathcal{F}'(H_u, H_v)$ to denote the event that $\mathcal{F}$ holds with a specific choice of H_u, H_v. We use $\mathcal{F}_0 = (\mathcal{F}' \cap \mathcal{A}) \setminus \mathcal{B}$.

For every $G \in \mathcal{F}'$, if (i) holds let $S_u = S_u(G), S_v = S_v(G)$ denote the set of unexplored vertices in H_u, H_v respectively when their boundary size first reaches ϵm. If (ii) holds, the definition of S_u is unchanged but define $S_v = \{w \in N(v) : d(w) \geq 2\}$ instead. If (ii) holds, we only consider H_v satisfying the event $\{|S_v| \geq \sqrt{m}(\log m)^{-5}\}$ as for other choices, $\mathcal{F}_0$ is empty. Relabelling u and v if necessary, we can assume that for every graph in $\mathcal{F}_0$, no vertex in the component containing u has degree exceeding $\sqrt{m}(\log m)^6$.

For $i \geq 0$ let $\mathcal{F}_{i+1}$ denote the set of graphs with the given choice of H_u and H_v for which the number of edges between S_u and S_v is between i and $2i$, and there is a set $Z \subseteq [n] \setminus S_v$ of at most $2i$ vertices each of degree at most $\sqrt{m}(\log m)^6$, and a set Y of at most $2i$ edges incident to S_v, such that the following hold:

1. There is no path from S_u to the union of S_v and the set of vertices of degree exceeding $\sqrt{m}(\log m)^6$ in $G - Z$, and
2. All edges between Z and S_v are in Y.

For $0 \leq i < m^{1/3}$, if $G \in \mathcal{F}_i$ then there are at least $\epsilon m - 2i(\sqrt{m}(\log m)^6 + 1)$ unexplored edges $u'x$ with $u' \in S_u \setminus Z$ and $x \notin Z$, and at least $\sqrt{m}(\log m)^{-5} - 2i$ unexplored edges with $v' \in S_v$ and $yv' \notin Y$. Note that $u'v'$ is not an edge, as there is no path from S_u to S_v in $G - Z$; since $v'y \notin Y$, y is not in Z so yx is not an edge, as $u'xyv'$ cannot be a path from S_u to S_v in $G - Z$. So the switching is valid. By the definition of Z, both u' and x have degree at most $\sqrt{m}(\log m)^6$, and neither vertex is in S_v. Furthermore, after switching $\{u'x, yv'\}$ to get $\{u'v', yx\}$, the only edge between u' and S_v is $u'v'$, and the only possible edge between x and S_v is yx, since there were no such edges before the switching. Thus the resulting graph is in $\mathcal{F}_{i+1}$ by putting u', x' in Z and (at most) $u'v', xy$ in Y. So for every $G \in \mathcal{F}_i$ there are at least $m^{13/9}$ switchings to graphs in $\mathcal{F}_{i+1}$.

There are at most $2(i+1)$ edges between S_u, S_v, so there are at most $4(i+1)m$ switchings from $\mathcal{F}_{i+1}$ to $\mathcal{F}_i$. So

$$\mathbf{P}\{\mathcal{F}_0 \mid H_u, H_v\} < \left(\frac{4m(m^{1/3})}{m^{13/9}}\right)^{m^{1/3}} < m^{-m^{1/4}}.$$

Thus $\mathbf{P}\{\mathcal{F} \cap \mathcal{A} - \mathcal{B}\} < m^{-m^{1/4}}$. □

We now move onto studying the process X_i for any starting vertex $v \in [n]$. The goal is to show that if v is in a linear sized component, then for some i, X_i will be $\Omega(m)$ or H_i will contain a high degree vertex, via a martingale argument. Since $X_i - X_{i-1}$ does not have a good upper bound, we study a truncated version of $X_i - X_{i-1}$, that we call X_i^*. For $1 \leq i \leq m$, if $X_j = 0$ for some $j \leq i-1$, set $X_i^* = 0$, and otherwise define:

$$X_i^* = \begin{cases} \min(3d_i(v_i), X_i - X_{i-1}), & \text{if } d_i(v_i) < \frac{\sqrt{m}}{(\log m)^2}; \\ 0.9 d_i(v_i), & \text{if } d_i(v_i) \geq \frac{\sqrt{m}}{(\log m)^2}. \end{cases}$$

Define $Y_0 = 0$, and for $1 \leq i \leq m$ let

$$Y_i = Y_{i-1} + X_i^* - \mathbf{E}[X_i^* \mid H_{i-1}].$$

Lemma 3.7. *If $\mathcal{D}$ is such that m is large enough, then for any vertex $v \in [n]$, for all $0 \leq i \leq m$, $\mathbf{P}\left\{Y_i \leq -\frac{m}{\log m}\right\} < e^{-m^{1/10}}$.*

Proof. By definition $\{Y_i\}_{i=0}^m$ is a martingale with respect to $\{H_i\}_{i=0}^m$, and $|Y_i - Y_{i-1}| = |X_i^* - \mathbf{E}[X_i^* \mid H_{i-1}]|$ is 0 if $d_i(v_i) > \frac{\sqrt{m}}{(\log m)}$ and is at most $3d_i(v_i) \leq \frac{3\sqrt{m}}{(\log m)^2}$ otherwise.

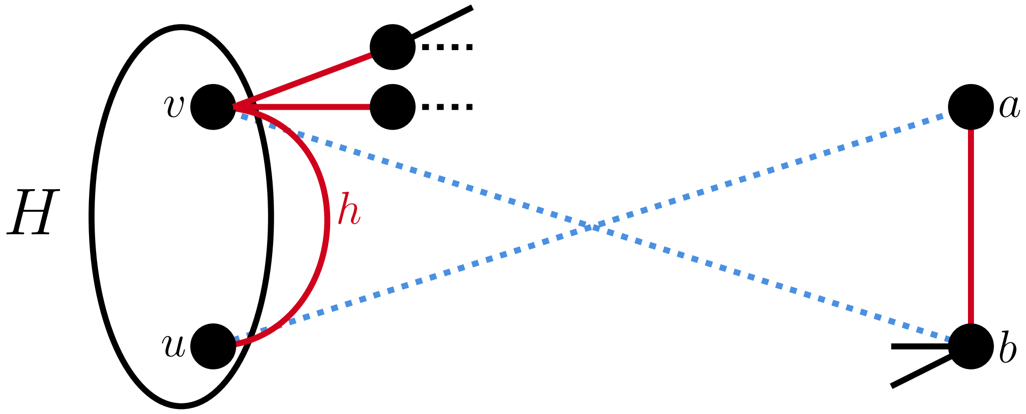

FIGURE 10. Switching used in the proof of Lemma 3.8. Here, the three types of "bad" edges are all illustrated: an edge not in $E(H)$ from v back into $V(H)$, or onto a degree 1 or 2 vertex. Vertex b has degree at least 3.

Furthermore,

$$\sum_{j\le i: d_j(v_j)<\frac{\sqrt{m}}{(\log m)^2}} (3d_j(v_j))^2 \le \frac{3\sqrt{m}}{(\log m)^2}\sum_{j\le i} d_j(v_j) \le \frac{3m^{3/2}}{(\log m)^2}$$

so by Azuma's inequality [1],

$$\mathbf{P}\left\{Y_i < -\frac{m}{\log m}\right\} < \exp\left\{\frac{-m^2(\log m)^{-2}}{3m^{3/2}(\log m)^{-2}}\right\} < e^{-m^{1/10}}. \qquad \square$$

Now, it remains to show that (1) X_i^* is a good proxy for $X_i - X_{i-1}$, and (2) lower bound $\mathbf{E}[X_i^* \mid H_{i-1}]$. Our proofs of both (1) and (2) rely on the fact that the probability that a single open half-edge gets matched either onto a vertex of degree no more than 2, or onto $V(H_{i-1})$, is small. This fact is shown by the following lemma.

Lemma 3.8. *Fix $0 < \epsilon < 10^{-6}$. Let $\mathcal{D}$ be such that m is large enough and $n_1, n_2 \le 10^{-6}m$. Let H be a possible subgraph of $G(\mathcal{D})$ and let h be a half-edge out of $v \in V(H)$ and not matched in $E(H)$. Suppose H is such that the following hold:*

(1) $\sum_{w\in V(H)} d(w) \le \epsilon m$, and
(2) Either $D^* \le \epsilon m$, or $d(w) < \sqrt{m}(\log m)^{-2}$ for all $w \in V(H)$.

Let $\mathcal{F}$ denote the event that h is matched onto $V(H)$ or onto a vertex of degree at most 2. Let $\mathcal{A}$ denote the event that no neighbour of v outside of $V(H)$ has degree $10d(v)$ or more. Then $\mathbf{P}\{\mathcal{F}\cap\mathcal{A} \mid H \subseteq G(\mathcal{D})\} < 0.001$. If $D^* \le \epsilon m$, then $\mathbf{P}\{\mathcal{F} \mid H \subseteq G(\mathcal{D})\} < 0.001$.

Proof. We consider the switching from $\mathcal{F}\cap\mathcal{A}$ (or from $\mathcal{F}$ if $D^* \le \epsilon m$) into $\mathcal{F}^{\complement}$ swapping $\{vu, ab\}$ with $\{vb, au\}$, where h is matched onto u, which is either in $V(H)$ or has degree at most 2, and ab is such that b has degree at least 3 and is not in $V(H) \cup N(v)$, and a is neither u nor a neighbour of u (see Figure 10). Note that vu, ab are not in $E(H)$ so H being a subgraph is preserved.

Let $G \in \mathcal{F}\cap\mathcal{A}$ (or $G \in \mathcal{F}$ if $D^* \le \epsilon m$) such that $H \subseteq G$. For the choice of ab, we exclude the edges in $E(G)\setminus E(H)$ which are incident to $V(H)\cup N(v)$ or to a vertex of degree at most 2, or have both endpoints in $N(u)\cup\{u\}$. Orient any non-excluded edge ab such that $a \notin N(u)\cup\{u\}$; then b is not in $V(H)\cup N(v)$ and has degree at least 3, so such ab is a valid choice for switching with vu. There are:

- at most $\sum_{w\in V(H)} d(w) \le \epsilon m$ edges incident to $V(H)$;
- at most $\sum_{w\in N(v)} d(w)$ edges incident to $N(v)$;

- at most $\max\left(\binom{d(u)+1}{2}, \sum_{w\in N(u)} d(w)\right)$ edges inside $N(u)\cup\{u\}$; and
- at most $n_1 + 2n_2 < 10^{-5}m$ edges incident to a vertex of degree at most 2.

Now, if $D^* \le \epsilon m$, then $\sum_{w\in N(v)} d(w), \sum_{w\in N(u)} d(w) \le D^* \le \epsilon m$. Otherwise $d(w) < \sqrt{m}(\log m)^{-2}$ for all $w \in V(H)$, so since $u \in V(H)$ or has degree at most 2, $d(v), d(u) < \sqrt{m}(\log m)^{-2}$. Furthermore, in this case $G \in \mathcal{A}$ so every $w \in N(v)$ has degree at most $10d(v)$, so $\sum_{w\in N(v)} d(w) \le 10(d(v))^2 < 10m(\log m)^{-4} < \epsilon m$ and $\binom{d(u)+1}{2} < \epsilon m$. In both cases, we forbid at most $(3\epsilon + 10^{-5})m$ edges from $E(G)\setminus E(H)$. So there at least $(1-4\epsilon-10^{-5})m$ choices of $\{vu, ab\}$ to switch from $\mathcal{F}$ to $\mathcal{F}^{\complement}$.

From $\mathcal{F}^{\complement}$ to $\mathcal{F}$, since au is an edge incident to $V(H)$ or to a vertex of degree at most 2, there are at most $n_1 + 2n_2 + \sum_{w\in V(H)} d(w) < (10^{-5}+\epsilon)m$ choices of $\{vb, au\}$. Thus

$$\mathbf{P}\{\mathcal{F}\cap\mathcal{A} \mid H \subseteq G(\mathcal{D})\} \le \mathbf{P}\left\{\mathcal{F}^{\complement} \,\middle|\, H \subseteq G(\mathcal{D})\right\} \frac{(10^{-5}+\epsilon)m}{(1-4\epsilon-10^{-5})m} < 0.001$$

and similarly $\mathbf{P}\{\mathcal{F} \mid H \subseteq G(\mathcal{D})\} < 0.001$ if $D^* \le \epsilon m$. □

Corollary 3.9. *Fix $0 < \epsilon < 10^{-6}$. Let $\mathcal{D}$ be such that m is large enough and $n_1, n_2 \le 10^{-6}m$. Let $v \in [n]$ be fixed. If, for $i \ge 1$, after exposing the first $i-1$ iterations, we have:*

(1) $X_{i-1} \ne 0$,
(2) $2|E(H_{i-1})| + |X_{i-1}| \le \epsilon m$,
(3) Either $D^* \le \epsilon m$ or $d(w) < \sqrt{m}(\log m)^{-2}$ for all $w \in V(H_{i-1})$

then $\mathbf{E}[X_i^* \mid H_{i-1}] \ge 0.9 d_i(v_i)$.

Proof. We condition on the first $i-1$ iterations such that (1), (2), (3) hold, and the choice of v_i (which are completely determined by H_{i-1}) and work within this space.

If $d_i(v_i) \ge \sqrt{m}(\log m)^{-2}$ then $X_i^* = 0.9d_i(v_i)$ by the definition of X_i^* so we assume this is not the case. We need to show $\mathbf{E}[\min(3d_i(v_i), X_i - X_{i-1})] \ge 0.9d_i(v_i)$. We let Bad be the event that v_i is not adjacent to a vertex outside H_{i-1} of degree $5d_i(v_i)$. For each unexplored half-edge $h \in S$ where S consists of the $d_i(v_i)$ unexplored half-edges incident to v_i, we let E_h be the event that h is adjacent to a vertex of degree at most 2 or a vertex of H_{i-1}. We note that $d_i(v_i) - \sum_{h\in S} 3(\mathbf{1}_{[E_h]}) \le \min(3d_i(v_i), X_i - X_{i-1})$ deterministically. When Bad fails, $\min(3d_i(v_i), X_i - X_{i-1}) \ge 3d_i(v_i)$. Thus, the expectation is at least

$$\begin{aligned}&3d_i(v_i)(1-\mathbf{P}\{\text{Bad}\}) + d_i(v_i)\mathbf{P}\{\text{Bad}\} - \sum_{h\in S} 3\mathbf{P}\{E_h \cap \text{Bad}\}\\ &\ge d_i(v_i) - \sum_{h\in S} 3\mathbf{P}\{E_h\cap\text{Bad}\}.\end{aligned}$$

Since $|S| = d_i(v_i)$, it suffices to show for each $h \in S$, $\mathbf{P}\{E_h \cap \text{Bad}\} < 0.1$; and we can see that this holds by applying the first bound of Lemma 3.8. □

Corollary 3.10. *Fix $0 < \epsilon < 10^{-6}$. Let $\mathcal{D}$ be such that m is large enough and $n_1, n_2 \le 10^{-6}m$. Let $v \in [n]$. Then for any $1 \le i \le m$, if i is such that $d_i(v_i) > \sqrt{m}(\log m)^{-2}$ and*

(1) $X_{i-1} \ne 0$,
(2) $2|E(H_{i-1})| + |X_{i-1}| \le \epsilon m$, and
(3) $D^* \le \epsilon m$,

then $\mathbf{P}\{X_i - X_{i-1} < 0.9d_i(v_i) \mid H_{i-1}\} < m^{-m^{1/3}}$.

Proof. Let $H'_0 = H_i$. For $1 \le j \le d_i(v_i)$, let h_j be the lowest index half-edge on v_i out of H'_{j-1}, and form H'_j from H'_{j-1} by adding the edge containing h_j. Let $\mathcal{F}_j$ denote the event that h_j is matched onto $V(H'_{j-1})$ or onto a vertex of degree at most 2. By Lemma 3.8, if (1) (2) (3) hold then $\mathbf{P}\left\{\mathcal{F}_j \mid H'_{j-1}\right\} < 0.001$. Let $\mathcal{B}_i$ denote the event that at least $0.01d_i(v_i)$ events among $\mathcal{F}_1, \dots, \mathcal{F}_{d_i(v_i)}$ occur. Then

$$\mathbf{P}\left\{X_i - X_{i-1} < 0.9d_i(v_i) \mid H_{i-1}\right\} \leq \mathbf{P}\left\{\mathcal{B}_i \mid H_{i-1}\right\} < \binom{d_i(v_i)}{0.01d_i(v_i)} 0.001^{0.01d_i(v_i)}$$
$$< \left(\frac{0.001e}{0.01}\right)^{0.01d_i(v_i)} < m^{-m^{1/3}},$$

the final bound holding when $d_i(v_i) > \sqrt{m}(\log m)^{-2} > 100m^{1/3}\log_{10/e} m$. □

We are now ready to combine our lemmas to prove if v is in a linear sized component then X_i will reach $\Omega(m)$ or the exploration will reach a high degree vertex.

Corollary 3.11. *Fix $0 < \epsilon < 10^{-6}$ and let $\mathcal{D}$ be such that m is large enough and $n_1, n_2 \leq 10^{-6}m$. Let $v \in [n]$ be fixed. Then the probability that there exists an i with $1 \leq i \leq m$ such that $2|E(H_{i-1})| + |X_{i-1}| \leq \epsilon m$, and such that neither*
(a) $X_i \geq 0.9\sum_{j=1}^{i} d_j(v_j) - \frac{m}{\log m}$, nor
(b) $D^* \geq \epsilon m$ and H_i contains a vertex of degree at least $\sqrt{m}(\log m)^{-2}$
holds, is at most $e^{-m^{1/11}}$.

Proof. Unless there is some $j \leq i$ such that $X_j^* > X_j - X_{j-1}$ then by Corollary 3.9 we have

$$X_i \geq \sum_{j=1}^{i} X_j^* = Y_i + \sum_{j=1}^{i} \mathbf{E}\left[X_j^* | H_{j-1}\right] \geq Y_i + 0.9\sum_{j=1}^{i} d_j(v_j).$$

So, by our observation, the probability we are bounding is at most

$$\mathbf{P}\left\{Y_i < \frac{-m}{\log m}\right\} + \mathbf{P}\left\{\exists 1 \leq j \leq i : X_j^* > X_j - X_{j-1}\right\}.$$

Lemma 3.7 implies $\mathbf{P}\left\{Y_i < \frac{-m}{\log m}\right\} < e^{-m^{1/10}}$, and Corollary 3.10 and the definition of X_j^* together imply $\mathbf{P}\left\{\exists 1 \leq j \leq i : X_j^* > X_j - X_{j-1}\right\} \leq m(m^{-m^{1/3}})$. □

Corollary 3.12. *The probability that there is a vertex v in a component with at least ϵm edges for which both*
(a) X_i is never more than $\frac{\epsilon m}{8}$, and
(b) $D^* < \epsilon m$ or the component contains no vertex of degree greater than $\sqrt{m}(\log m)^2$
hold, is at most $me^{-m^{1/11}}$.

Proof. Note that $|E(H_{i-1})|$ is at most $\sum_{j=1}^{i-1} d_j(v_j)$ for all $i \geq 1$. We consider the first i for which $\sum_{j=1}^{i} d_j(v_j) \geq \frac{\epsilon m}{4}$. Observe that if $2|E(H_{i-1})| + X_{i-1} \geq \epsilon m$, (a) fails. Applying Corollary 3.11 we are done. □

We are now in a position to prove the main result of this section, Lemma 3.1.

Proof of Lemma 3.1. First suppose $D^* < \epsilon m$. By Corollary 3.12, with probability at least $1 - me^{-m^{1/11}}$, every vertex in a component with at least ϵm edges has its X_i reach $\frac{\epsilon m}{8}$, so by Lemma 3.6 these vertices are also all in the same component with probability at least $1 - m^{-7\log\log m + 2}$.

Now suppose $D^* \geq \epsilon m$ but $d(n) \geq \sqrt{m}(\log m)^6$. By Corollary 3.4, all vertices of degree at least $\sqrt{m}(\log m)^{-2}$ are in the same component with probability at least $1 - m^{-8\log\log m}$. If $D^* \geq \epsilon m$ but $d(n) < \sqrt{m}(\log m)^6$, then by Lemma 3.5, again all vertices of degree at least $\sqrt{m}(\log m)^{-2}$ are in the same component with probability at least $1 - m^{-8\log\log m}$. By Corollary 3.12, with probability at least $1 - me^{-m^{1/11}}$, every vertex in a component with ϵm edges is either in a component with a vertex of degree greater than $\sqrt{m}(\log m)^{-2}$, or X_i reaches $\frac{\epsilon m}{8}$. Finally, by Lemma 3.6, the probability that there exists a vertex whose exploration reaches $\frac{\epsilon m}{8}$ in a different component from any vertex of degree $\sqrt{m}(\log m)^{-2}$ or from another vertex whose exploration also reaches $\frac{\epsilon m}{8}$ is at most $m^{-7\log\log m+2}$. □

4. Analyzing Iteration by Iteration

In this section we carry out our second angle of analysis of the exploration $(X_i)_{i\geq 0}$ out of an arbitrary fixed vertex $v \in [n]$, as defined in Section 2. We consider each iteration in turn and focusing on the the random variable $X_i - X_{i-1}$ especially on its behaviour when it is negative. The results we obtain will be used both to complete the proof of Theorem 1.3, which is done in the next section, and to bound the probability that there is a component of size at most $4(\log m)^4$ which we do in the section after that.

We recall that $X_i - X_{i-1}$ is at least $d_i(v_i) - 2J_i - K_i - 3L_i$. Thus our focus is on the behaviour of J_i, K_i, and L_i. We handle L_i separately from J_i and K_i. Given an iteration $i \geq 1$ and $j,k \geq 0$ integers, define $\text{Bad}_{j,k}^{(i)}$ to be the event that $X_i \leq X_{i-1}$, $J_i \geq j$, and $K_i \geq k$.

We note that if $X_i \leq X_{i-1}$ then $J_i + K_i + L_i \geq \frac{d_i(v_i)}{3}$, which implies either $L_i \geq \frac{d_i(v_i)}{6}$ or $\text{Bad}_{j,k}^{(i)}$ occurs for some nonnegative j,k which sum to at least $\frac{d_i(v_i)}{6}$. We will now bound the probability of $\text{Bad}_{j,k}^{(i)}$ conditioned on H_{i-1}.

Lemma 4.1. *Let $0 < \epsilon < 10^{-6}$ and assume $\mathcal{D}$ is such that m is large enough and that $n_1, n_2 < \epsilon m$. Suppose that for some $i \geq 1$, $|E(H_{i-1})| \leq \frac{m}{10^5}$ and $X_{i-1} < 10^{-4}m$. Then for any nonnegative j,k with $j+k \leq \frac{\sqrt{m}}{1000}$, the probability, conditioning on H_{i-1}, of the event $Bad_{j,k}^{(i)}$ is at most*

$$\frac{jm}{jm - 2ed_i(v_i)n_1}\left(\frac{2ed_i(v_i)n_1}{jm}\right)^j \frac{km}{km - 10ed_i(v_i)n_2}\left(\frac{10ed_i(v_i)n_2}{km}\right)^k \tag{4.1}$$

if $j,k > 0$, $2ed_i(v_i)n_1 < jm$, and $10ed_i(v_i)n_2 < km$; at most

$$\frac{km}{km - 10ed_i(v_i)n_2}\left(\frac{10ed_i(v_i)n_2}{km}\right)^k \tag{4.2}$$

if $j = 0$ and $10ed_i(v_i)n_2 < km$; and at most

$$\frac{jm}{jm - 2ed_i(v_i)n_1}\left(\frac{2ed_i(v_i)n_1}{jm}\right)^j \tag{4.3}$$

if $k = 0$ and $2ed_i(v_i)n_1 < jm$. Furthermore, for any $\ell \geq 1$, the probability, conditioning on H_{i-1}, of the event $Bad_{j,k}^{(i)} \cap \{L_i \geq \ell\}$ is at most

$$p_{j,k}^{(i)}\left(\frac{2d_i(v_i)X_{i-1}}{\ell m}\right), \tag{4.4}$$

where $p_{j,k}^{(i)}$ (depends on j, k, and H_{i-1} via $d_i(v_i)$) is equal to the minimum of 1 *and* (4.1), (4.2), *or* (4.3) *if their respective assumptions are satisfied.*

Proof. Fix $i \geq 1$ where $|E(H_{i-1})| \leq \frac{m}{10}$ and $X_{i-1} < 10^{-4}m$ and denote $\text{Bad}_{j,k}^{(i)}$. For every $j', k' \geq 0$ we define $\mathcal{F}_{j',k'}$ to be the event that $X_i \leq X_{i-1} < 10^{-4}m$, $E(H_{i-1}) \leq \frac{m}{10}$, $J_i = j'$, $K_i = k'$.

For every (j',k') with $j' \geq j, k' \geq k$, for each (p,q) with $0 \leq p \leq j', 0 \leq q \leq k'$, and $p+q \leq \frac{\sqrt{m}}{1000}$, we let $\mathcal{A}_{p,q} = \mathcal{A}_{p,q}^{(j',k')}$ be the event that $V(H_i) \setminus V(H_{i-1})$ (i.e. the neighbours of v_i added to H during iteration i) has total degree at most $20(p+q)\sqrt{m} + 2d_i(v_i)$, that

$J_i = j' - p$, and that $K_i = k' - q$. We note that $X_i \le X_{i-1}$ implies $V(H_i) \setminus V(H_{i-1})$ has total degree at most $2d_i(v_i)$, so $\mathcal{F}_{j',k'} \subseteq \mathcal{A}_{0,0}^{(j',k')}$.

For every graph in $\mathcal{A}_{p,q}$, we say an edge is a *valid switch edge* if the following hold: it has one endpoint of degree at most $10\sqrt{m}$, and both of its endpoints are not neighbours of v_i, have degree at least three, and are not in $V(H_{i-1})$. For a graph for which $\mathcal{A}_{p,q}$ occurs, there are at most:

- $\frac{m}{50}$ edges both of whose endpoints have degree exceeding $10\sqrt{m}$;
- $20(p+q)\sqrt{m} + 2d_i(v_i) \le \frac{m}{25}$ edges out of neighbours of v_i outside of $V(H_{i-1})$;
- $n_1 + 2n_2 \le 10^{-5}m$ edges incident to vertices of degree less than 3; and
- $\frac{m}{10^5} + 10^{-4}m < \frac{m}{500}$ edges incident to $V(H_{i-1})$.

Thus there are at least $\frac{7m}{8}$ valid switch edges in any such G. Also note that

$$2m = n_1 + 2n_2 + 3(n - n_1 - n_2) = 3n - 2n_1 - n_2 > 3n - 3\epsilon m \implies n < \frac{(2+3\epsilon)m}{3}.$$

For every $(j', k'), (p, q)$ as specified, with $0 \le p \le j' - 1$, we consider switchings from $\mathcal{A}_{p,q}$ to $\mathcal{A}_{p+1,q}$ using switchings swapping $\{v_i u, ab\}$ with $\{v_i b, au\}$, where $d(u) = 1$ and ab is a valid switch edge, oriented so that For any $G \in \mathcal{A}_{p,q}$, there are at least $(j'-p)\frac{7m}{8}$ such switchings. Now, there are no more than $d_i(v_i)n_1$ choices of $\{v_i b, au\}$ in $\mathcal{A}_{p+1,q}$. Hence

$$\frac{|\mathcal{A}_{p,q}|}{|\mathcal{A}_{p+1,q}|} \le \frac{2d_i(v_i)n_1}{(j'-p)m}.$$

So $\mathbf{P}\{\mathcal{A}_{0,q}\} \le \frac{1}{j'!}(\frac{2d_i(v_i)n_1}{m})^{j'} \le (\frac{2ed_i(v_i)}{j'm})^{j'}$. Very similarly, for every $(j', k'), (p, q)$ as specified, with $0 \le q \le k' - 1$, we consider switchings from $\mathcal{A}_{p,q}$ to $\mathcal{A}_{p,q+1}$ using switchings swapping $\{v_i u, ab\}$ with $\{v_i b, au\}$, where $d(u) = 2$ and ab is a valid switch edge such that neither of a, b is the other neighbour of u, oriented so that $d(b) \le 10\sqrt{m}$. Since less than $n < \frac{(2+3\epsilon)m}{3}$ edges are incident to the other neighbour of u, we get at least $(k'-q)(\frac{7m}{8} - n) > (k'-q)\frac{m}{5}$ switchings from $\mathcal{A}_{p,q}$ and at most $d_i(v_i)2n_2$ switchings from $\mathcal{A}_{p,q+1}$. Thus

$$\frac{|\mathcal{A}_{p,q}|}{|\mathcal{A}_{p,q+1}|} \le \frac{10d_i(v_i)n_2}{(k'-q)m}.$$

So $\mathbf{P}\{\mathcal{A}_{p,0}\} \le \frac{1}{k'!}(\frac{10d_i(v_i)n_1}{m})^{k'} \le (\frac{10ed_i(v_i)}{k'm})^{k'}$. Thus, all conditioning on H_i,

$$\mathbf{P}\left\{\mathcal{F}_{j',k'}\right\} \le \mathbf{P}\left\{\mathcal{A}_{0,0}^{(j',k')}\right\} \le \begin{cases} \left(\frac{2ed_i(v_i)n_1}{j'm}\right)^{j'} \left(\frac{10ed_i(v_i)n_2}{k'm}\right)^{k'}, & j', k' > 0; \\ \left(\frac{10ed_i(v_i)n_2}{k'm}\right)^{k'}, & j' = 0; \\ \left(\frac{2ed_i(v_i)n_1}{j'm}\right)^{j'}, & k' = 0. \end{cases}$$

Union bounding over all $j' \ge j$ then over all $k' \ge k$, we get (when the fractions in parentheses are strictly less than one, as in the lemma hypotheses)

$$\mathbf{P}\left\{\mathrm{Bad}_{j,k}^{(i)}\right\} \le \begin{cases} \frac{jm}{jm-2ed_i(v_i)n_1} \left(\frac{2ed_i(v_i)n_1}{jm}\right)^{j} \frac{km}{km-10ed_i(v_i)n_2} \left(\frac{10ed_i(v_i)n_2}{km}\right)^{k}, & j, k > 0; \\ \frac{km}{km-10ed_i(v_i)n_2} \left(\frac{10ed_i(v_i)n_2}{km}\right)^{k}, & j = 0; \\ \frac{jm}{jm-2ed_i(v_i)n_1} \left(\frac{2ed_i(v_i)n_1}{jm}\right)^{j}, & k = 0. \end{cases}$$

Finally, for any $\ell \ge 1$, we consider swaps from graphs in $\mathcal{A}_{j',k'} \cap \{L_i = \ell\}$ into graphs in $\{L_i = \ell - 1\}$, conditioning on H_i. We can use any pair consisting of one of the ℓ direct back edges and one of the at least $\frac{m}{2}$ valid switch edges to switch from $\mathrm{Bad}_{j,k}^{(i)} \cap \{L_i = \ell\}$, and have at most $d_i(v_i)X_i$ switches from $\{L_i = \ell - 1\}$. Thus if $L_i \ge 1$ then we get an additional factor of $\frac{2d_i(v_i)X_i}{\ell m}$, as claimed. $\square$

Corollary 4.2. *For all $0 < \gamma < 1$ and $0 < \epsilon < 10^{-8}$, if m is large enough and $n_1 \leq m^{1-\gamma}$ while $n_2 \leq \epsilon m$ and $2|E(H_i)| + X_i \leq \epsilon m$, then*

$$\mathbf{P}\left\{ X_i < X_{i-1} - 12\sqrt{X_{i-1}} \;\middle|\; H_{i-1} \right\} \leq \left(\frac{d_i(v_i) n_1}{10\sqrt{X_{i-1}} m} \right)^{10\sqrt{X_{i-1}}}.$$

Proof. We have $X_i - X_{i-1} \geq d(v_i) - 2J_i - K_i - 3L_i \geq -J_i - 2L_i$ and $L_i \leq \sqrt{X_{i-1}}$. So if $X_i < X_{i-1} - 12\sqrt{X_i}$ then $J_i \geq 10\sqrt{|X_i|}$. Lemma 4.1 implies that the probability this occurs is at most $(\frac{d_i(v_i)n_1}{10\sqrt{X_i}m})^{10\sqrt{X_i}}$. □

Techniques similar to those used in the proof of Lemma 4.1 can be pushed further when we are considering edges between low degree vertices. Specifically, defining

$$\delta^* = \min\left(10^6, \max\left(k \geq 1 : \sum_{j=1}^{k} n_j \leq \frac{n}{10^{24}}\right)\right), \tag{4.5}$$

they allow us to prove the following lemma.

Lemma 4.3. *Fix $0 < \epsilon < \frac{1}{100}$. Assume H_{i-1} is such that $X_{i-1} + 2|E(H_{i-1})| \leq 10^6$. Then for any (δ^*+1)-tuple $(s_1, \ldots, s_{\delta^*+1})$, the probability conditioned on H_{i-1} that v_i has s_j neighbours outside H_{i-1} of degree j for each $1 \leq j \leq \delta^* + 1$ is $O(\prod_{j=1}^{\delta^*+1} (\frac{n_j}{m})^{s_j})$. If $d_i(v_i) \leq \delta^* + 1$ then the probability that this occurs and $L_i = \ell$ is $O(\prod_{j=1}^{\delta^*+1} (\frac{n_j}{m})^{s_j} (\frac{1}{m})^{\ell})$.*

Proof. We note that our bound on X_i implies that the probability we are bounding is 0 unless the sum of the s_j is at most 10^6. So, if $n_{\delta^*+1} > \frac{m}{10^{18}}$ then $(\frac{n_{\delta^*+1}}{m})^{s_{\delta^*+1}} = \Omega(1)$. So, in this case, the lemma is equivalent to the statement obtained from it by replacing the upper bound on j by δ^*. We let δ' be δ^* in this case, and otherwise we set $\delta' = \delta^* + 1$. So we need only prove the statement obtained from the lemma by replacing the upper bound on j by δ'.

Let B be the set of vertices of degree at least $\delta' + 1$. By our choice of δ^*, there are at most $\frac{n}{10^{12}} < \frac{m}{10^{12}}$ edges incident to vertices of degree at most δ^*, in particular when $\delta' = \delta^*$. By our choice of δ', if $\delta' = \delta^* + 1$ then there are at most $\frac{m}{10^{12}}$ edges incident to vertices of degree δ'. So the sum of the degrees of the vertices in B is at least $\frac{199m}{100}$. For any set S of at most δ' vertices and large m, the sum of the degrees of the vertices in S is at most $\binom{\delta'}{2} + m < \frac{26m}{25}$. Hence the sum of the degrees of the vertices of $B \setminus S$ is at least $\frac{4m}{5}$. Every vertex in $B \setminus S$ has degree exceeding δ' and hence at least one neighbour outside of S, so the number of edges which have one endpoint in $B \setminus S$ and the other in $V \setminus S$ is at least $|B \setminus S| \geq \frac{4m}{5(\delta'+1)} > \frac{m}{10^7}$. By our upper bound $\frac{m}{10^{12}}$ on the number of edges incident to $V \setminus B$ we see there are $\Omega(m)$ edges within $B \setminus S$.

Fix H_{i-1} as in the statement of the lemma. For any $(\delta'+1)$-tuple $(\ell_1, \ldots, \ell_{\delta'}, \ell)$ where $0 \leq \ell_j \leq s_j$ for each $1 \leq j \leq \delta'$ and $0 \leq \ell \leq d_i(v_i)$, let $\mathcal{F}_{(\ell_1,\ldots,\ell_{\delta'},\ell)}$ denote the event that H_{i-1} is explored and v_i has ℓ_j neighbours outside H_{i-1} of degree j for each $1 \leq j \leq \delta'$, and $L_i = \ell$.

We consider first switchings which reduce the value of some $\ell_j \geq 1$ for some $j \leq \delta'$. We consider an edge $v_i w$ where w is outside H_{i-1} and has degree j. We set $S = N(w) - v$ and note $|S| \leq \delta' - 1$. As noted above, there are $\Omega(m)$ edges within $B \setminus S$. At most 10^{12} have both their endpoints in $N(v_i)$ and at most $2(10)^{12}$ have an endpoint in $V(H_{i-1})$. So there are $\Omega(m)$ valid switches using $v_i w$ and an xy where $x \in B \setminus (S \cup V(H_{i-1}))$ and $y \in B \setminus (S \cup N(v_i) \cup V(H_{i-1}))$. Since $d_i(v_i), d(w) \leq 10^6$, there are $d_i(v_i) j n_j = O(n_j)$ swaps in the opposite direction. Letting $L = (\ell_1, \ldots, \ell_{\delta^*}, \ell)$ and L' denote L with ℓ_j replaced by $\ell_j - 1$, then

$$\mathbf{P}\left\{\mathcal{F}_L\right\} \leq \mathbf{P}\left\{\mathcal{F}_{L'}\right\} \frac{O(n_j)}{\Omega(m)}$$

and the first statement follows.

We turn to the second statement, so we can assume $d_i(v_i) \leq \delta^* + 1$. We consider switchings which reduce $L_i = \ell \geq 1$. We let $v_i w$ be an edge of G with $w \in V(H_{i-1})$. We let S be the neighbours of v_i outside $V(H_{i-1})$. So, $|S| \leq \delta'$. There are $\Omega(m)$ edges within $B \setminus S$. At most 10^{12} have both their endpoints in $N(w)$ and at most $2(10)^{12}$ have an endpoint in $V(H_{i-1})$. So there $\Omega(m)$ valid switches using $v_i w$ and an xy where $x \in B \setminus (S \cup V(H_i))$ and $y \in B \setminus (S \cup N(w) \cup V(H_{i-1}))$. Since $X_{i-1} \leq 2(10)^{12}$, there are $X_{i-1}^2 = O(1)$ swaps in the opposite direction. Letting $L = (\ell_1, \ldots, \ell_{\delta^*}, \ell)$ and let L' denote L with ℓ replaced by $\ell - 1$, then

$$\mathbf{P}\{\mathcal{F}_L\} \leq \mathbf{P}\{\mathcal{F}_{L'}\} \frac{O(1)}{\Omega(m)}$$

and the second statement follows. □

5. Analyzing The Later Stages of The Exploration

In this section, we consider the behaviour of the exploration after it has survived for some $C = C(m)$ iterations. By the end of the section we will have developed enough machinery to prove Theorem 1.3.

We first prove:

Lemma 5.1. *For any m large enough, $0 < \gamma < 1$, $0 < \epsilon < 10^{-10}$, $n_1 \leq m^{1-\gamma}$, $n_2 < \epsilon m$, $0 < C \leq m^{\gamma/20}$, $1 \leq i$, $0 \leq r \leq C(\log m)^2$, the following holds: if $2|E(H_{i-1})| + X_{i-1} \leq \epsilon m$ then*

$$\mathbf{P}\left\{ \sum_{t=i}^{i+r-1} J_t \mathbf{1}_{[X_t \leq X_{t-1}]} \geq \frac{C}{3} \text{ and } \forall i \leq t \leq i+r-1 : X_{t-1} \leq 2C \;\middle|\; H_{i-1} \right\} < m^{-\gamma C/4}.$$

Proof. If the event whose conditional probability we aim to bound occurs, we can choose one of a set of at most $r^{C/3}$ sequences of nonnegative integers J'_t which sum to $\frac{C}{3}$ such that for each $i \leq t \leq i+r-1$, $J'_t \leq J_t$, and $J'_t > 0$ only when $X_t \leq X_{t-1}$. We can choose such a set by repeatedly increasing some J'_t which is less than J_t. For any t with $i \leq t \leq i+r-1$ we have by the hypotheses and that fact that $d_s(v_s) < X_s$ that $2|E(H_{t-1})| + X_{t-1} \leq 2|E(H_{i-1})| + 2\sum_{s=i}^{t-1} d_s(v_s) + X_{t-1} \leq \frac{m}{10^5}$. We may thus apply the bound (4.3) from Lemma 4.1 to the positive $J'_t \leq \frac{C}{3} \leq \sqrt{m}/1000$. Doing so, and then applying our bounds on r, n_1 and C, and the fact $d_t(v_t) \leq X_{t-1} < 2C$, we obtain that

$$\begin{aligned}
&\mathbf{P}\left\{ \sum_{t=i}^{i+r-1} J_t \mathbf{1}_{[X_t \leq X_{t-1}]} \geq \frac{C}{3} \text{ and } \forall i \leq t \leq i+r-1 : X_{t-1} \leq 2C \;\middle|\; H_{i-1} \right\} \\
&\leq \sum_{\{J'_t\}_{t=i}^{i+r-1}} \prod_{t=i}^{i+r-1} \left(\frac{2e(2C)n_1}{J'_t m} \right)^{J'_t} \leq r^{C/3} \left(\frac{2e(2C)n_1}{m} \right)^{C/3} \\
&\leq \left(\frac{4e(C\log m)^2 n_1}{m} \right)^{C/3} \leq (m^{\frac{-8\gamma}{9}})^{C/3}.
\end{aligned}$$

□

Lemma 5.2. *For any m large enough, $0 < \gamma < 1$, $0 < \epsilon < 10^{-6}$, $n_1 < \epsilon m$, $n_2 \leq 10^{-6} m$, $0 < C \leq m^{\gamma/20}$, $i \geq 1$, and $0 \leq r \leq C(\log m)^2$, the following holds: if $2|E(H_{i-1})| + X_{i-1} \leq \epsilon m$ then, conditioned on H_{i-1},*

$$\mathbf{P}\left\{ |\{i \leq t \leq i+r-1 : K_t = d_t(v_t) > 0\}| > \frac{r}{2} \text{ and } \forall i \leq t \leq i+r-1 : X_{t-1} \leq 2C \right\} \leq 14^{-r}.$$

Proof. We have $C > 0$ by hypothesis and we can assume $n_2 > 0$ as otherwise every $K_t = 0$. For any t with $i \le t \le i + r - 1$ we have by the hypotheses that $2|E(H_{t-1})| + X_{t-1} \le 2|E(H_{i-1})| + X_{t-1} + \sum_{s=i-1}^{t-1} d_s(v_s) \le \frac{m}{10^5}$. The bound (4.2) from Lemma 4.1 and our hypotheses imply that the probability that $0 < K_t = d_t(v_t) \le 2C < \sqrt{m}/1000$ conditioned on H_{t-1} is at most $\frac{1}{800}$. So on the event that $2|E(H_{i-1})| + X_{i-1} \le \epsilon m$ and that for all $i \le t \le i + r$, $X_{t-1} \le 2C$, the size of the set $\{i \le t \le i + r - 1 : K_t = d_t(v_t) > 0\}$ is stochastically dominated by a Binomial random subset of a set of size r, and thus the expected number of subsets of $\{i, \dots, i + r - 1\}$ of size at least $\frac{r}{2}$ such that for each t in the set $K_t = d_t(v_t) > 0$ is at most $2^r 800^{-r/2} < 14^{-r}$. □

Lemma 5.3. *For any m large enough, $0 < \gamma < 1$, $0 < \epsilon < 10^{-6}$, $n_1, n_2 < \epsilon m$, $C \le m^{\gamma/20}$, $1 \le i$, $r \le C(\log m)^2$ the following holds, if $2|E(H_{i-1})| + X_{i-1} \le \epsilon m$ then*

$$\mathbf{P}\left\{ \sum_{t=i}^{i+r-1} L_t \mathbf{1}_{[X_t \le X_{t-1}]} \ge \frac{C}{15} \text{ and } \forall i \le t \le i + r - 1 : X_{t-1} \le 2C \;\middle|\; H_{i-1} \right\} \le m^{-\sqrt{C}/25}.$$

Proof. Note $L_t < \sqrt{X_{t-1}} \le \sqrt{2C}$ deterministically. Thus

$$\sum_{t=i}^{i+r-1} L_t \mathbf{1}_{[X_t \le X_{t-1}]} \le \sqrt{2C} \sum_{t=i}^{i+r-1} \mathbf{1}_{[X_t \le X_{t-1},\ L_t > 0]}.$$

and we need only bound the probability that $\sum_{t=i}^{i+r-1} \mathbf{1}_{[X_t \le X_{t-1},\ L_t > 0]} \le \frac{\sqrt{C}}{15\sqrt{2}}$.

Since $|E(H_{t-1})| \le |E(H_{i-1})| + \sum_{s=i-1}^{t-1} d_s(v_s)$ the hypotheses imply that for all t with $i \le t \le i + r - 1$, $|E(H_{t-1})| + X_{t-1} \le \frac{m}{10^5}$. So, by the bound (4.4) from Lemma 4.1 with $p_{j,k}^{(i)} \le 1$, since $d_t(v_t), X_{t-1} \le 2C$,

$$\mathbf{P}\left\{ \sum_{t=i}^{i+r-1} L_t \mathbf{1}_{[X_t \le X_{t-1}]} \ge \frac{C}{15} \text{ and } \forall i \le t \le i + r - 1 : X_{t-1} \le 2C \;\middle|\; H_{i-1} \right\}$$
$$< r^{\frac{\sqrt{C}}{15\sqrt{2}}} \left(\frac{8C^2}{m}\right)^{\frac{\sqrt{C}}{15\sqrt{2}}} \le \left(\frac{8C^3(\log m)^2}{m}\right)^{\frac{\sqrt{C}}{15\sqrt{2}}} < m^{-\sqrt{C}/25}$$ □

Corollary 5.4. *For any m large enough, $0 < \epsilon < 10^{-8}$, $0 < \gamma < 1$, if $n_1 \le m^{1-\gamma}, n_2 \le 10^{-6} m$, and $3 \le C \le m^{\gamma/20}$, then the probability that for a specific i, $|E(H_{i-1})| + X_{i-1} \le \epsilon m$ and $0 < X_t < 2C$ for all $i \le t < i + C(\log m)^2$ is at most $m^{-\gamma\sqrt{C}/26}$.*

Proof. Recall that $X_s - X_{s-1} \ge d_s(v_s) - 2J_s - K_s - 3L_s$. So, if $d_s(v_s) \ne K_s$ then either $J_s > 0$, $L_s > 0$, or $X_s > X_{s-1}$. Hence letting $r = \lfloor C(\log m)^2 \rfloor$ and Y be the number of s with $i \le s \le i + r - 1$ for which $K_s \ne d_s(v_s)$, we have that the number of open edges increases by at least $Y - \sum_{s=i}^{i+r-1} J_s \mathbf{1}_{[X_s \le X_{s-1}]} - \sum_{s=i}^{i+r-1} L_s \mathbf{1}_{[X_s \le X_{s-1}]}$ of the iterations between i and $i + r - 1$. Furthermore, the total decrease over those of these r iterations in which the number of open edges decreases is at most $\sum_{s=i}^{i+r-1} J_s \mathbf{1}_{[X_s \le X_{s-1}]} + \sum_{s=i}^{i+r-1} 2L_s \mathbf{1}_{[X_s \le X_{s-1}]}$. So, $X_{i+r} \ge X_i + Y - 2\sum_{s=i}^{i+r-1} J_s \mathbf{1}_{[X_s \le X_{s-1}]} - 3\sum_{s=i}^{i+r-1} L_s \mathbf{1}_{[X_s \le X_{s-1}]}$.

Let $\mathcal{A}$ denote the event that $0 < d_t(v_t) < 2C$ for all $i \le t \le i + r$. So we need to show $\mathbf{P}\{\mathcal{A}\} \le m^{-\gamma\sqrt{C}/26}$. We note that if $\mathcal{A}$ holds then for every t with $i \le t \le i + r - 1$, $|E(H_t)| + X_t \le \frac{m}{10^6}$. Applying Lemma 5.2, the probability $\mathcal{A}$ holds and $Y < 3C < \frac{r}{2}$ is at most $14^{-r} < m^{-C}$. Applying Lemma 5.1, the probability $\mathcal{A}$ holds and $\sum_{s=i}^{i+r-1} J_s \mathbf{1}_{[X_s \le X_{s-1}]} > \frac{C}{3}$ is at most $m^{-\gamma C/4}$, Finally, by Lemma 5.3, the probability $\mathcal{A}$ holds and $\sum_{s=i}^{i+r-1} L_s \mathbf{1}_{[X_s \le X_{s-1}]} > \frac{C}{15}$ is $m^{-\sqrt{C}/25}$. But if $\mathcal{A}$ holds while $Y \ge 3C$, $\sum_{s=i}^{i+r-1} J_s \le \frac{C}{3}$, and $\sum_{s=i}^{i+r-1} L_s \mathbf{1}_{[X_s \le X_{r-1}]} \le \frac{C}{15}$ then $X_r \ge 3C - \frac{2C}{3} - \frac{C}{5} > 2C$, contradicting the fact $\mathcal{A}$ holds; so the probability of this event is 0. □

Corollary 5.5. *For any m large enough, if $0 < \gamma < 1$, $0 < \epsilon < 10^{-8}$, $n_1 \le m^{1-\gamma}$, $n_2 \le 10^{-6}m$, and $3 \le C \le m^{\gamma/20}$, then the probability given that for a specific $i \ge 1$, $X_{i-1} \ge C$ and there is no $t > i$ such that either $X_t \ge 2C$ or $2|E(H_{t-1})| + X_{t-1} \ge \epsilon m$ is $2m^{-\gamma\sqrt{C}/26}$.*

Proof. Fix iteration $i \ge 1$ such that $X_{i-1} \ge C$ and condition all probabilities on H_{i-1}. For $r = \lceil C(\log m)^2 \rceil$, we let $\mathcal{A}$ be the event that for all t with $i \le t \le i+r-1$, $X_t < 2C$ and $2|E(H_{t-1})| + X_{t-1} \le \epsilon m$. So it suffices to show $\mathbf{P}\{\mathcal{A} \mid H_{i-1}\} \le m^{-\gamma\sqrt{C}/20}$. We note that since $2|E(H_{i-1})| + X_{i-1} < \epsilon m$, the assumptions of Lemmas 5.1, 5.2, 5.3 and Corollary 5.4 are satisfied.

We first compute the probability of the intersection of $\mathcal{A}$ and the event $\{X_{i+r-1} \ne 0\}$. This implies $X_t > 0$ for all $i \le t \le i+r-1$. Applying Corollary 5.4, we have that the conditional probability of $\mathcal{A} \cap \{X_r > 0\}$ is at most $m^{-\gamma\sqrt{C}/26}$.

We next compute the probability of the intersection of $\mathcal{A}$ and $\{X_r = 0\}$. The latter event is contained in the union of $\{\sum_{s=i}^{i+r-1} J_s \mathbf{1}_{[X_s \le X_{s-1}]} > \frac{C}{2}\}$ and $\{\sum_{s=i}^{i+r-1} 2L_s \mathbf{1}_{[X_t \le X_{t-1}]} > \frac{C}{2}\}$. Applying Lemma 5.1, the (conditional) probability of $\mathcal{A} \cap \{\sum_{s=i}^{i+r-1} J_s \mathbf{1}_{[X_s \le X_{s-1}]} > \frac{C}{2}\} < m^{\gamma C/4}$. Applying Lemma 5.3, the (conditional) probability of the event $\mathcal{A} \cap \{\sum_{s=i}^{i+r-1} 2L_s \mathbf{1}_{[X_t \le X_{t-1}]} > \frac{C}{2}\}$ is $m^{-\sqrt{C}/22}$. So the (conditional) probability of $\mathcal{A} \cap \{X_r = 0\}$ is also at most $m^{-\gamma\sqrt{C}/26}$.

□

Corollary 5.6. *For any $\gamma > 0$ and sufficiently small $\epsilon > 0$, if $n_1 \le m^{1-\gamma}, n_2 \le 10^{-6}m$ then the probability that $\sum_{t=1}^{i} d_t(v_t) \ge 4(\log m)^4$ for some i, and v does not lie in a component with at least $\frac{\epsilon m}{2}$ edges is $o(m^{-(\log m)^{1/3}})$.*

Proof. We consider the largest $t \le \lfloor \frac{\gamma \log m}{20} \rfloor$ such that $X_i \ge 2^t$ for some i, and the largest i for which this holds for our choice of t. If $t \le \lfloor \log\log m \rfloor$ then there is no i such that $X_i > 2\log m$ and hence no i such that $d_i(v_i) > 2\log m$. In this case, in order for $\sum_{t=1}^{i} d_t(v_t) \ge 4(\log m)^4$ to occur for some i, in particular X_t must remain above 0 for the first $\lceil 2(\log m)^3 \rceil$ iterations without reaching $2\log m$. Applying Corollary 5.4 with $i = 1$ (note that the role of i in that corollary is not the same as the role of i in this one) and $C = 2\log m$, we see that the probability this occurs is $o(m^{-(\log m)^{1/3}})$. Applying Corollary 5.5 for $\lfloor \log\log m \rfloor < t < \lfloor \frac{\gamma \log m}{20} \rfloor$ times, we see that the probability we have $\lfloor \log\log m \rfloor < t < \lfloor \frac{\gamma \log m}{20} \rfloor$ is $o(m^{-(\log m)^{1/3}})$.

If $t = \lfloor \frac{\gamma \log m}{20} \rfloor$ then applying Corollary 5.5, to the first i for which $X_i \ge 2^t$, we see that the probability we have no $i' > i$ for which either $X_{i'} \ge 2^{t+1}$ or $2|E(H_{i'-1})| + X_{i'-1} \ge \epsilon m$ is $o(m^{-(\log m)^{1/3}})$. We further apply Corollary 5.5, to every i for which $2^t \le X_i < 2^{t+1}$ and $X_{i-1} \ge 2^{t+1}$, combined with the union bound, this yields that the probability that there is such an i for which there is no $i' > i$ with $X_{i'} \ge 2^{t+1}$ or $2|E(H_{i'-1})| + X_{i'-1} \ge \epsilon m$ is $o(m^{-(\log m)^{1/3}})$. In particular, this implies we need only bound the probability of the event that there is an i such that $X_i \le 2^t$ and $X_{i-1} \ge 2^{t+1}$. Since this means $X_i < X_{i-1} - 12\sqrt{X_{i-1}}$, applying Corollary 4.2, we are done. □

We can now give the proof of Theorem 1.3.

Proof of Theorem 1.3. We simply combine n applications of Corollary 5.6 with one application of Lemma 3.1. □

6. Analyzing The Start of The Exploration

We now prove Theorem 1.1.

We recall that the theorem states that for any feasible degree sequence the probability that $G(\mathcal{D})$ is disconnected is

$$O\left(u_{edge} + u_{\Delta} + u_{\Delta+1} + u_{K_4-e} + u_{K_4} + u_{K_5^+}\right).$$

where

$$u_{edge} = \frac{\max(n_1-1,0)^2}{m},\ u_{\Delta} = \frac{\max(n_2-2,0)^3}{m^3},\ u_{\Delta+1} = \frac{n_1\max(n_2-1,0)^2 n_3}{m^4},$$
$$u_{K_4-e} = \frac{\max(n_2-1,0)^2\max(n_3-1,0)^2}{m^5},\ u_{K_4} = \frac{\max(n_3-3,0)^4}{m^6},\ u_{K_5^+} = \frac{n}{m^6}.$$

Proof of Theorem 1.1. The theorem is trivially true unless $n_1 < 10^{-6}\sqrt{m}$ and $n_2 \le 10^{-6}m$ so we assume this is the case. We prove the theorem by exploring out of every vertex $v \in [n]$. By Theorem 1.3, the probability that the component containing v has at least $2(\log m)^4$ edges but $G(\mathcal{D})$ is not connected, is $o(m^{\log\log m})$. Therefore it suffices to bound, for every $v \in [n]$, the probability that the exploration from v dies out before reaching a step i with $2|E(H_i)|+X_i \ge 4(\log m)^4$. We refer to the period before $2|E(H_i)|+X_i \ge 4(\log m)^4$ as the early stage and note that it consists of at most $2(\log m)^4$ iterations. For any vertex v which dies out in the early stage, we define i^* as the minimum of the i for which $X_i = 0$, so $i^* \le 2(\log m)^4$. Our precise results on the early stage of the exploration depend on the shape of the degree distribution and on the degree of v.

We first compute the probability that a vertex v of degree 1 is the endpoint of a path component of length i^*+1 by considering the exploration from it. For this to occur, we must have $d_i(v_i) = X_{i-1} = 1 = K_i$ for all $1 \le i < i^*$ (if $i^* = 1$ then there are no such i) and $d_{i^*}(v_{i^*}) = X_{i^*-1} = J_{i^*} = 1$. Applying Lemma 4.1, we obtain that the probability this occurs is at most

$$\frac{10\max(n_1-1,0)}{m}\left(\frac{20n_2}{m}\right)^{i^*}.$$

Hence, summing over all vertices of degree 1, the probability a path component is created in the early stages is at most

$$n_1\sum_{0\le i^*\le 2(\log m)^2}\frac{10\max(n_1-1,0)10^{-4i^*}}{m} \le \frac{20\max(n_1-1,0)^2}{(1-10^{-4})m} < \frac{21max(n_1-1,0)^2}{m}.$$

We next compute the probability that a vertex v of degree 2 lies on a cycle of length i^*+1 by considering the exploration from it. If this occurs then $i^* \ge 2$, $X_0 = d_1(v_1) = K_1 = 2$, $d_i(v_i) = K_i = 1$ and $X_i = 2$ for all $2 \le i < i^*$ (if $i^* = 2$ then there are no such i), and $d_{i^*}(v_{i^*}) = L_{i^*} = 1$ while $X_{i^*-1} = 2$. Applying Lemma 4.1, we obtain that the probability this occurs is no more than

$$\left(\frac{30\max(n_2-2,0)}{m}\right)^2\left(\frac{30\max(n_2-2,0)}{m}\right)^{i^*-2}\frac{4}{m} \le \frac{3600\max(n_2-2,0)^2(10^{-4i^*+8})}{m^3}.$$

So the probability there is a vertex in such a cycle component is at most

$$n_2\sum_{j\ge 0}\frac{3600\max(n_2-2)^2(10^{-4j})}{m^3} \le \frac{10800\max(n_2-2,0)^3}{(1-10^{-4})m^3} < \frac{11000\max(n_2-2,0)^3}{m^3}.$$

Finally, we consider explorations from v of degree at least 3 which die out in the early stages. We note that for any iteration i, if $L_i \ge 2$ then $X_i \ge L_i(d_i(v_i)-1) > 0$, and if $L_i = 1$ then $X_i \ge d_i(v_i)-1 > 0$ unless $d_i(v_i) = 1$. So either $L_{i^*} = 0$ (and $J_{i^*} = X_{i^*-1} = d_{i^*}(v_{i^*})$), or $L_{i^*} = d_{i^*}(v_{i^*}) = 1$ and $X_{i^*-1} = 2$ (and $J_{i^*} = K_{i^*} = 0$).

It is straightforward to show that if, for some i, $X_i > 4000$ but the exploration from v dies out in the early stages, then either

(i) there is a $j > i$ for which $X_{j-1} > 500$ and $X_j < X_{j-1} - 12\sqrt{X_{j-1}}$, or

(ii) there are at least 13 $j > i$ for which $X_j < X_{j-1}$, or
(iii) there are at least 7 $j > i$ for which $X_j < X_{j-1} - 1$.

Furthermore if $X_i < X_{i-1}$ at most 12 times, no X_i exceeds 4000, and (i) does not hold, then there are at most 10000 i for which $X_i > X_{i-1}$. So to deal with explorations from vertices of degree at least three it is enough to show that

(A) For every such vertex v, the probability that $X_i < X_{i-1}$ for more than 12 iterations $i \leq i^*$ or $X_i < X_{i-1} - 1$ for more than 6 iterations $i \leq i^*$ is $O(m^{-6})$ (cases (ii) (iii));
(B) For every such vertex v, the probability that for some $i < i^*$ with $X_i > 500$, $X_{i-1} < X_i - 12\sqrt{X_{i-1}}$ is $O(m^{-6})$ (case (i)); and
(C) The probability that there is some component such that when exploring from every vertex v in the component, no X_i exceeds 4000 and there are at most 12 iterations i for which $X_i < X_{i-1}$ and at most 6 iterations i for which $X_i < X_{i-1} - 1$, is
$$O\left(u_{edge} + u_{\Delta} + u_{\Delta+1} + u_{K_4-e} + u_{K_4} + u_{K_5^+}\right).$$

In order that $X_i \leq X_{i-1} - 1$, we must either have $J_i \geq 1$ or $L_i \geq 1$. Applying Lemma 4.1, we see that the probability that this occurs during an iteration in the early stage conditioned on the exploration to that point is $O(\frac{\max(n_1,(\log m)^8)}{m})$ which is $O(m^{-1/2})$. In the same vein, in order that $X_i \leq X_{i-1} - 2$, we must either have $J_i \geq 2$ or $L_i \geq 1$. Applying Lemma 4.1, we see that the probability that this occurs in the early stage is $O(\frac{(\log m)^8}{m})$. So, the probability that in the early stage there are more than twelve iterations in which X_i decreases or more than six in which X_i decreases by at least 2 is $o(m^{-6})$ and we are done with (A).

Next, Corollary 4.2 implies that the probability that for some iteration i in the early stage, $X_i > 500$ and $X_i < X_{i-1} - 12\sqrt{X_{i-1}}$ is $o(m^{-6})$ and we are done with (B),

We now turn to proving (C). For any $0 \leq i^* \leq \lfloor 4(\log m)^4 \rfloor$ and disjoint subsets $S^-, S^+ \subseteq S = \{1, 2, \ldots, \lfloor 4(\log m)^4 \rfloor\}$ with $|S^-| \leq 12$ and $|S^+| \leq 10000$, we can compute the probability that $X_{i^*} = 0$ but $0 < X_i < 4000$ for all $1 \leq i < i^*$, $X_i - X_{i-1} > 0$ only for $i \in S^+$ and $X_i - X_{i-1} < 0$ only for $i \in S^-$. For each of the remaining $i < i^*$ in which $X_i = X_{i-1}$ either $K_i = d_i(v_i)$ or $J_i + L_i > 0$. So, applying Lemma 4.1, the probability that $X_{i-1} = X_i$ is at most $\max(m^{-1/3}, \frac{20n_2}{m}) \leq 10^{-4}$, while the number of choices for S^+ and S^- and L_i, J_i, K_i for each $i \in S^- \cup \{1, \ldots, 300\}$ is less than $(i^*)^{10012}(4000)^{936}$. So the probability that such an S^- and S^+ exist is of no larger order than the maximum, over all such choices of $\{L_i, J_i, K_i\}_{i \in S^- \cup \{1,\ldots,300\}}$, of the product of the conditional probabilities given H_{i-1} that J_i, K_i, L_i take the chosen values for each i in $\{1, \ldots, 300\} \cup S^-$.

We note that if $|S^-| = 1$ and the events of the preceding paragraph occur, then $X_{i^*-1} - X_{i^*} \geq X_0 - 0 \geq 3$ so we must have $L_{i^*} = 0$ and $J_{i^*} = d_{i^*}(v_{i^*}) \geq 3$.

If $n_1 \geq 2$ then $u_{edge} = \frac{\max(n_1-1,0)^2}{m} \geq \frac{n_1^2}{2m}$, and Lemma 4.1 then implies that the probability that $|S^-| \geq 2$ or $J_{i^*} \geq 3$ is $O(\frac{n_1^2}{m^2})$; so (C) holds in this case.

We therefore only need to prove (C) when $n_1 \leq 1$ (recalling the definition of δ^* from (4.5), we note that this implies $\delta^* + 1 \geq 2$ which will be relevant when we apply Lemma 4.3). Applying Lemma 4.1 we see that this implies that the probability $X_i < X_{i-1}$ is $O(\frac{1}{m})$ so (C) holds for explorations for which $|S^-| > 5$. Therefore we need only consider explorations with $|S^-| \leq 5$. Furthermore, either $L_{i^*} = 1$ and $J_{i^*} = 0$ or $J_{i^*} = d_{i^*}(v_{i^*}) = 1$ and $L_{i^*} = 0$. In either case, $X_{i^*-1} \leq 2$ and so $|S^-| \geq 2$. Hence, if $n_2 = \Omega(n)$ then we are done by considering u_{Δ}, so we can assume this is not the case and so $\delta^* + 1 \geq 3$. Furthermore, we are done if $n_2 > 2$ and $|S^-| > 3$ so we can assume this is not the case.

We now let $i_1 > i_2 \cdots > i_{|S^-|}$ be the elements of S^-, so $i_1 = i^*$. If $X_{i_2-1} > 6$ then $2L_{i_2} + J_{i_2} \geq 5$. Thus, we have $d_{i_2}(v_{i_2}) \geq 3$ and since $J_{i_2} \leq n_1 \leq 1$, $L_{i_2} \geq 2$, but since $X_{i_2} \leq X_{i^*-1} \leq 2$, this contradicts our choice of v_{i_2}. So, $X_{i_2-1} \leq 6$. A similar proof then shows $X_{i_3-1} \leq 12$. Continuing in the same vein, we obtain that $X_{i_4-1} \leq 20$ and

$X_{i_5-1} \leq 30$. Thus we need only consider explorations where every $X_i \leq 30$ and hence every vertex explored has degree at most 31.

Now, by Lemma 4.1, since $n_1 \leq 1$, the (conditional) probability that $0 < X_i \leq X_{i-1}$ for $i \leq 300$ is $O(\frac{\max(n_2,1)}{m})$. It follows that the probability that there are 5 such i and $0 = X_{i^*} < X_{i^*-1}$ is $O((\frac{\max(n_2,1)}{m})^6)$ $=O(\frac{\max(u_\Delta, u_{K_5^+})}{n})$ and (C) holds in this case. But if $i^* > 500$ then either there are 5 such i or some X_i exceeds 30, and (C) holds in both cases. So, we need only consider explorations for which $i^* \leq 500$. Hence the number of choices for the degrees of the vertices added to the graph and the L_i edges added within the graph at every iteration is a fixed constant and we need only bound the maximum, over any such choice, of the products of the conditional probability that in each iteration i, the degrees of the vertices added to the tree and back edges are as claimed.We refer to such a choice of the number of vertices of every degree up to 31 and back edges for each iteration up to 500 as the *specified choice.*

We note that since $d_i(v_i) \leq X_{i-1} < 30$ in every iteration, we have $2|E(H_{i-1})| + X_i < 10^6$ in every iteration.

Claim 6.1. *The conditional probability, given H_{i-1}, of the specified choice for a fixed iteration $1 \leq i \leq 500$ occurring, is $O\left(m^{-\left\lceil\frac{\min(X_{i-1},11)-X_i}{2}\right\rceil}\right)$.*

Proof. Recall that $J_i + 2L_i \geq X_{i-1} - X_i$. So, applying Lemma 4.1 proves the claim unless $L_i \geq 2$. Furthermore, since $\delta^* + 1 \geq 3$, applying Lemma 4.3 proves the claim unless $d_i(v_i) \geq 4$. If $L_i > 2$ this implies $X_i \geq L_i(d_i(v_i) - 1) \geq 9$ and again we are done by applying Lemma 4.1.

So, $L_i = 2$, $X_{i-1} \geq (L_i + 1)d_i(v_i) \geq 12$ and $11 - X_i \leq 4$. So we are done by Lemma 4.1 unless $J_1 = 0$ and $X_i \leq 8$. This implies $d_i(v_i) = 4, L_i = 2,\ K_i = 2$ and $X_i = 8$. So, applying Lemma 4.1 proves the claim unless $n_2 \geq 3$. This implies $|S^-| \leq 3$ which is impossible as $X_i > 6$. □

Corollary 6.2. *For any fixed iteration $1 \leq i \leq 500$, the conditional probability, given H_{i-1}, of the specified choice for every iteration after i occurring, is $O\left(m^{-\left\lceil\frac{\min(X_{i-1},11)}{2}\right\rceil}\right)$.*

This corollary implies that the probability that there is any exploration dying out in the first 500 iterations for which some $X_i \geq 11$ is $O(\frac{n}{m^6}) = O(u_{K_5^+})$, so we need only count choices for which every $X_i \leq 10$.

Claim 6.3. *Claim (C) holds if we weaken its statement by requiring the component to contain a vertex of degree at least* 4.

Proof. We consider exploring from a vertex v of degree at least 4.

We first consider the case $n_2 \geq 3$. We know that $X_1 \geq 2d(v) - 2J_1 - K_1$ or equivalently that $\frac{X_1}{2} - J_1 \leq -d(v) + \frac{K_1}{2}$. Applying Lemma 4.3 we know the probability that the first iteration behaves as claimed is $O(\frac{\max(n_2-2,1)^{K_1}}{m^{J_1+K_1}})$. So, applying Corollary 6.2 and the fact that $K_1 \leq d(v)$ we have that the probability the specified exploration occurs is

$$O(\max(n_2 - 2, 1)^{K_1} m^{-\frac{X_1}{2} - J_1 - K_1}) = O(\max(n_2 - 2, 1)^{K_1} m^{-d(v) - \frac{K_1}{2}}).$$

Since $d(v) \geq max(K_1, 4)$, this is $O(\frac{u_\Delta}{m})$.

It remains to consider the case $n_2 \leq 2$, which implies $\max(n_2 - 2, 1) = O(1)$. Since $n_1 \leq 1$, if some vertex of degree at least 4 has a neighbour of degree at least 4, then one of these two vertices has no neighbour of degree 1. Thus a component containing a vertex of degree at least 4 either contains a vertex of degree at least 4 adjacent to no vertex of degree at least 4, or a vertex of degree at least 4 adjacent to no vertex of degree 1.

We consider first the probability of a specified exploration from v in which v has no neighbour of degree at least 4, and hence has a neighbour of degree three. We can assume $X_1 \leq 10$ so applying Lemma 4.3 with $\delta^* + 1 = 3$ to the first iteration and Corollary 6.2 to the remainder of the exploration, we obtain that the probability of the specified exploration is

$$O(n_3^{d(v)-J_1-K_1} m^{-d(v)-\lceil \frac{X_1}{2} \rceil}).$$

We have that $X_1 \geq 2d(v) - 2J_1 - K_1 \geq 4$ deterministically since $J_1 \leq 1$ and $K_1 \leq 2$. So, if $n_3 \leq 3$ the probability of the specified exploration is $O(\frac{u_{K_5^+}}{n})$ and we are done. If $X_1 \geq 5$ the probability of the specified exploration is $O(\frac{u_{K_4}}{m})$ So $X_1 = 4$ and we must have $J_1 = 1$ and $K_2 = 2$. Thus, the probability of the specified exploration is $O(\frac{u_{\Delta+1}}{m})$.

We consider next the possibility of a specified exploration from v in which v has no neighbour of degree 1. So, letting T_1 be the number of neighbours of v of degree 3,we have have $X_1 \geq 3d(v) - 2K_1 - T_1$ which implies $\lceil \frac{X_1}{2} \rceil + K_1 + T_1 \geq \lceil \frac{3}{2} d(v) + \frac{T_1}{2} \rceil$. Hence, applying Lemma 4.3 to the first iteration followed by Corollary 6.2 to the rest of the exploration , we have that the probability the specified exploration occurs is

$$O(\max(n_3 - 1, 1)^{T_1} m^{-K_1 - T_1 - \lceil \frac{X_1}{2} \rceil}) = O(\max(n_3 - 1, 1)^{T_1} m^{-\lceil \frac{3}{2} d(v) + \frac{T_1}{2} \rceil}).$$

If $n_3 < 4$ or $T_1 = 0$ this is $O(\frac{u_{K_5^+}}{n})$. If $n_3 \geq 4$ and $0 < T_1 < 5$ this is $O(\frac{u_{K_4}}{m})$.

If $n_3 \geq 4$ and $T_1 \geq 5$ setting $a = d(v) - 4$ and $a' = T_1 - 4$, we have $a \geq a'$ which implies that $\frac{a+a'}{2} \geq a'$ and

$$-\frac{3}{2} d(v) - \frac{T_1}{2} \leq -4 - \frac{a + 4 + a' + 4}{2} = -8 - \frac{a + a'}{2},$$

so the probability of the specified exploration is

$$O(n_3^{4+a'} m^{-8-\frac{a+a'}{2}}) = O(\max(n_3 - 3, 0)^4 m^{-8})$$

So, the probability there is such an exploration from v is $O(\frac{u_{K_4}}{m})$.

□

So, to complete the proof of Claim (C), we need only show:

Claim 6.4. *Claim (C) holds if we weaken its statement by requiring the component to contain no vertex of degree at least* 4.

Proof. Let H denote the component subgraph corresponding to the specified choices. Recall H contains at least one vertex of degree 3, at most one vertex of degree 1 and only vertices of degree at most 3. Applying Lemma 4.3 with $\delta^* + 1 \geq 3$, we see that the probability the specified exploration occurs somewhere is at most $m^{-|E(H)|} \prod_{v \in V(H)} n_{d(v)}$. We note that this is at most $m^{-\frac{1}{2} \sum_{v \in S} d(v)} \prod_{v \in S} n_{d(v)}$ for any $S \subseteq V(H)$ such that $V(H) \setminus S$ contains no degree 1 vertices. Recall that $n_1 \leq 1$.

So, if H contains:

(1) At least four vertices of degree 3, then since the number of vertices of odd degree is even, this probability is $O(u_{K_4})$;
(2) One vertex of degree 1 and at least two vertices of degree 2, then this probability is $O(u_{\Delta+1})$;
(3) No vertices of degree 1 and at least two vertices of degree 2, then since the number of vertices of odd degree is even, H contains at least two vertices of degree 3, so this probability is $O(u_{K_4 - e})$;

(4) One vertex of degree 1 and exactly one vertex of degree 2, then H contains two adjacent vertices of degree 3, so since the number of vertices of odd degree is even H contains at least three vertices of degree 3, so this is

$$O(m^{-6}n_2 \max(n_3 - 2, 0)^3) = O(u_{\Delta+1} + u_{K_4} + u_{K_5^+});$$

(5) No vertices of degree 1 and exactly one vertex of degree 2, then H contains at least three vertices of degree 3, so it contains at least four by parity;

(6) One vertex of degree 1 and no vertices of degree 2, then H contains at least three vertices of degree 3, but $(1, 3, 3, 3)$ is infeasible, so H contains at least four vertices of degree 3;

(7) No vertices of degree 1 or 2, then H contains at least four vertices of degree 3.

□

This completes the proof of Theorem 1.1. □

7. A Remark on Optimality

We show now that our bounds are tight up to a constant factor if $D^* \leq \frac{m}{3}$ and any of the following hold: $n_1 > 1$, $n_2 > 2$, or $n_3 = \Omega(n^{1/4})$. We need the following which is the part of Theorem 1 of [5] restricted to the case $H_2 = \emptyset$, which lower-bounds the probability of the existence of an edge (they use $J(\mathcal{D})$ for what we call D^* and $M = 2m$, we use ij in place of their uv, we have dropped any reference to H_2 and replaced H_1 by H, they use H^+ for the event that $H \subseteq G$ and Δ for $d(n)$).

Theorem 7.1. *Let H be a graph on $[n]$ such that H is a possible subgraph on a graph with degree sequence $\mathcal{D}$, and $uv \notin E(H)$. Then,*

$$\mathbf{P}\{ij \in E(G) \mid H \subseteq G\} \geq \left(1 - \frac{2D^* + 6d(n)}{2m - 2|E(H)|}\right) \frac{(d(i) - d_H(i))(d(j) - d_H(j))}{2m - 2|E(H)| + (d(i) - d_H(i))(d(j) - d_H(j))}.$$

Applying this theorem yields the following bound.

Corollary 7.2. *If $D^* \leq \frac{m}{3}$, then conditioned on the existence of any subgraph H of G with at most 8 vertices, each of which has degree at most three in G, we have for any two vertices i and j such that $ij \notin H$, $10 > d(i) > d_H(i)$, and $10 > d(j) > d_H(j)$, the conditional probability that $ij \in E(G)$ given that $H \subseteq G$ is $\Omega(\frac{1}{m})$.*

Proof. We have $D^* \geq 2d(n)$ so $2D^* + 6d(n) \leq \frac{5m}{3}$. Also, $|E(H)| \leq 24$ so for large m, $\frac{11m}{6} < 2m - |E(H)| < 2m$. So, $\frac{(d(i)-d_H(i))(d(j)-d_H(j))}{2m-|E(H)|} = \Theta(\frac{1}{m})$ and $\left(1 - \frac{2D^*+6d(n)}{2m-2|E(H)|}\right) = \Omega(1)$. □

It follows that under the assumption $D^* \leq \frac{m}{3}$, the expected number of edge components is $\Omega(u_{edge})$, the expected number of triangle components is $\Omega(u_\Delta)$,the expected number of components consisting of a degree 1 vertex attached to a triangle is $\Omega(u_{\Delta+1})$, the expected number of $K_4 - e$ components is $\Omega(u_{K_4-e})$ and the expected number of K_4 components is $\Omega(u_{K_4})$.

For each $J \in \{edge, \Delta, \Delta + 1, K_4 - e, K_4\}$, one can then show the expected number of pairs of components inducing J is $O(u_J)^2$ using Lemma 4.3. It then follows by Chebyshev's inequality that the probability there is a component inducing J is $\Omega(u_J)$. This shows our bound is tight unless no u_j is $\Omega(n/m^6)$, which is not the case if $n_1 > 2$, $n_2 > 3$, or $n_4 = \Omega(n^{1/4})$.

8. Acknowledgements

This work began while the first two authors were in residence at the Simons–Laufer Mathematical Sciences Institute (SLMath) during the Spring 2025 semester, supported by NSF grant DMS-1928930, and the third author was visiting SLMath. The first author acknowledges support from an NSERC Discovery Grant and from the Canada Research Chairs program.

References

[1] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357 – 367, 1967. 14

[2] A. Bekessy, P. Bekessy, and Janos Komlos. Asymptotic enumeration of regular matrices. *Studia Scientiarum Mathematicarum Hungarica*, 7, 01 1972. 3

[3] Colin Cooper, Alan Frieze, and Bruce Reed. Random regular graphs of non-constant degree: Connectivity and hamiltonicity. *Combinatorics, Probability and Computing*, 11(3):249–261, 2002. 5

[4] Lorenzo Federico and Remco Van der Hofstad. Critical window for connectivity in the configuration model. *Combinatorics, Probability and Computing*, 26(5):660–680, 2017. 5

[5] Pu Gao and Yuval Ohapkin. Subgraph probability of random graphs with specified degrees and applications to chromatic number and connectivity. *Random Structures & Algorithms*, 62(4):911–934, 2023. 5, 27

[6] Václav Havel. A remark on the existence of finite graphs. *Casopis Pest. Mat.*, 80:477–480, 1955. 3

[7] Felix Joos, Guillem Perarnau, Dieter Rautenbach, and Bruce Reed. How to determine if a random graph with a fixed degree sequence has a giant component. *Probability Theory and Related Fields*, 170(1):263–310, 2018. 5

[8] Tomasz Luczak. Sparse random graphs with a given degree sequence. In *Random Graphs*, volume 2, pages 165–182, 1992. 5

[9] Brendan D McKay. Subgraphs of random graphs with specified degrees. In *Congressus Numerantium Volume 33*, pages 213–223, 1981. 3

[10] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995. 5

[11] James K. Senior. Partitions and their representative graphs. *American Journal of Mathematics*, 73(3):663–689, 1951. 3

[12] Nicholas C Wormald. The asymptotic connectivity of labelled regular graphs. *Journal of Combinatorial Theory, Series B*, 31(2):156–167, 1981. 2, 3, 5

[13] Nicholas C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *Journal of Combinatorial Theory, Series B*, 31(2):168–182, 1981. 1