Maximum matchings in sparse random graphs: Karp-Sipser re-visited

Jonathan Aronson* Alan Frieze† Boris G. Pittel‡

November 21, 1997

Abstract

We study the average performance of a simple greedy algorithm for finding a matching in a sparse random graph $G_{n,c/n}$, where $c > 0$ is constant. The algorithm was first proposed by Karp and Sipser [12]. We give significantly improved estimates of the errors made by the algorithm. For the *sub-critical* case where $c < e$ we show that the algorithm finds a maximum matching with high probability. If $c > e$ then with high probability the algorithm produces a matching which is within $n^{1/5+o(1)}$ of maximum size.

1 Introduction

A matching in a graph $G = (V, E)$ is a set of edges in E which are vertex disjoint. A standard problem in algorithmic graph theory is to find the largest possible matching in a graph. The first polynomial time algorithm to solve this problem was devised by Edmonds in 1965 and runs in time $O(|V|^4)$ [10]. Over the years, many improvements have been made. Currently the fastest such algorithm is that of Micali and Vazirani which dates back to 1980. Its running time is $O(|E|\sqrt{|V|})$ [17]. These algorithms are rather complicated and there is a natural interest in the performance of simpler heuristic algorithms which should find large, but not necessarily maximum matchings. One well studied class of heuristics goes under the general title of the GREEDY heuristic.

GREEDY

> **begin**
> $M \leftarrow \emptyset$;
> **while** $E(G) \neq \emptyset$ **do**
> **begin**
> ***A***: Choose $e = \{x, y\} \in E$
> $G \leftarrow G \setminus \{x, y\}$;

*Department of Mathematics, Carnegie Mellon University, Pittsburgh PA15213, U.S.A.

†Department of Mathematics, Carnegie Mellon University, Pittsburgh PA15213, U.S.A. Supported in part by NSF grants CCR-9225008 and CCR-9530974

‡Department of Mathematics, Ohio State University, Columbus OH43210, U.S.A. Supported in part by NSA grant MDA 904-96-1-0053

$M \leftarrow M \cup \{e\}$
end;
Output M
end

($G \setminus \{x, y\}$ is the graph obtained from G by deleting the vertices x, y and all incident edges.)

There are many variations on this theme which depend on the exact choice of edge in Step $\boldsymbol{A}$. There is also the issue of whether to study the worst-case or the average-case. In this paper we study the average case – for work on the worst-case see Korte and Hausmann [14], Dyer and Frieze [8] or Aronson, Dyer, Frieze and Suen [1]. For the average case we need a model of a random graph. We will use the standard model $G_{n,M}$ which has vertex set $[n]$ and M random edges. We will consider the sparse case where $M = \lfloor cn/2 \rfloor$ and $c > 0$ is an absolute constant. (The average vertex degree then is $c + O(n^{-1})$.) The simplest method of choosing an edge in Step $\boldsymbol{A}$ is to choose e uniformly at random from all remaining edges. This algorithm was analysed by Dyer, Frieze and Pittel [9]. They proved (among other things) that $X(n, M)$, the size of the matching produced when GREEDY is aplied to $G_{n,M}$, is asymptotically normal with mean $n\phi(c)$ and variance $n\psi(c)$; here

$$\begin{aligned}\phi(c) &= \frac{c}{2(c+1)} \\ \psi(c) &= \frac{c^2(c+3)}{6(c+1)^4}.\end{aligned}$$

As one should expect, $\phi(c) \to 1/2$ as $c \to \infty$, which corresponds to a (near) perfect matching.

One can deduce from earlier results of Frieze [11] that this version of GREEDY is *not* asymptotically optimal, i.e. as $n \to \infty$ the ratio of the size of the matching produced to the size of the maximum matching does *not* tend to one in probability . On the other hand in an earlier seminal paper, Karp and Sipser [12] describe a modification which gives a remarkable improvement in average performance over GREEDY. We refer to their algorithm as KSGREEDY:

KSGREEDY

begin
$M \leftarrow \emptyset$;
while $E(G) \neq \emptyset$ **do**
begin
$\boldsymbol{A1}$: If G has a vertex of degree one, choose one, x say, randomly.
Let $e = \{x, y\}$ be the unique edge of G incident with x;
$\boldsymbol{A2}$: Otherwise, (no vertices of degree one) choose
$e = \{x, y\} \in E$ randomly
$G \leftarrow G \setminus \{x, y\}$;
$M \leftarrow M \cup \{e\}$
end;
Output M
end

The idea here is that while there are vertices of degree one, it is correct to choose edges incident with such vertices. By correct we mean that for any edge chosen in this way, there will be a maximum size matching containing this edge. Let Phase 1 of the algorithm end, and Phase 2 begin, at the first instance when G has minimum degree at least two. We exclude isolated vertices in this computation and note that x, y are considered to be isolated after Step A2. Thus the minimum degree of G is considered to be either one or at least two. Then

Fact 1 *No mistakes are made by KSGREEDY in Phase 1.*

Therefore the size of the maximum matching in G is equal to the number of edges in M at the end of Phase 1 plus the size of the maximum matching in the graph left after Phase 1.

To describe the results of Karp and Sipser requires a little notation. Initially $G = G_{n,M}$ is likely to have a large number $\approx nce^{-c}$ of vertices of degree one.

Let $R(n, M)$ denote the number of vertices remaining in the graph at the start of Phase 2. Let $L(n, M)$ denote the number of vertices in the graph at the start of Phase 2 which are not covered by the final matching. Note that Fact 1 implies that the final matching produced is within $L(n, M)/2$ of optimal in size. Karp and Sipser found a function $r(c)$ such that the following is true:

Theorem 1 *Let $M = \lfloor cn/2 \rfloor$ and let $\epsilon > 0$ be arbitrary. Then*

(a) $\lim_{n\to\infty} \mathbf{Pr}\left[\left|\frac{R(n,M)}{n} - r(c)\right| > \epsilon\right] = 0.$

(b) $\lim_{n\to\infty} \mathbf{Pr}\left[\frac{L(n,M)}{n} > \epsilon\right] = 0.$

(c) *If $c < e$ (the basis of natural logarithms) then*

$$\lim_{n\to\infty} \mathbf{Pr}\left[\frac{R(n,M)}{n} > \epsilon\right] = 0.$$

Part (b) means that KSGREEDY is **whp** (*with high probability* i.e. with probability $1 - o(1)$) asymptotically optimal! The surprising threshold result described in (c) is referred to as the e-phenomenon. The main aim of this paper is provide more detail on the size of the error terms. We prove the following theorems:

Theorem 2 *If $c < e$ then $R(n, M)$ converges in distribution to the sum R of a sequence of independent random variables $k\mathit{Poisson}(\gamma^k/2k)$, $k \geq 3$, where γ is the solution to the equation $c = \gamma e^\gamma$. In particular*

- *At the end of Phase 1 the graph is* **whp** *a collection of vertex disjoint cycles.*
- $\mathbf{E}(R) = \frac{\gamma^3}{2(1-\gamma)}$.
- $\mathbf{Pr}(R = 0) = ((1-\gamma)e^{-\gamma-\gamma^2/2})^{1/2}$.
- *KSGREEDY finds a maximum matching* **whp**.

Thus in this case, instead of concluding that $R(n, M) = o(n)$ **whp**, we can conclude the much stronger result that $R(n, M)$ is bounded in probability and that **whp** KSGREEDY finds a maximum matching.

Theorem 3 *If $c > e$ then* **whp**

$$\Omega(n^{1/5}/(\log n)^{75/2}) \leq \mathbf{E}(L(n,M)) \leq O(n^{1/5}(\log n)^{12}).$$

Remark: There is a gap of order $(\log n)^{99/2}$ between the upper and lower bound here. We have not been particularly careful in reducing the exponent of $\log n$, but on the other hand we do not know how to remove it completely.

Conjecture: $\mathbf{E}(L(n,M))/n^{1/5} \to \xi(c)$ as $n \to \infty$ for some function $\xi(c)$.

Remark: The reader will notice another gap. What happens when $c = e$? Karp and Sipser show that $R(n,M) = o(n)$ and $L(n,M) = o(n)$ **whp**.

Problem: determine the likely growth rate of $R(n,M)$ and $L(n,M)$ when $c = e$.

As a corollary of our work, we obtain the following tight estimate for the size $\mu(n,M)$ of the largest matching in $G_{n,M}$. This result was already obtained by Karp and Sipser (in different notation), but we repeat it here because of its importance.

Theorem 4 *Let $\epsilon > 0$ be fixed. Then*

$$\lim_{n\to\infty} \mathbf{Pr}\left[\left|\frac{\mu(n,M)}{n} - \left(1 - \frac{\gamma^* + \gamma_* + \gamma^*\gamma_*}{2c}\right)\right| \geq \epsilon\right] = 0.$$

Here γ_* is the smallest root of the equation $x = c\exp(-ce^{-x})$ and $\gamma^* = ce^{-\gamma_*}$.

More precisely, we show that the Karp-Sipser ϵ can be replaced by $n^{-1/6+\epsilon}$ for $c < e$, and by $n^{-1/7+\epsilon}$ for $c > e$.

For $c < 1$, when $\gamma_* = \gamma^* = \gamma$, Pittel [19] proved that $\mu(n,M)$ is asymptotically Gaussian with mean $n[1 - (2\gamma + \gamma^2)/(2c)]$, and variance $n\sigma^2(c)$. It seems plausible that $\mu(n,M)$ is Gaussian in the limit for every c.

Among the techniques used in this paper, we derive two systems of differential equations whose solutions provide a deterministic approximation for the dynamics of the deletion process. These seemingly complicated equations have unexpectedly simple integrals which lead to an alternative proof of the e-phenomenon. We use the integrals to construct certain supermartingales, and to provide the probabilistic bounds for the deviations of the actual realizations of the process from those solutions. Analogous techniques had been used by Pittel, Spencer, and Wormald [20]. The notion of differential equations as an approximation tool in random processes has long been known, of course, but their first serious application in the random graph setting seems to have been Karp and Sipser's use of Kurtz's Theorem [15].

We should also mention a recent work by Bollobás and Brightwell [5] on the independence number of a sparse bipartite graph. Their approach is technically quite different since it is based on analysis of an algorithm that deletes, at each step, all the vertices of degree one. This was the approach used in Aronson [2].

2 Random Sequence Model

A slight change of model will simplify the analysis. Given a sequence $\mathbf{x} = x_1, x_2, \ldots, x_{2M}$ of integers between 1 and n we can define a (multi)-graph $G_{\mathbf{x}}$ with vertex set $[n]$ and edge set $\{(x_{2i-1}, x_{2i}) : 1 \le i \le M\}$. If $\mathbf{x}$ is chosen randomly from $[n]^{2M}$ then $G_{\mathbf{x}}$ is close in distribution to $G_{n,M}$. Indeed,

Lemma 1 **(a)** *Conditional on being simple, $G_{\mathbf{x}}$ is distributed as $G_{n,M}$.*

(b) $\lim_{n\to\infty} \mathbf{Pr}(G_{\mathbf{x}}$ *is simple*$) = \exp\left\{-\frac{c}{2} - \frac{c^2}{4}\right\}$, *if* $M = \lfloor cn/2 \rfloor$.

Proof (a) If $G_{\mathbf{x}}$ is simple then it has vertex set $[n]$ and M edges. Also, there are $M!2^M$ distinct equally likely values of $\mathbf{x}$ which yield the same graph.

(b) Let $N = \binom{n}{2}$. Then

$$\begin{aligned} \mathbf{Pr}(G_{\mathbf{x}} \text{ is simple}) &= \left(1 - \frac{1}{n}\right)^M \prod_{j=0}^{M-1} \left(1 - \frac{j}{N}\right) \\ &= \exp\left\{-\frac{c}{2} - \frac{c^2}{4} + O\left(\frac{1}{n}\right)\right\}. \end{aligned} \qquad (1)$$

□

Given the above lemma, we will be able to analyse the likely evolution of KSGREEDY on $G_{n,M}$ by changing the input to $\mathbf{x}$. (We will show later how to translate our results back to $G_{n,M}$). As the algorithm progresses it produces some conditioning. Consider the first iteration. When an edge is removed we will replace it in $\mathbf{x}$ by a pair of $\star$'s. This goes for all of the edges removed at an iteration, not just the matching edge $\{x, y\}$. Thus at the end of this and subsequent iterations we will have a sequence in $Z = ([n] \cup \{\star\})^{2M}$ where for all i, $x_{2i-1} = \star$ if and only if $x_{2i} = \star$. We call such sequences *proper*.

For a proper $\mathbf{z} \in Z$ and vertex $j \in [n]$ we let its *degree* $d_{\mathbf{z}}(j)$ be the number of occurrences of j in $\mathbf{z}$. We denote $\mathbf{d_z} = \{d_{\mathbf{z}}(j)\}_{j\in[n]}$, and call $\mathbf{d_z}$ the *degree sequence* of $\mathbf{z}$. Let $V(\mathbf{z}) = \{j : d_{\mathbf{z}}(j) > 0\}$ and $S(\mathbf{z}) = \{i : z_{2i-1} = z_{2i} = \star\}$. For a tuple $\mathbf{v} = (v_0, v_1, v, 2m)$ with $v_0 + v_1 + v = n$ and $m \le M$ we let $Z_{\mathbf{v}}$ denote the set of proper $\mathbf{z} \in Z$ with

- v_0 vertices of degree 0,
- v_1 vertices of degree 1,
- v vertices of degree at least 2 and
- $2m$ non-$\star$ entries.

This corresponds to a multi-graph $G_{\mathbf{z}}$ with m edges. Note that

$$v_0 + v_1 + v = n,$$

since x, y are treated as vertices of degree zero after Step 2 of KSGREEDY.

Given $\mathbf{z} \in Z_{\mathbf{v}}$ and a permutation π of $[2M]$ let $\mathbf{z}_\pi = z_{\pi(1)}, z_{\pi(2)}, \ldots, z_{\pi(2M)}$. We call π *proper* if the $\star$ entries of $\mathbf{z}$ are fixed under π. For a proper π, $\mathbf{z}_\pi$ is certainly proper. Let $A_{\mathbf{v}}(\mathbf{z}) = \{\mathbf{z}_\pi : \pi \text{ is proper}\}$. The sets $\mathcal{A} = \{A_{\mathbf{v}}(\mathbf{z})\}$ partition $Z_{\mathbf{v}}$ into the equivalence classes. If two proper sequences $\mathbf{z}'$ and $\mathbf{z}''$ are equivalent then $\mathbf{d}_{\mathbf{z}'} = \mathbf{d}_{\mathbf{z}''}$, and $\{z'_{2t-1}, z'_{2t}\} = \{\star, \star\}$ iff $\{z''_{2t-1}, z''_{2t}\} = \{\star, \star\}$. So the $\mathbf{z}$ from the same class have the *starred* pairs at the same locations. Clearly, the size of an equivalence class with a common degree $\mathbf{d} = \{d(j)\}_{j\in[n]}$ is $(2m)!/[\prod_j d(j)!]$. The following simple fact and its immediate Corollary will be instrumental in our proofs.

Fact 2 *For a given* $\mathbf{v}$, *let* A *be a fixed equivalence class from* $\mathcal{A}$. *If* $\mathbf{z}$ *is a random member of* $Z_{\mathbf{v}}$, *then conditional on* $\mathbf{z} \in A$, $\mathbf{z}$ *is a random member of* A. *In other words, the conditional probability of each feasible value of* $\mathbf{z}$ *is the same, namely* $[(2m)!/\prod_j d(j)!]^{-1}$, *where* $\mathbf{d}$ *is the degree of the sequences from* A.

Recall that $(x)_a = x(x-1)\cdots(x-a+1)$.

Corollary 1 *Given* A, *let* $\mathbf{t} = \{t_1 < \ldots < t_r\}$ *be such that all* $\{2t_s - 1, 2t_s\}$ *are non-starred. Let* $i_s, j_s \in [n]$ *be given,* $s \in [r]$. *Denote by* $P(\mathbf{t}; \mathbf{i}, \mathbf{j})$ *the conditional probability that* $(z_{2t_s-1}, z_{2t_s}) = (i_s, j_s)$, $s \in [r]$, *given* $\mathbf{z} \in A$. *Then*

$$
\begin{aligned}
P(\mathbf{t};\mathbf{i},\mathbf{j}) &= \frac{\prod_{k\in[n]}(d(k))_{\mu_k}}{(2m)_{2r}}, \qquad (2)\\
\mu_k = \mu_k(\mathbf{i},\mathbf{j}) : &= |\{s : i_s = k\}| + |\{s : j_s = k\}|,
\end{aligned}
$$

We now study the random sequence $\mathbf{z}(0) = \mathbf{x}, \mathbf{z}(1), \mathbf{z}(2), \ldots,$ of sequences produced by KS-GREEDY and the corresponding sequence $\mathbf{v}(0), \mathbf{v}(1), \mathbf{v}(2), \ldots,$ where $\mathbf{z}(t) \in Z_{\mathbf{v}(t)}$.

We let $A(t)$ be the equivalence class of $\mathbf{z}(t)$ and $G(t) = G_{\mathbf{z}(t)}$ for $t \geq 0$.

Lemma 2 *Suppose that* $\mathbf{z}(0)$ *is a random member of* $Z_{\mathbf{v}(0)}$. *Then given* $\mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t)$, *the vector* $\mathbf{z}(t)$ *is a random member of* $Z_{\mathbf{v}(t)}$ *for all* $t \geq 0$, *that is, the conditional distribution of* $\mathbf{z}(t)$ *is uniform.*

Proof We prove this by induction on t. It is trivially true for $t = 0$. Fix $t \geq 0$, $\mathbf{v} = \mathbf{v}(t)$, $\mathbf{v}' = \mathbf{v}(t+1)$. We start by proving

Claim: Each $\mathbf{z}' \in Z_{\mathbf{v}'}$ arises by a transition of KSGREEDY from the same number $D(\mathbf{v}, \mathbf{v}')$ of $\mathbf{z} \in Z_{\mathbf{v}}$. Suppose for example that $v_1 > 0$, so an edge incident to a pendant vertex is to be deleted, together with other incident edges. To recover $\mathbf{z} \in Z_{\mathbf{v}}$ from $\mathbf{z}' \in Z_{\mathbf{v}'}$, we are

1. choosing a subset $J \subseteq S(\mathbf{z}')$ of cardinality $m - m'$.
2. choosing element $x \in V^c(\mathbf{z}')$.
3. choosing element $y \in V^c(\mathbf{z}') \setminus \{x\}$.
4. choosing $i \in J$ and assigning $\{z_{2i-1}, z_{2i}\} = \{x, y\}$.
5. choosing $\ell = v_0' - v_0 - 2$ vertices $u_1, u_2, \ldots, u_\ell$ from $V^c(\mathbf{z}') \setminus \{x, y\}$.
6. choosing $m'' :\ \ell \leq m'' \leq m - m' - 1$, (unless $\ell = 0$, in which case $m'' = 0$).

   [m'' is the number of edges joining y to vertices other than x which become isolated by the step.]
7. choosing $\alpha \geq 0$ such that $\ell' = v_1' - v_1 + \alpha + 1 + \chi_{\{m=m'+1\}} \geq 0$.

   [α is the number of of $u_1, u_2, \ldots, u_\ell$ which are of degree one in $\mathbf{z}$]

   [ℓ' is the number of vertices which are of degree 1 in $\mathbf{z}'$, but of degree > 1 in $\mathbf{z}$.]
8. choosing a surjection $\phi : [m''] \to [\ell]$ such that $\alpha = |\{u \mid \exists\, e \text{ s.t. } \phi(e) = u\}|$.

9. choosing indices $K = \{k_1, k_2, \ldots, k_{m''}\} \subseteq J \setminus \{i\}$ and assigning pairs $\{z_{2k_r-1}, z_{2k_r}\} = \{u_{\phi(r)}, y\}$ for $1 \le r \le m''$.
10. choosing vertices $w_1, w_2, \ldots, w_{\ell'}$ of degree 1 in $\mathbf{z}'$.
11. choosing $m''' : \ \ell' \le m''' \le m - m' - m'' - 1$ (unless $\ell' = 0$, in which case $m''' = 0$) and a surjection $\psi : [m'''] \to [\ell']$.
12. choosing ℓ' indices $h_1, h_2, \ldots, h_{\ell'}$ in $J \setminus (\{i\} \cup K)$ and assigning pairs $\{z_{2h_s-1}, z_{2h_s}\} = \{w_{\psi(s)}, y\}$ for $1 \le s \le \ell'$.
13. assigning to each of the remaining $m - m' - m'' - m''' - 1$ pairs of $\star$'s a set $\{y, y\}$ or $\{\omega, y\}$ where each ω is of degree at least 2 in $\mathbf{z}'$.

In each step, the number of options is the same for all $\mathbf{z}' \in Z_{\mathbf{v}'}$. The statement (for $v_1 > 0$) follows since the total number of ways to recover $\mathbf{z} \in Z_{\mathbf{v}}$ equals the product of those numbers. The exact value $D(\mathbf{v}, \mathbf{v}')$ of the product will not be important to us.

A similar accounting is possible when $v_1 = 0$ which we leave to the reader. This completes the proof of our claim.

If $\mathbf{z}' \in Z_{\mathbf{v}(t+1)}$ then the inductive assumption and the Markov property of the process $\{\mathbf{z}(t)\}$ implies — via conditioning on $\mathbf{v}(t)$ — that

$$\mathbf{Pr}(\mathbf{z}(t+1) = \mathbf{z}' \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t)) = \frac{1}{|Z_{\mathbf{v}}|} \sum_{\mathbf{z} \in Z_{\mathbf{v}}} \mathbf{Pr}(\mathbf{z}(t+1) = \mathbf{z}' \mid \mathbf{z}(t) = \mathbf{z}).$$

Now, let $N(\mathbf{v})$ denote the number of choices of transition T available for KSGREEDY on sequence $\mathbf{z} \in Z_{\mathbf{v}}$. The notation underscores the fact that this number depends on $\mathbf{v}$ only. Indeed, it equals v_1 if $v_1 > 0$ and m otherwise. Hence, if T refers to the t-th choice of transition

$$\mathbf{Pr}(\mathbf{z}(t+1) = \mathbf{z}' \mid \mathbf{z}(t) = \mathbf{z}) = \frac{1}{N(\mathbf{v})} \sum_T 1_{\{\mathbf{z}' \text{arises from } \mathbf{z}, T\}}.$$

Using our claim, we obtain

$$\mathbf{Pr}(\mathbf{z}(t+1) = \mathbf{z}' \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t)) = \frac{D(\mathbf{v}, \mathbf{v}')}{N(\mathbf{v})|Z_{\mathbf{v}}|}.$$

This probability is independent of $\mathbf{z}' \in Z_{\mathbf{v}'}$. But then so is $\mathbf{Pr}(\mathbf{z}(t+1) = \mathbf{z}' \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t+1))$, since it equals the ratio of the above probability and $\mathbf{Pr}(\mathbf{z}(t+1) \in Z_{\mathbf{v}'} \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t))$. □

As a consequence

Lemma 3 *The random sequence* $\mathbf{v}(t)$, $t = 0, 1, 2, \ldots,$ *is a Markov chain.*

Proof Slightly abusing notation,

$$\begin{aligned}\mathbf{Pr}(\mathbf{v}(t+1) \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t)) &= \sum_{\mathbf{z}' \in Z_{\mathbf{v}(t+1)}} \mathbf{Pr}(\mathbf{z}' \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t)) \\ &= \sum_{\mathbf{z}' \in Z_{\mathbf{v}(t+1)}} \sum_{\mathbf{z} \in Z_{\mathbf{v}(t)}} \mathbf{Pr}(\mathbf{z}', \mathbf{z} \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t))\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\mathbf{z}'\in Z_{\mathbf{v}(t+1)}} \sum_{\mathbf{z}\in Z_{\mathbf{v}(t)}} \mathbf{Pr}(\mathbf{z}' \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t-1), \mathbf{z}) \\
&\quad \times \mathbf{Pr}(\mathbf{z} \mid \mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(t)) \\
&= \sum_{\mathbf{z}'\in Z_{\mathbf{v}(t+1)}} \sum_{\mathbf{z}\in Z_{\mathbf{v}(t)}} \mathbf{Pr}(\mathbf{z}' \mid \mathbf{z}) |Z_{\mathbf{v}(t)}|^{-1},
\end{aligned}
$$

which depends only on $\mathbf{v}(t), \mathbf{v}(t+1)$. □

We will also need the following corollary of Lemma 2 and Lemma 4.

Corollary 2 *Let $\mathcal{T}$ be a stopping time adapted to $\{\mathbf{v}(t)\}$. Conditioned on $\mathbf{v}(\mathcal{T})$, the sequence $\mathbf{z}(\mathcal{T})$ is distributed uniformly on $Z_{\mathbf{v}(\mathcal{T})}$.*

The proof of this intuitively clear statement is simple, and we omit it for brevity. □

3 Transition Probabilities

In the light of Lemma 2, we will discuss the following problem: let tuple $\mathbf{v}$ be given. Suppose $\mathbf{z}$ is chosen randomly from $Z_{\mathbf{v}}$ and one iteration of KSGREEDY is carried out. This will yield $\mathbf{z}' \in Z_{\mathbf{v}'}$. What can we say about $\mathbf{v}'$? As a preparation we discuss the degree sequence of $\mathbf{z}$.

Lemma 4 *Let $\mathbf{z}$ be chosen randomly from $Z_{\mathbf{v}}$. Let $J = J(\mathbf{z}) = \{j \in [n] : d_{\mathbf{z}}(j) \geq 2\}$ and let X_j $(j \in J)$ denote the degrees of vertices from J. Let Z_j $(j \in J)$ be independent copies of a* truncated Poisson *random variable Z, where*

$$
\mathbf{Pr}(Z = k) = \frac{z^k}{k!f(z)}, \qquad k = 2, 3, \ldots .
$$

Here

$$
f(z) = e^z - 1 - z
$$

and z satisfies

$$
\frac{z(e^z - 1)}{f(z)} = \zeta, \tag{3}
$$

where

$$
\zeta = \frac{s}{v}, \quad s := 2m - v_1 \tag{4}
$$

being the total degree of vertices from J. Then $\{X_j\}_{j\in J}$ is distributed as $\{Z_j\}_{j\in J}$ conditional on $\sum_{j\in J} Z_j = s$.

Proof Note first that the value of z in (3) is chosen so that

$$
\mathbf{E}(Z) = \zeta.
$$

Assume without loss of generality that $J = [v]$. Let

$$
S = \left\{ \vec{x} \in [s]^v \,\middle|\, \sum_{1\leq j\leq v} x_j = s \text{ and } \forall j,\, x_j \geq 2 \right\}.
$$

Fix $\vec{\xi} \in S$. Then, by the definition of $\mathbf{z}$ and $\{X_j\}_{j\in[v]}$,

$$\mathbf{Pr}(\vec{X} = \vec{\xi}) = \left(\frac{s!}{\xi_1!\xi_2!\dots\xi_v!}\right) \Big/ \left(\sum_{\vec{x}\in S} \frac{s!}{x_1!x_2!\dots x_v!}\right).$$

On the other hand,

$$\begin{aligned}
\mathbf{Pr}\left(\vec{Z} = \vec{\xi} \,\middle|\, \sum_{1\le j\le v} Z_j = s\right) &= \left(\prod_{1\le j\le v} \frac{z^{\xi_j}}{(e^z-1-z)\xi_j!}\right) \Big/ \left(\sum_{\vec{x}\in S}\prod_{1\le j\le v} \frac{z^{x_j}}{(e^z-1-z)x_j!}\right) \\
&= \left(\frac{(e^z-1-z)^{-v}z^s}{\xi_1!\xi_2!\dots\xi_v!}\right) \Big/ \left(\sum_{\vec{x}\in S} \frac{(e^z-1-z)^{-v}z^s}{x_1!x_2!\dots x_v!}\right) \\
&= \mathbf{Pr}(\vec{X} = \vec{\xi}).
\end{aligned}$$

□

The reader has certainly noticed that the statement holds for every $z > 0$. Intuitively, one wants to avoid conditioning on a "thin" event, and our z almost maximizes $\mathbf{Pr}(\sum_{j\in J} Z_j = s)$ in the interesting cases. Estimational advantages of our choice will be seen shortly. The chosen parameter z is extremely important to the ensuing analysis. It is a measure of the density of the residual graph. It is initially (close to) c and we expect it to tend to zero, in which case G has mainly vertices of degree two. A simple first moment calculation shows that **whp** $G(n, M)$ contains no subgraph with average degree $3c$ or more. We therefore restrict our attention from now on to $\mathbf{v}$ with $2m - v_1 \le 3cv$. For those $\mathbf{v}$'s, according to (3) and (4) the parameter z is bounded above by an absolute constant.

To use Lemma 4 for approximation of vertex degrees distributions we need to have sharp estimates of the probability that $\sum_{1\le j\le v} Z_j$ is close to its mean s. In particular we need sharp estimates of $\mathbf{Pr}(\sum_{1\le j\le v} Z_j = s)$ and $\mathbf{Pr}(\sum_{2\le j\le v} Z_j = s - k)$, for $k = o(v)$. Those estimates are possible precisely because $\mathbf{E}(\sum_{1\le j\le v} Z_j) = s$! The well known local limit theorem (Durrett [7], Theorem 5.2, p 113) is not sufficient, since we need to cover the case $\sigma^2 := \mathbf{Var}(Z) = \frac{z(e^z-1)^2 - z^3e^z}{f^2(z)} = o(1)$. Fortunately, using the special properties of Z, we can refine a standard argument to show (Appendix 1) that for $v\sigma^2 \to \infty$

$$\mathbf{Pr}\left(\sum_{j=1}^{v} Z_j = s\right) = \frac{1}{\sigma\sqrt{2\pi v}}(1 + O(v^{-1}\sigma^{-2})) \tag{5}$$

and

$$\mathbf{Pr}\left(\sum_{j=2}^{v} Z_j = s - k\right) = \frac{1}{\sigma\sqrt{2\pi v}}\left(1 + O((k^2+1)v^{-1}\sigma^{-2})\right), \tag{6}$$

if, in addition, $k = O(v^{1/2}\sigma)$. **Whp** the maximum degree of $G(n, M)$ is $o(\log n)$ and our process will be closely analysed only while $vz \ge n^{1/5+o(1)}$. When applied, the offset k represents the sum of one or two vertex degrees. Since $\sigma^2 \approx z/3$ as $z \to 0$ we see that $k = O(v^{1/2}\sigma)$ is where our interest lies.

Borrowing a term from Knuth, Motwani, and Pittel [13], we say that an event $\mathcal{E} = \mathcal{E}(n)$ occurs quite surely (**qs**, in short) if $\mathbf{Pr}(\mathcal{E}) = 1 - O(n^{-a})$ for any constant $a > 0$.

Lemma 4, (5) and (6) plus a standard tail estimate for the binomial distribution shows that there is an absolute constant $K > 0$ such that the following event $\mathcal{D}(t)$ occurs **qs**: let $v_k = v_k(t)$ denote the number of vertices of degree k in $G(t)$; then

$$\mathcal{D}(t) = \left\{ \left| v_k - \frac{vz^k}{k!f(z)} \right| \leq K_1 \left(1 + \sqrt{vz^k/(k!f(z))} \right) \log n,\ 2 \leq k \leq \log n \right\}. \tag{7}$$

A simple first moment calculation shows that **whp** $G(n, M)$ contains no subgraph with average degree $3c$ or more. We therefore restrict our attention from now on to $\mathbf{v}$ with $2m - v_1 \leq 3cv$.

Lemma 5 **(a)** *Assume that* $\log n = O((vz)^{1/2})$. *For every* $j \in J$ *and* $2 \leq k \leq \log n$,

$$\mathbf{Pr}(X_j = k \mid \mathbf{v}) = \frac{z^k}{k!f(z)} \left(1 + O\left(\frac{k^2+1}{vz} \right) \right).$$

Furthermore, for all $j_1, j_2 \in J$, $j_1 \neq j_2$, *and* $2 \leq k_1, k_2 \leq \log v$,

$$\mathbf{Pr}(X_{j_1} = k_1, X_{j_2} = k_2 \mid \mathbf{v}) = \frac{z^{k_1}}{k_1!f(z)} \frac{z^{k_2}}{k_2!f(z)} \left(1 + O\left(\frac{\log^2 v}{vz} \right) \right).$$

(b) *For all* $k \geq 2$

$$\mathbf{Pr}(X_j = k \mid \mathbf{v}) = O\left((vz)^{1/2} \frac{z^k}{k!f(z)} \right).$$

Proof Assume that $J = [v]$ and $j = 1$. Then

$$\begin{aligned} \mathbf{Pr}(X_1 = k \mid \mathbf{v}) &= \frac{\mathbf{Pr}\left(Z_1 = k \text{ and } \sum_{i=1}^{v} Z_i = s \right)}{\mathbf{Pr}\left(\sum_{i=1}^{v} Z_i = s \right)} \\ &= \frac{z^k}{k!f(z)} \frac{\mathbf{Pr}\left(\sum_{i=2}^{v} Z_i = s - k \right)}{\mathbf{Pr}\left(\sum_{i=1}^{v} Z_i = s \right)}. \end{aligned}$$

Likewise, with $j_1 = 1, j_2 = 2$,

$$\mathbf{Pr}(X_1 = k_1, X_2 = k_2 \mid \mathbf{v}) = \frac{z^{k_1}}{k_1!f(z)} \frac{z^{k_2}}{k_2!f(z)} \frac{\mathbf{Pr}\left(\sum_{i=3}^{v} Z_i = s - k_1 - k_2 \right)}{\mathbf{Pr}\left(\sum_{i=1}^{v} Z_i = s \right)}.$$

The statements (a),(b) now follow immediately from (5) and (6). □

Our aim now is to use this lemma to compute $\mathbf{E}(\mathbf{v}' - \mathbf{v} \mid \mathbf{v})$ for both Step A1 and Step A2 of KSGREEDY.

Lemma 6 *Assuming that* $\log n = O((vz)^{1/2})$ *and* $0 < v_1 = O(v)$, *and abbreviating* $f = f(z)$,

$$\begin{aligned} \mathbf{E}[v_1' - v_1 | \mathbf{v}] &= -1 - \frac{v_1}{2m} + \frac{v^2 z^4 e^z}{(2mf)^2} - \frac{v_1 v z^2 e^z}{(2m)^2 f} + O\left(\frac{\log^2 v}{vz} \right), \\ \mathbf{E}[v' - v | \mathbf{v}] &= -1 + \frac{v_1}{2m} - \frac{v^2 z^4 e^z}{(2mf)^2} + O\left(\frac{\log^2 v}{vz} \right), \\ \mathbf{E}[m' - m | \mathbf{v}] &= -1 - \frac{v z^2 e^z}{2mf} + O\left(\frac{\log^2 v}{vz} \right). \end{aligned}$$

$$\Omega\left(\frac{v_1 - 1}{m} \right) \leq \mathbf{E}[v_0' - v_0 - 2 | \mathbf{v}] \leq O\left(\frac{v_1}{m} \right),$$

where the lower bound assumes $v_1 \leq m$.

Proof

Let x, y be as in Step A1 of KSGREEDY. Introduce the parameters d_r, $(1 \le r \le 3)$; they are the total number of neighbors of y in the multigraph $G_{\mathbf{z}}$ that have degree 1, 2, and at least 3 respectively. Set $d := d_1 + d_2 + d_3$. Let ℓ, δ denote the number of loops and multiple edges incident with y, and let $\tilde{d} = 2\ell + \delta$ (an edge of multiplicity k counts $k-1$ towards δ). Thus $d + \tilde{d}$ is the degree of y.

We show first that

$$\mathbf{E}(\tilde{d} \mid \mathbf{v}) = O\left(\frac{1}{m}\right). \tag{8}$$

Let A be the equivalence class that contains $\mathbf{z}$, and let $\mathbf{d} = \{d(j)\}_{j\in[n]}$ denote the corresponding degree. Using Corollary 1, we compute

$$\begin{aligned}\mathbf{E}(2\ell \mid A) &= \frac{2m(2m-2)}{(2m)_4}\sum_k (d(k))_3\\ &= \frac{1}{(2m-1)(2m-3)}\sum_k (d(k))_3.\end{aligned}$$

(Here $2m(m-1)$ counts the number of ways to pair a fixed vertex of degree one and a fixed vertex k, and to form a pair $\{k,k\}$ in the sequence $\mathbf{z}$.) Likewise

$$\begin{aligned}\mathbf{E}(\delta \mid A) &\le \frac{2m\binom{m-1}{2}2^2}{(2m)_6}\sum_{j,k}(d(j))_2(d(k))_3\\ &= \frac{1}{2(2m-1)(2m-3)(2m-5)}\sum_{j,k}(d(j))_2(d(k))_3.\end{aligned}$$

(The number of multiple edges joining j and k is at most the number of ways to sample—without order and replacement—two pairs among $\{z_{2t-1}, z_{2t}\} = \{j,k\}$.) Combining the last two estimates with Lemma 5 we obtain

$$\begin{aligned}\mathbf{E}(\tilde{d} \mid \mathbf{v}) &= O\left(\frac{v^2}{m^3}\mathbf{E}(X_1^3 \mid \mathbf{v})\mathbf{E}(X_1^2 \mid \mathbf{v})\right)\\ &= O\left(\frac{1}{m}\right).\end{aligned} \tag{9}$$

The estimate (8) will allow us to handle easily the annoying complexity of possible neighborhoods of y due to loops and multiple edges incident with it.

Consider two possible alternatives.

Case (a) $d \ge 2$. Then the deletion of all the edges incident with y leads to the new state $\mathbf{v}' = (v_0', v_1', v', 2m')$, where

$$\begin{aligned}v_1' &= v_1 - d_1 + d_2 + O(\tilde{d}),\\ v' &= v - (1 + d_2) + O(\tilde{d}),\\ 2m' &= 2m - 2d + O(\tilde{d}).\end{aligned} \tag{10}$$

Case (b) $d = 1$. Then

$$v_1' = v_1 - 2, \tag{11}$$

$$\begin{aligned} v' &= v, \\ 2m' &= 2m-2. \end{aligned}$$

Let $\chi_{\mathbf{a}}$, $\chi_{\mathbf{b}}$ be the indicator of the event **Case (a)** and of the event **Case (b)** respectively. Let us compute $\mathbf{E}(d_r\chi_{\mathbf{a}} \mid \mathbf{v})$, $(r = 1, 2, 3)$. We start with the conditional expectations, given the equivalence class A. Let $\mathbf{d} = \mathbf{d}(A)$ be the degree of $\mathbf{z}$'s from A. By (2), the (conditional) probability that a vertex i with degree 1 is incident to a vertex j is

$$\frac{2m}{(2m)_2} d(j) = \frac{d(j)}{2m-1}. \tag{12}$$

This formula implies directly that

$$\begin{aligned} \mathbf{E}(\chi_{\mathbf{a}} \mid A) &= 1 - \mathbf{E}(\chi_{\mathbf{b}} \mid A) \\ &= 1 - \frac{v_1 - 1}{2m-1} = \frac{2m - v_1}{2m-1}. \end{aligned} \tag{13}$$

Furthermore, the probability that a vertex i of degree 1 and a vertex ℓ are neighbors of a vertex j, $(d(j) \geq 2)$, is

$$\begin{aligned} \frac{d(\ell)(d(j))_2}{(2m-1)(2m-3)} &- O\left(\frac{(d(\ell))_2(d(j))_3}{(2m-1)(2m-3)(2m-5)}\right) \\ &= \frac{d(\ell)(d(j))_2}{(2m-1)(2m-3)} - O(m^{-3}d(\ell)^2 d(j)^3). \end{aligned} \tag{14}$$

We have used

$$\sum_s \mathbf{Pr}(B_s) - \sum_{s_1<s_2} \mathbf{Pr}(B_{s_1} \cap B_{s_2}) \leq \mathbf{Pr}(\cup_s B_s) \leq \sum_s \mathbf{Pr}(B_s). \tag{15}$$

So, using (12) and the last relation with $d(\ell) = 1$,

$$\begin{aligned} \mathbf{E}(d_1\chi_{\mathbf{a}} \mid A) &= \frac{1}{2m-1} \sum_{j:d(j)\geq 2} d(j) + \frac{v_1 - 1}{(2m-1)(2m-3)} \sum_{j:d(j)\geq 2} (d(j))_2 \\ &= \frac{2m - v_1}{2m-1} + \frac{v_1 - 1}{(2m-1)(2m-3)} \sum_j (d(j))_2, \end{aligned} \tag{16}$$

within an error $O(m^{-2}\sum_j d^3(j))$. Thus, by Lemma 5,

$$\begin{aligned} \mathbf{E}(d_1\chi_{\mathbf{a}} \mid \mathbf{v}) &= \frac{2m - v_1}{2m-1} + \frac{(v_1-1)v}{(2m-1)(2m-3)} \mathbf{E}((X_1)_2) + O(m^{-1}) \\ &= \frac{2m - v_1}{2m-1} + \frac{(v_1-1)v}{(2m-1)(2m-3)} \left(\sum_{d=2}^{\lfloor \log n \rfloor} \frac{d(d-1)z^d}{d!f(z)} \left(1 + O\left(\frac{\log^2 v}{vz}\right)\right)\right. \\ &+ \left. O\left((vz)^{1/2} \sum_{d>\lfloor \log n \rfloor} d^2 \frac{z^d}{d!f(z)}\right)\right) + O(m^{-1}) \\ &= \frac{2m - v_1}{2m-1} + \frac{(v_1-1)v}{(2m-1)(2m-3)} \frac{z^2 e^z}{f(z)} \left(1 + O\left(\frac{\log^2 v}{vz}\right)\right) \\ &\quad + O(m^{-1}) \\ &= 1 - \frac{v_1}{2m} + \frac{v_1 v z^2 e^z}{(2m)^2 f(z)} + O\left(\frac{\log^2 v}{vz}\right). \end{aligned} \tag{17}$$

Next, using (14) with $d(\ell)=2$,

$$\begin{aligned}
\mathbf{E}(d_2\chi_{\mathbf{a}} \mid A) &= \frac{2}{(2m-1)(2m-3)} \sum_{j\neq\ell:d(\ell)=2,d(j)\geq 2} (d(j))_2 \\
&= \frac{2}{(2m-1)(2m-3)} \sum_j (v_2 - \chi_{\{d(j)=2\}})(d(j))_2,
\end{aligned}$$

within an error $O(m^{-2}\sum_j d^3(j))$; here v_2 is the total number of vertices of degree 2. So, by Lemma 5 again,

$$\begin{aligned}
\mathbf{E}(d_2\chi_{\mathbf{a}} \mid \mathbf{v}) &= \frac{2v^2}{(2m)^2}\frac{z^2/2}{f(z)}\mathbf{E}[(X_1)_2]\left(1+O\left(\frac{\log^2 v}{vz}\right)\right)+O(m^{-1}) \\
&= \frac{v^2z^4e^z}{(2mf(z))^2}+O\left(\frac{\log^2 v}{vz}\right). \qquad (18)
\end{aligned}$$

Indeed,

$$v_2 = \sum_j \chi_{\{d(j)=2\}},$$

and, for $j\neq k$,

$$\begin{aligned}
\mathbf{E}(\chi_{\{d(j)=2\}}(d(k))_2 \mid \mathbf{v}) &= \left(1+O\left(\frac{\log^2 v}{vz}\right)\right)\left[\frac{z^2/2}{f(z)}\sum_{d=2}^{\lfloor \log n\rfloor}\frac{d(d-1)z^d}{d!f(z)}\right] \\
&+ O\left((vz)^{1/2}\sum_{d>\lfloor \log n\rfloor} d^2\frac{z^d}{d!f(z)}\right) \\
&= \frac{z^2/2}{f(z)}\frac{z^2e^z}{f(z)}+O\left(\frac{\log^2 v}{vz}\right).
\end{aligned}$$

Likewise, summing the right hand side of (14) over $\ell\neq j$ such that $d(j)\geq 2$ and $d(\ell)\geq 3$,

$$\begin{aligned}
\mathbf{E}(d_3\chi_{\mathbf{a}} \mid A) &= \frac{1}{(2m-1)(2m-3)}\sum_{\ell\neq j:d(\ell)\geq 3,d(j)\geq 2} d(\ell)(d(j))_2 \qquad (19)\\
&= \frac{1}{(2m-1)(2m-3)}\sum_{j:d(j)\geq 2}(2m-v_1-2v_2-d(j)\chi_{\{d(j)\geq 3\}})(d(j))_2,
\end{aligned}$$

within an error $O[m^{-3}\sum_\ell d^2(\ell)\sum_j d^3(j)]$. Then, as twice before,

$$\begin{aligned}
\mathbf{E}(d_3\chi_{\mathbf{a}} \mid \mathbf{v}) &= \frac{1}{(2m)^2}\left[v(2m-v_1)\frac{z^2e^z}{f(z)}-2\frac{vz^2/2}{f(z)}\frac{z^2e^z}{f(z)}\right]\left(1+O\left(\frac{\log^2 v}{vz}\right)\right) \\
&\quad +O(m^{-1}) \\
&= \frac{v^2z^3e^z}{(2m)^2f(z)}+O\left(\frac{\log^2 v}{vz}\right), \qquad (20)
\end{aligned}$$

after using (4).

We use (8), (10), (13), and (17)–(20) to compute $\mathbf{E}[(v_1' - v_1)\chi_{\mathbf{a}} \mid \mathbf{v}]$ Then we combine (8), (11), and (13) for a very simple computation of $\mathbf{E}[(v_1' - v_1)\chi_{\mathbf{b}} \mid \mathbf{v}]$, and complete the proof for v_1 via

$$\mathbf{E}[v_1' - v_1 \mid \mathbf{v}] = \mathbf{E}[(v_1' - v_1)\chi_{\mathbf{a}} \mid \mathbf{v}] + \mathbf{E}[(v_1' - v_1)\chi_{\mathbf{b}} \mid \mathbf{v}].$$

The same holds for computing $\mathbf{E}[v' - v \mid \mathbf{v}]$ and $\mathbf{E}[m' - m \mid \mathbf{v}]$.

We now turn to $v_0' - v_0$. Notice first that

$$\tilde{d} = 0 \Rightarrow v_0' - v_0 = d_1 + 1. \tag{21}$$

Then,

$$\begin{aligned} \mathbf{E}(v_0' - v_0 \mid \mathbf{v}) &\leq \mathbf{E}(v_0' - v_0 \mid \mathbf{v}, \tilde{d} = 0) + v_1 \mathbf{Pr}(\tilde{d} \neq 0) \\ &= \mathbf{E}(v_0' - v_0 \mid \mathbf{v}, \tilde{d} = 0) + O\left(\frac{v_1}{m}\right), \end{aligned}$$

by (9).

It follows from (9) and (21) that

$$\mathbf{E}(v_0' - v_0 \mid \mathbf{v}, \tilde{d} = 0) = (1 + O(1/m))(1 + \mathbf{E}(d_1 \mid \mathbf{v})).$$

Now, by (16),

$$\mathbf{E}(d_1\chi_{\mathbf{a}} \mid A) \leq 1 - \frac{v_1 - 1}{2m - 1} + O\left(\frac{v_1}{m^2}\sum_j (d(j))_2\right).$$

It follows from (7) that **qs** $\sum_j (d(j))_2 = O(m)$ and so

$$\begin{aligned} \mathbf{E}(d_1 \mid \mathbf{v}) &= \mathbf{E}(d_1\chi_{\mathbf{a}} \mid \mathbf{v}) + \mathbf{E}(\chi_{\mathbf{b}} \mid \mathbf{v}) \\ &= 1 - \frac{v_1 - 1}{2m - 1} + O\left(\frac{v_1}{m}\right) + \frac{v_1 - 1}{2m - 1} \\ &= 1 + O\left(\frac{v_1}{m}\right). \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathbf{E}((d_1 - 1)\chi_{\mathbf{a}} \mid A) &= \sum_{k \geq 2} \frac{kv_k}{2m - 1} \frac{(k-1)(v_1 - 1)}{2m - 3} \\ &= \Omega\left(\frac{(v_1 - 1)(2m - v_1)}{m^2}\right) \\ &= \Omega\left(\frac{v_1 - 1}{m}\right), \end{aligned}$$

provided $v_1 \leq m$. □

Corollary 3 *Under the assumptions of Lemma 6*

$$\mathbf{E}(v_1' - v_1) \leq -\min\left\{\frac{z^2}{200}, \frac{1}{20000}\right\} + O\left(\frac{(\log n)^2}{vz}\right).$$

Proof We observe first that (3) and (4) imply

$$\frac{vz}{2m} \leq \frac{f}{e^z - 1} \tag{22}$$

and then that for $z \geq 0$

$$\frac{d}{dz}\left(\frac{z^2 e^z}{(e^z-1)^2}\right) \leq 0. \tag{23}$$

¿From Lemma 6

$$\begin{aligned}
\mathbf{E}(v_1' - v_1 \mid \mathbf{v}) &\leq -1 + \frac{v^2 z^4 e^z}{(2mf)^2} + O\left(\frac{(\log n)^2}{vz}\right) \\
&\leq -1 + \frac{z^2 e^z}{(e^z-1)^2} + O\left(\frac{(\log n)^2}{vz}\right) \qquad \text{by (22)} \qquad (24)\\
&= -\frac{1}{(e^z-1)^2}\sum_{k=4}^{\infty} \frac{2^k - 2 - k(k-1)}{k!} z^k + O\left(\frac{(\log n)^2}{vz}\right) \\
&\leq -\frac{z^2}{200} + O\left(\frac{(\log n)^2}{vz}\right) \qquad (25)
\end{aligned}$$

for $z \leq 1/10$. For $z > 1/10$ we use (23) and (24). □

We now compute the conditional expected changes in $\mathbf{v}$ for Step A2 of KSGREEDY, that is for the case $v_1 = 0$.

Lemma 7 *Assuming that* $\log n = O((vz)^{1/2})$ *and* $v_1 = 0$*,*

$$\begin{aligned}
\mathbf{E}[v_0' - v_0|\mathbf{v}] &= 2 + O\left(\frac{1}{m}\right), \\
\mathbf{E}[v_1' - v_1|\mathbf{v}] &= \frac{v^2 z^4 e^z}{2m^2 f^2} + O\left(\frac{\log^2 v}{vz}\right), \\
\mathbf{E}[v' - v|\mathbf{v}] &= -2 - \frac{v^2 z^4 e^z}{2m^2 f^2} + O\left(\frac{\log^2 v}{vz}\right), \\
\mathbf{E}[m' - m|\mathbf{v}] &= -1 - \frac{vz^2 e^z}{mf} + O\left(\frac{\log^2 v}{vz}\right).
\end{aligned}$$

Proof In Step A2, KSGREEDY chooses a random edge $\{y_1, y_2\}$ of the multigraph $G_{\mathbf{z}}$ and deletes it together with all other edges incident to y_1 or y_2, including loops at y_1, y_2 and other edges that join y_1 and y_2, if any are present.

First consider the case where $y_1 \neq y_2$; we refer to it as **Case (c)**. Introduce the parameters d_2, d_3; they are the total number of neighbors of y_1 or y_2, or both, in $G_{\mathbf{z}}$ that have degree 2 and at least 3 respectively. Let d'' denote the total number of loops and multiple edges incident to y_1 or y_2 *and* of the common neighbors of y_1 and y_2. The new state $\mathbf{v}'$ is then given by

$$\begin{aligned}
v_1' &= d_2 + O(d''), \qquad (26)\\
v' &= v - 2 - d_2 + O(d''), \\
m' &= m - 1 - d_2 - d_3 + O(d'').
\end{aligned}$$

Analogously to (8),

$$\mathbf{E}(d'' \mid \mathbf{v}) = O\left(\frac{1}{m}\right). \tag{27}$$

So it remains to compute $\mathbf{E}(d_r \chi_{\mathbf{c}} \mid \mathbf{v})$, $(r = 2, 3)$. We use again (2) and (15). Conditional on the equivalence class A, the probability that two vertices j_1, j_2 are joined by an edge and that a vertex $i \neq j_1, j_2$ is a neighbor of j_1 or j_2 is (cf. (14))

$$\begin{aligned} \mathbf{Pr}(i; \{j_1, j_2\}) &= \frac{d(i)}{(2m-1)(2m-3)}((d(j_1))_2 d(j_2) + d(j_1)(d(j_2))_2) \\ &+ O(m^{-3} d^2(i) d^4(j_1) d^4(j_2)). \end{aligned} \tag{28}$$

So

$$\begin{aligned} \mathbf{E}(d_2 \chi_{\mathbf{c}} \mid A) &= \frac{1}{m} \sum_{i, \{j_1, j_2\}: d(i)=2, i \neq j_1, j_2, j_1 \neq j_2} P(i; \{j_1, j_2\}) \\ &= \frac{1}{2m^3} v_2 \sum_{j_1} (d(j_1))_2 \sum_{j_2} d(j_2) + O[m^{-3} (\sum_j d^4(j))^2]. \end{aligned} \tag{29}$$

Consequently (cf. (17)), with the help of Lemma 5,

$$\begin{aligned} \mathbf{E}(d_2 \chi_{\mathbf{c}} \mid \mathbf{v}) &= \frac{v^2 z^4 e^z}{2m^2 f^2(z)} \left(1 + O\left(\frac{\log^2 v}{vz}\right)\right) + O(m^{-1}) \\ &= \frac{v^2 z^4 e^z}{2m^2 f^2(z)} + O\left(\frac{\log^2 v}{vz}\right). \end{aligned} \tag{30}$$

Analogously, we also obtain from (28)

$$\begin{aligned} \mathbf{E}(d_3 \chi_{\mathbf{c}} \mid A) &= \frac{2m - 2v_2}{4m^3} \sum_{j_1} (d(j_1))_2 \sum_{j_2} d(j_2) \\ &+ O(m^{-4} \sum_i d^2(i) (\sum_j d^4(j))^2). \end{aligned}$$

Therefore (cf. (20))

$$\begin{aligned} \mathbf{E}(d_3 \chi_{\mathbf{c}} \mid \mathbf{v}) &= \frac{2m - 2\frac{vz^2/2}{f(z)}}{2m^2} \frac{vz^2 e^z}{f(z)} \left(1 + O\left(\frac{\log^2 v}{vz}\right)\right) + O(m^{-1}) \\ &= \frac{v^2 z^3 e^z}{2m^2 f(z)} + O\left(\frac{\log^2 v}{vz}\right). \end{aligned} \tag{31}$$

Here we have used

$$\frac{z(e^z - 1)}{f(z)} = \frac{2m}{v}, \tag{32}$$

when $v_1 = 0$.

The explicit formulas for $\mathbf{E}[(\mathbf{v}' - \mathbf{v})\chi_{\mathbf{c}} \mid \mathbf{v}]$ follow immediately from (26), (27), (30) and (31). They are the right-hand side expressions in the statement of the lemma and it can be easily proved that

$$\mathbf{E}[(\mathbf{v}' - \mathbf{v})\chi_{\mathbf{c}^c} \mid \mathbf{v}] = O(m^{-1}).$$

To handle $v_0' - v_0 - 2$ observe that $z \neq x, y$ becomes isolated after the deletion of x, y only if x, y, z is a triangle in $G(0)$. The expected number of triangles in $G(0)$ is $O(1)$ and so the probability that x, y lie in one is $O(1/m)$. □

To handle the technical problem of the (unlikely) existence of t with $|\mathbf{v}(t+1) - \mathbf{v}(t)| \geq \log n$ ($|\ |$ denoting Euclidean norm) we define the event

$$\mathcal{L}(t) = \{|\mathbf{v}(t+1) - \mathbf{v}(t)| \leq \log n\}.$$

Now $|\mathbf{v}(t+1) - \mathbf{v}(t)| = O(\Delta(G(0)))$ and Lemma 5 implies that $\Delta(G(0)) = o(\log n)$ **qs** and so $\bigcap_t \mathcal{L}(t)$ occurs **qs**.

It is also convenient to introduce a stopping time

$$\mathcal{T}_L = \begin{cases} \min\{t : |\mathbf{v}(t+1) - \mathbf{v}(t)| \geq \log n \text{ or } \mathcal{D}(t) \text{ does not hold} & \text{if such } \tau \text{ exist,} \\ n & \text{otherwise.} \end{cases}$$

Note. We choose $\mathcal{T}_L = n$ when the unlikely events do not occur simply because T_n the total number of steps is at most $[n/2]$. If not otherwise stipulated, n will be the "closing" value for other stopping times in the sequel. It is also convenient to define $\mathbf{v}(t) = \mathbf{v}(T_n)$ for $t \in [T_n, [n/2]]$.

As a bit of notation, whenever we write v, m etc. without specifying an argument, we will mean by default $v(t), m(t)$ etc..

4 Approximation by Differential Equations

Lemma 6 suggests that, for Phase 1, the random sequence $\{\mathbf{v}(t)\}$ must be close to the solution of the following system of differential equations.

$$\begin{aligned} \frac{dv_1}{dt} &= -1 - \frac{v_1}{2m} + \frac{v^2 z^4 e^z}{(2mf)^2} - \frac{v_1 v z^2 e^z}{(2m)^2 f}, \\ \frac{dv}{dt} &= -1 + \frac{v_1}{2m} - \frac{v^2 z^4 e^z}{(2mf)^2}, \\ \frac{dm}{dt} &= -1 - \frac{v z^2 e^z}{2mf}. \end{aligned} \tag{33}$$

We need to integrate these equations subject to the initial conditions

$$v_1(0) = ce^{-c} n, \quad v(0) = p(c)n, \quad m(0) = cn/2, \tag{34}$$

where as usual $f = f(y) = e^y - 1 - y$, $p(y) = e^{-y} f(y)$. The conditions and the equations (3), (4) imply that $z(0) = c$. (Indeed, the degree of a vertex in $G_{n,M}$, $M = cn/2$, is in the limit Poisson with the parameter c.)

Lemma 8 *The solution to equations (33) is:*

$$\begin{aligned} 2m &= \frac{n}{c} z^2, \\ v &= np(z)\beta(z), \\ v_1 &= \frac{n}{c}\left[z^2 - zc\beta(z)(1 - e^{-z})\right], \\ t &= \frac{n}{c}\left[c(1 - \beta(z)) - \frac{1}{2}\log^2 \beta(z)\right], \end{aligned} \tag{35}$$

where

$$\beta e^{c\beta} = e^z. \tag{36}$$

So, the solution is given in a parametric form, as functions of z, the hidden parameter.

Proof First observe that

$$\begin{aligned}
\mathbf{E}[Z(z)] &= \frac{zf'(z)}{f(z)},\\
\mathbf{Var}[Z(z)] &= E[Z(z)(Z(z)-1)] + \mathbf{E}[Z(z)] - \mathbf{E}^2[Z(z)]\\
&= \frac{z^2e^z}{f(z)} + \frac{zf'(z)}{f(z)} - \left(\frac{zf'(z)}{f(z)}\right)^2,
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
\left(\frac{zf'(z)}{f(z)}\right)' &= \frac{f'(z)}{f(z)} + \frac{zf''}{f(z)} - z\left(\frac{f'(z)}{f(z)}\right)^2\\
&= \frac{1}{z}\mathbf{Var}[Z(z)].
\end{aligned} \tag{38}$$

Explicitly

$$\mathbf{Var}[Z(z)] = \frac{z(e^z-1)^2 - z^3e^z}{f^2(z)}. \tag{39}$$

Consequently

$$\frac{d}{dt}\left(\frac{2m-v_1}{v}\right) = \frac{1}{z}\mathbf{Var}[Z(z)]\frac{dz}{dt}. \tag{40}$$

On the other hand, using the differential equations (33),

$$\begin{aligned}
\frac{d}{dt}\left(\frac{2m-v_1}{v}\right) &= -\frac{2m-v_1}{v^2}\left(-1 + \frac{v_1}{2m} - \frac{v^2z^4e^z}{(2mf)^2}\right)\\
&\quad + \frac{1}{v}\left[2\left(-1 - \frac{vz^2e^z}{2mf}\right) + 1 + \frac{v_1}{2m} - \frac{v^2z^4e^z}{(2mf)^2} + \frac{v_1vz^2e^z}{(2m)^2f}\right]\\
&= \frac{1}{2m}\left[\left(\frac{2m-v_1}{v}\right)^2 - \frac{2m-v_1}{v} - \frac{z^2e^z}{f}\right]\\
&\quad + (2m-v_1)\frac{z^4e^z}{(2mf)^2} + \left(-\frac{z^2e^z}{2mf} + \frac{v_1z^2e^z}{(2m)^2f}\right) - \frac{vz^4e^z}{(2mf)^2}\\
&= -\frac{1}{2m}\mathbf{Var}[Z] + \frac{zf'}{f}\frac{vz^4e^z}{(2mf)^2} - \frac{zf'}{f}\frac{vz^2e^z}{(2m)^2f} - \frac{vz^4e^z}{(2mf)^2}\\
&= -\frac{1}{2m}\mathbf{Var}[Z] + \frac{vz^2e^z}{(2m)^2f}\frac{z^3e^z - z(e^z-1)^2}{f^2}\\
&= -\frac{1}{2m}\mathbf{Var}[Z] - \frac{vz^2e^z}{(2m)^2f}\mathbf{Var}[Z].
\end{aligned}$$

Comparing this with (40) we obtain

$$\begin{aligned}
\frac{1}{z}\frac{dz}{dt} &= -\frac{1}{2m}\left(1 + \frac{vz^2e^z}{2mf}\right)\\
&= \frac{1}{2m}\frac{dm}{dt},
\end{aligned} \tag{41}$$

see the third equation in (33). Integrating,

$$\frac{z^2}{2m} \equiv \text{constant},$$

so by (34)

$$\frac{z^2 n}{2m} = c. \tag{42}$$

Next, we rewrite the second equation in (33):

$$\frac{dv}{dt} = -\frac{v}{2m}\frac{zf'}{f} - \frac{v^2 z^4 e^z}{(2mf)^2}.$$

Here, using the third equation in (33) and (41),

$$\begin{aligned}\frac{dv}{dt} &= \frac{dv}{dm}\frac{dm}{dt} \\ &= \frac{dv}{dm}\left(-1 - \frac{vz^2e^z}{2mf}\right) \\ &= \frac{z}{2m}\frac{dv}{dz}\left(-1 - \frac{vz^2e^z}{2mf}\right).\end{aligned}$$

So the equation for v becomes

$$\frac{dv}{dz}\left(-1 - \frac{vz^2e^z}{2mf}\right) = -\frac{v}{f}f' - \left(\frac{v}{f}\right)^2 \frac{z^3e^z}{2m}. \tag{43}$$

The form of this equation suggests a substitution

$$v = np(z)\beta(z), \quad p(z) = e^{-z}f(z). \tag{44}$$

Then (see (34))

$$\beta(c) = 1. \tag{45}$$

Plugging this formula into (43) and using (42) and $p'(z) = ze^{-z}$, we obtain, after cancelling $-\beta^2(e^{-z}z^3n/2m)$ on both sides,

$$-e^{-z}z\beta - p\frac{d\beta}{dz} - cp\beta\frac{d\beta}{dz} = -\beta(1 - e^{-z}).$$

Since

$$1 - e^{-z} - ze^{-z} = e^{-z}(e^z - 1 - z) = p(z),$$

the equation for β simplifies to

$$\frac{1}{\beta}\frac{d\beta}{dz} + c\frac{d\beta}{dz} = 1.$$

Integrating and using (45) we obtain

$$\beta e^{c\beta} = e^z. \tag{46}$$

Next, using

$$\frac{zf'(z)}{f(z)} = \frac{2m - v_1}{v},$$

we obtain from (42), (44),

$$v_1 = \frac{n}{c}z^2 - nz(1-e^{-z})\beta(z).$$

Finally, from the third equation in (33), and (44), (41), (46),

$$\begin{aligned} dt &= -\frac{dm}{1+\frac{vz^2e^z}{2mf}} \\ &= -\frac{(2m/z)dz}{1+\frac{n\beta z^2}{2m}} \\ &= -\frac{n}{c}\frac{z\,dz}{1+c\beta}. \end{aligned} \tag{47}$$

Here, denoting $y = c\beta$, and using (46),

$$dz = d(-\log c + \log y + y) = \frac{1+y}{y}dy,$$

so that

$$dt = -\frac{n}{c}\frac{(-\log c + \log y + y)}{y}dy.$$

Integrating, and using the initial condition $y(0) = c\beta(0) = c$, we arrive at

$$t = \frac{n}{c}\left[c(1-\beta) - \frac{1}{2}\log^2\beta\right].$$

□

Let us see for which values of c the solution $\mathbf{v}(t)$ reaches a point such that $v_1 = 0$, but v, z are still positive. At this moment, call it t^*, by the formula for v_1 we can write

$$\begin{aligned} z^* &= \gamma^* - \gamma_*, \\ \gamma^* &:= c\beta, \\ \gamma_* &:= c\beta e^{-z^*}. \end{aligned} \tag{48}$$

Observe that, from (36),

$$\gamma_* = ce^{-\gamma^*}, \quad \gamma^* = ce^{-\gamma_*}. \tag{49}$$

Indeed

$$\begin{aligned} ce^{-\gamma^*} &= ce^{-c\beta} = c\beta e^{-z^*} = \gamma_*, \\ ce^{-\gamma_*} &= ce^{z^*-c\beta} = c\beta = \gamma^*. \end{aligned}$$

Since $\gamma_* < \gamma^*$, the equation (49) can be satisfied iff $c > e$. Indeed, $\gamma^* e^{-\gamma^*} = \gamma_* e^{-\gamma_*}$ which implies $\gamma_* < 1 < \gamma^*$ and $\gamma^*\gamma_* < 1$. Thinking of γ^*, c as functions of γ_* we get $dc/d\gamma_* = (1-\gamma^*\gamma_*)e^{\gamma^*}/(1-\gamma^*) < 0$. $c > e$ then follows from $c(1) = e$. See also Lemmas 9 and 10 below. The formula for t^* determines the total number of matched pairs in Phase 1, and using γ_*, γ^*, we can write it as follows:

$$t^* = \frac{n}{c}\left[c - \gamma^* - \frac{1}{2}\gamma_*^2\right]. \tag{50}$$

In this case (see 49) γ_* and γ^* are the smallest root and the largest root among the three roots of

$$x = c\exp(-ce^{-x}). \tag{51}$$

The middle root, denote it γ, is the only root of

$$x = ce^{-x}. \tag{52}$$

For $c < e$, $v_1(t)$ remains positive until $z(t)$ (whence $v(t)$ and $m(t)$) reaches the zero value. According to (36), the terminal value β satisfies $\beta = e^{-c\beta}$, so that

$$\beta = \frac{\gamma}{c}.$$

So, by (35), the likely number of matched pairs in Phase 1 (whence the likely maximum matching number) is given by

$$t^* = \frac{n}{c}\left(c - \gamma - \frac{1}{2}\gamma^2\right). \tag{53}$$

If $c > e$ then Phase 1 ends with v of order n, so Phase 2 should be expected to deliver many more matched pairs. Let us "derive" the differential equations that should provide an accurate approximation of the actual Markov chain. To this end, we observe first that, for the moments t such that $v_1(t) = 0$, we apparently have to use the equations

$$\begin{aligned} \frac{dv_1}{dt} &= \frac{v^2 z^4 e^z}{2m^2 f^2(z)}, \\ \frac{dv}{dt} &= -2 - \frac{v^2 z^4 e^z}{2m^2 f^2(z)}, \\ \frac{dm}{dt} &= -1 - \frac{v z^2 e^z}{m f(z)}, \end{aligned} \tag{54}$$

suggested by Lemma 6. Since $dv_1/dt > 0$, the representing point $\mathbf{v}(t)$ is pushed back into the region $\{\mathbf{v} : v_1 > 0\}$, so instantly we have to switch to the equations (33) with v_1 set equal 0:

$$\begin{aligned} \frac{dv_1}{dt} &= -1 + \frac{v^2 z^4 e^z}{(2mf(z))^2}, \\ \frac{dv}{dt} &= -1 - \frac{v^2 z^4 e^z}{(2mf(z))^2}, \\ \frac{dm}{dt} &= -1 - \frac{v z^2 e^z}{2mf(z)}. \end{aligned} \tag{55}$$

Here, by (32),

$$\begin{aligned} \frac{dv_1}{dt} &= -1 + \frac{z^2 e^z}{(e^z - 1)^2} \\ &= \left(\frac{z/2}{\sinh z/2}\right)^2 - 1 < 0, \end{aligned} \tag{56}$$

which means that $\mathbf{v}(t)$ moves back toward the boundary set $B = \{\mathbf{v} : v_1 = 0\}$. These instantaneously alternating attractions to, and repulsions from B strongly suggest that a proper system of differential equations is obtained by mixing the right hand expressions of (54) and (55) with "weights" (relative time frequences) $1 - \omega(t)$, $\omega(t)$. Here $\omega(t)$ has to be such that the resulting equations admit a solution $\mathbf{v}(t)$ with $v_1(t) \equiv 0$, that is $dv_1/dt \equiv 0$. (In the language of dynamic systems control theory, we have encountered here a so-called sliding mode (trajectory)). Explicitly, see the first equations in (54), (55),

$$\omega(t)\left[-1+\frac{v^2z^4e^z}{(2mf(z))^2}\right]+(1-\omega(t))\frac{v^2z^4e^z}{2m^2f^2(z)}\equiv 0,$$

so that

$$\omega(t)=\frac{\frac{v^2z^4e^z}{2m^2f^2(z)}}{1+\frac{v^2z^4e^z}{4m^2f^2(z)}}. \tag{57}$$

Using $\omega(t)$ to mix the remaining equations in (54) with their counterparts in (55), we obtain

$$\begin{aligned}\frac{dv}{dt} &= -2, \\ \frac{dm}{dt} &= -1-\frac{\frac{vz^2e^z}{mf(z)}}{1+\frac{v^2z^4e^z}{4m^2f^2(z)}}.\end{aligned} \tag{58}$$

The first equation is strikingly simple; it means that, on average, we lose —one way or another— exactly two heavy vertices (i.e. of degree two or more) per step of Phase 2. Since at each step two vertices get matched, we can see that the total number of matched pairs delivered by Phase 2 must be asymptotic to $v(t^*)/2$. That is, almost all vertices present at the moment t^* get matched. Combining the second equation in (35) with (49)-(50), we obtain that for $c>e$ the maximum matching number of $G_{n,cn/2}$ is asymptotic to

$$n\left(1-\frac{\gamma^*+\gamma_*+\gamma^*\gamma_*}{2c}\right).$$

Of course, this argument is too superficial! The actual analysis of Phase 2 is much more technical, and revealing. Curiously our proof does not require justification of the first equation in (58), nor have we tried to prove it as the limit property, using our analysis.

Let us use the equations (58) to determine the parametric solution analogous to (35). We have (cf. Phase 1 computation): by (40),

$$\begin{aligned}\frac{1}{z}\mathbf{Var}(Z)\frac{dz}{dt} &= \frac{d}{dt}\left(\frac{2m}{v}\right)=-\frac{2m}{v^2}\cdot(-2)+\frac{2}{v}\left(-1-\frac{\frac{vz^2e^z}{mf(z)}}{1+\frac{v^2z^4e^z}{4m^2f^2(z)}}\right)\\ &= \frac{1}{m}\left[\left(\frac{2m}{v}\right)^2-\frac{2m}{v}-\frac{z^2e^z}{f(z)}\right]\\ &+ \frac{z^2e^z}{mf(z)}-\frac{\frac{2z^2e^z}{mf(z)}}{1+\frac{v^2z^4e^z}{4m^2f(z)^2}}\\ &= -\frac{1}{m}\mathbf{Var}(Z)-\frac{ze^z}{mf(z)}\cdot\frac{z(e^z-1)^2-z^3e^z}{z^2e^z+(e^z-1)^2}\\ &= -\frac{1}{m}\mathbf{Var}(Z)-\frac{ze^z}{mf(z)}\cdot\frac{f^2(z)\mathbf{Var}(Z)}{z^2e^z+(e^z-1)^2}.\end{aligned}$$

Cancelling $\mathbf{Var}(Z)$, we get

$$\frac{1}{z}\frac{dz}{dt}=-\frac{1}{m}\left(1+\frac{ze^zf(z)}{z^2e^z+(e^z-1)^2}\right).$$

Invoking the second equation in (58), we exclude t and obtain

$$\frac{dz}{z} \cdot \frac{z^2 e^z + (e^z - 1)^2 + 2z(e^z - 1)e^z}{z^2 e^z + (e^z - 1)^2 + ze^z(e^z - 1 - z)} = \frac{dm}{m},$$

or, after simple algebra,

$$\left(\frac{1}{z} + \frac{e^z}{e^z - 1} - \frac{ze^z}{e^z - 1 + ze^z} \right) dz = \frac{dm}{m}.$$

Integrating from $z^* := z(t^*)$, $m^* := m(t^*)$ to z, m, we obtain

$$\log \left[\frac{z(e^z - 1)}{m} \div \frac{z^*(e^{z^*} - 1)}{m^*} \right] = \int_{z^*}^{z} \frac{\xi e^\xi}{e^\xi (1 + \xi) - 1} \, d\xi. \tag{59}$$

The last relation and (32) provide the desired parametric description of the sliding trajectory $\{\mathbf{v}(t)\}_{t \geq t^*}$.

Note. We should technically use notation for the deterministic trajectory that differs from that for the random process, but this would have been cumbersome. However from now on we will refer to the latter as $\tilde{\mathbf{v}}(t)$ and $\tilde{\mathbf{v}}^*$ will stand for $\tilde{\mathbf{v}}(t^*)$.

The above analysis provides an insight into what is going on. We now have to carefully justify the fact that our process is likely to follow the trajectories computed above.

5 Analysis of Phase 1

We wish to show that random variables, m, v_1 and v, tend to stay close to the solution of differential equations (33).

As we showed in Section 4, along the trajectory (35) the following four functions remain constant.

$$\begin{aligned}
J_1(\mathbf{v}) &= \frac{m}{nz^2}, \\
J_2(\mathbf{v}) &= \frac{v}{np(z)\beta(z)}, \\
J_3(\mathbf{v}) &= \frac{v_1}{nh(z)}, \\
J_4(\mathbf{v}, t) &= \frac{t}{n} - g(z),
\end{aligned}$$

where

$$\begin{aligned}
h(z) &= z^2[1 - c\beta(z)(1 - e^{-z})/z], \\
g(z) &= 1 - \beta(z) - \frac{1}{2c} \log^2 \beta(z), \\
\beta e^{c\beta} &= e^z.
\end{aligned}$$

Initially **whp** $z \approx c$ and we expect $\mathbf{v}(t)$ to follow the trajectory (35) and so we expect $\mathbf{v}(t)$ to lie well within the set V of $\mathbf{v}$ satisfying

$$\begin{aligned}
\tfrac{1}{2}\tfrac{nz^2}{2c} &\leq m \leq 2\tfrac{nz^2}{2c}, \\
\tfrac{1}{2}np(z)\beta(z) &\leq v \leq 2np(z)\beta(z), \\
\tfrac{1}{2}\tfrac{n}{c}h(z) &\leq v_1 \leq 2\tfrac{n}{c}h(z).
\end{aligned}$$

Before proving that $\mathbf{v}(t)$ stays within these bounds **whp**, let us look at the behavior of the deterministic approximation $\tilde{\mathbf{v}}(t)$ as t increases. Let z^* be the largest nonnegative root of $h(z) = 0$, the same z^* as in (48). [Observe that $h(0) = 0$.]

Lemma 9 *For $c < e$, $h(z) > 0$ for $z > 0$ and so $z^* = 0$. Also $g'(0) = 0$.*

Proof If $h(z) = 0$ then $1 - c\beta(1 - e^{-z})/z = 0$. Since $\beta e^{c\beta} = e^z$ we get $r(\beta) = 0$, where

$$r(y) := \exp(-ce^{-cy}) - y.$$

Clearly the solution to $\beta = e^{-c\beta}$ is a root of $r(\beta) = 0$. This corresponds to $e^z = 1$ i.e. $z = 0$. For a positive root to exist we must have some root of $r'(\beta) = 0$. This is not possible as $c^2e^{-c\beta}\exp(-ce^{-c\beta}) < 1$. To see this let $x = e^{-c\beta}$. By simple calculus $c^2xe^{-cx} \leq \frac{c}{e} < 1$. Finally

$$g'(z) = -\frac{z\beta'}{c\beta},$$

yielding $g'(0) = 0$. □

Lemma 10 *If $c > e$ then $z^* > 0$.*

Proof Clearly $r(1) < 0$ and $r(\frac{\ln c}{c}) = \frac{1}{e} - \frac{\ln c}{c} > 0$. (See Karp and Sipser). A root must exist for some $\beta \in (\frac{\ln c}{c}, 1)$. This corresponds to $z^* \in (\ln\ln c, c)$. □

We will be able to show that $\tilde{\mathbf{v}}(t)$ provides **whp** a sharp approximation for the actual process at least as long as the hidden parameter z exceeds $z^* + n^{-a}$ where $a > 0$ is a constant, such that $a < 1/6$ if $c < e$, and $a < 1/4$ if $c > e$. When $c < e$ we have $m, v, v_1 = \Theta(nz^2)$ as $z \to z^* = 0$. For $c > e$ we will have $m, v = \Theta(n)$ as $z \to z^*$. The variable v_1 is another story, and to estimate it we will need to use $h(z) = h(z^*) + (z - z^*)h'(z^*) + O((z - z^*)^2h''(z^*))$. Using $\beta' = \frac{\beta}{1+c\beta}$ we have

$$h'(z) = 2z - c\beta\frac{e^z - 1}{e^z} - cz\frac{\beta}{1 + c\beta}\frac{e^z - 1}{e^z} - cz\beta\frac{1}{e^z}.$$

At $z = z^*$ we can use $\frac{z^*e^{z^*}}{e^{z^*}-1} = c\beta$ to get

$$h'(z^*) = \frac{z^*[(e^{z^*} - 1)^2 - z^{*2}e^{z^*}]}{(e^{z^*} - 1)(z^*e^{z^*} + e^{z^*} - 1)} > 0. \tag{60}$$

To see that this is strictly positive just compare the Taylor series of $e^{z^*} - 1$ and $z^*e^{z^*/2}$. Therefore

$$v_1 = \Theta(n(z - z^*)). \tag{61}$$

Note that $h'(z) > 0$ is associated with $\frac{dv_1}{dt} < 0$, at least along the trajectory (35), (cf. (47)).

Turn now to the random sequence $\{\mathbf{v}(t)\}$. Set $a \in (0, 1/6)$, $\alpha_0 = (1 - 6a)/2$ if $c < e$, and $a \in (0, 1/4)$, $\alpha_0 = (1 - 4a)/2$ if $c > e$.

Let

$$W = \{\mathbf{v} \in V : \ z \geq z^* + n^{-a}\},$$

and introduce

$$\mathcal{T}_W = \begin{cases} \min\{t \leq \mathcal{T}_L : \mathbf{v}(t) \notin W\} & \text{if such } \tau \text{ exist,} \\ n & \text{otherwise.} \end{cases}$$

Lemma 11 *Assume* $\mathbf{v}(0) \in W$. *If* $0 < \alpha < \alpha_0$ *then* **qs**,

$$\max_{t \leq \mathcal{T}_W} |J_i(\mathbf{v}(t)) - J_i(\mathbf{v}(0))| \leq n^{-\alpha}, \quad i = 1, 2, 3. \tag{62}$$

Consequently, if $\mathbf{v}(t)$ *ever leaves* W, *it happens* **qs** *only because* $z(t)$ *falls below* $z^* + n^{-a}$.

Proof To prove this we will use a technique from Pittel, Spencer and Wormald [20]. It is based on showing that $\{Q_i(\mathbf{v}(t))\}_{t \geq 0}$, where

$$Q_i(\mathbf{v}) = \exp\{L(J_i(\mathbf{v}) - J_i(\mathbf{v}(0)))\}, \quad i = 1, 2, 3,$$

is almost a supermartingale for $L = n^{\alpha'}$, $\alpha' \in (\alpha, \alpha_0)$, essentially because $J_i(\mathbf{v}(t))$ is constant along the deterministic trajectory.

Suppose $c < e$. We consider only Q_2 since the two other cases are very similar.

We define $Q(t) = Q_2(\mathbf{v}(t))$ if $t < \mathcal{T}_W$, and $Q(t) = 0$ if $t \geq \mathcal{T}_W$. We also let $J(t) = J_2(\mathbf{v}(t))$.

For $t - 1 \geq \mathcal{T}_W$, we obviously have $Q(t) = Q(t-1) = 0$. For $t - 1 < \mathcal{T}_W$ we can write

$$\mathbf{E}(Q(t) \mid \{\mathbf{v}(s)\}_{s<t}) \leq Q(t-1)\mathbf{E}\left\{ \mathbf{1}_{\{\mathcal{L}(t)\}} \exp\left[L(J(t) - J(t-1))\right] \middle| \mathbf{v}(t-1) \right\}. \tag{63}$$

Since $\mathbf{v}(t-1) \in W$, each of $m(t-1)$ and $v(t-1)$ is of order $n^{1-2a}/\log n$ at least. The same holds then for $\mathbf{v}(t) \in B(\mathbf{v}(t-1), \log n) = \{\mathbf{v} : |\mathbf{v} - \mathbf{v}(t-1)| \leq \log n\}$. Consequently $v(t)z(t)$ is of order $n^{1-3a}/\log n$ at least. Moreover, it can be easily verified that, uniformly for such $\mathbf{v}$ and $i = 1, 2, 3$, $x, y = v_0, v_1, v, m$,

$$\frac{\partial J}{\partial x} = O\left(\frac{1}{vz}\right), \tag{64}$$

$$\frac{\partial^2 J}{\partial x \partial y} = O\left(\frac{1}{v^2 z^2}\right). \tag{65}$$

So, assuming $\mathbf{v}(t) \in B(\mathbf{v}(t-1), \log n)$ and expanding the exponential function,

$$\exp\{L(J(t) - J(t-1))\} = [1 + L\nabla J(t)^*(\mathbf{v}(t) - \mathbf{v}(t-1)) + O(L^2(\log n)^2/(vz)^2)], \tag{66}$$

since

$$L \log n = o(vz). \tag{67}$$

Consequently, equation (63) becomes

$$\begin{aligned} \mathbf{E}(Q(t)|\{\mathbf{v}(s)\}_{s<t}) \leq\ & Q(t-1)\{1 + L\nabla J(t)^*\mathbf{E}[\mathbf{v}(t) - \mathbf{v}(t-1)|\mathbf{v}(t-1)]\} \\ & + O(Q(t-1)L^2(\log n)^2/(vz)^2). \end{aligned} \tag{68}$$

Here, denoting the vector-valued right-hand side of (33) by $\mathbf{F}(\mathbf{v})$, and using Lemma 6,

$$\begin{aligned} \nabla J(t)^*\mathbf{E}[\mathbf{v}(t) - \mathbf{v}(t-1)|\mathbf{v}(t-1)] &= \nabla J(t-1)^*\left[\mathbf{F}(\mathbf{v}(t-1)) + O((\log n)^2/vz)\right] \\ &= O\left(\|\nabla J(t-1)\|(\log n)^2/vz\right) \\ &= O\left(\frac{(\log n)^2}{(vz)^2}\right). \end{aligned} \tag{69}$$

($\nabla J(\mathbf{v}) \perp \mathbf{F}(\mathbf{v})$ since $J(\mathbf{v})$ is constant along the trajectory (35) of $d\mathbf{v}/dt = \mathbf{F}(\mathbf{v})$!)

Therefore, for $t-1 < \mathcal{T}_W$ and hence for all t,

$$\mathbf{E}(Q(t) \mid \{\mathbf{v}(s)\}_{s<t}) \le Q(t-1)\left(1 + O(L^2(\log n)^2/(vz)^2)\right).$$

So we can find $1 < \omega < 2 - 6a - 2\alpha'$ such that the random sequence

$$\{R(t)\} := \{(1+n^{-\omega})^{-t} Q(t)\}$$

is a *supermartingale*.

Introduce a stopping time

$$\mathcal{T}'_W = \begin{cases} \min\{t < \mathcal{T}_W : J(t) - J(0) > n^{-\alpha}/2\}, & \text{if such } t \text{ exist,} \\ \mathcal{T}_W, & \text{otherwise.} \end{cases}$$

Now, applying the Optional Sampling Theorem (Durrett [7]) to the supermartingale $\{R(t)\}$ and the stopping time $\mathcal{T}'_W$, and going back to $\{Q(t)\}$, we get

$$\begin{aligned} \mathbf{E}[Q(\mathcal{T}_W)] &\le (1+n^{-\omega})^n \cdot \mathbf{E}[Q(0)] \\ &= (1+n^{-\omega})^n \\ &= O(1), \text{ as } n \to \infty. \end{aligned} \tag{70}$$

Since obviously

$$\mathbf{E}[Q(\mathcal{T}'_W)] \ge e^{n^{\alpha'-\alpha}/2} \cdot \mathbf{Pr}\{\mathcal{T}'_W < \mathcal{T}_W\},$$

we have then

$$\begin{aligned} \mathbf{Pr}\{\max_{t<\mathcal{T}_W} [J(t) - J(0)] > n^{-\alpha}/2\} &= \mathbf{Pr}\{\mathcal{T}'_W < \mathcal{T}_W\} \\ &= O(e^{-n^{\alpha'-\alpha}/2}). \end{aligned}$$

Analogously,

$$\mathbf{Pr}\{\min_{t<\mathcal{T}_W} [J(t) - J(0)] < -n^{-\alpha}/2\} = O(e^{-n^{\alpha'-\alpha}/2}).$$

So **qs**

$$\max_{t<\mathcal{T}_W} |J(t) - J(0)| \le n^{-\alpha}/2.$$

It remains to notice that on $\mathcal{T}_W < \mathcal{T}_L$,

$$J(\mathcal{T}_W) - J(\mathcal{T}_W - 1) = O(n^{-(1-3a)+o(1)}) = o(n^{-\alpha}).$$

The case $c > e$ is essentially similar. The real difference is that the first order derivatives and the second order derivatives of J_3 in the ball $B(\mathbf{v}(t-1), \log n)$ are $O(n^{-1+2a})$, $O(n^{-2+4a})$, respectively. (This comes from $h(z^*) = 0$, $h'(z^*) > 0$.) For each of J_1, J_2, the orders are even better, $O(n^{-1})$ and $O(n^{-2})$. These better estimates result in (62) for $\alpha < \alpha_0 = (1-4a)/2$. □

We now turn to J_4 and prove that **qs** it also stays close to its initial value.

Lemma 12 *Assume* $\mathbf{v}(0) \in V$. *If* $0 < \alpha < \alpha_0$, *then* **qs**

$$|J_4(\mathbf{v}, t) - J_4(\mathbf{v}(0))| \le n^{-\alpha}, \quad (t \le \mathcal{T}_W).$$

Proof We consider $c < e$ only. We start by choosing $\alpha' \in (\alpha, \alpha_0)$, retaining $L = n^{\alpha'}$ and defining

$$Q_4(t) = \exp\{LJ_4(\mathbf{v}, t)\} = \exp\left\{L\left[\frac{t}{n} - g(z(t))\right]\right\},$$

if $t < \mathcal{T}_W$, and setting $Q_4(t) = 0$ for $t \geq \mathcal{T}_W$. Then, for $t < \mathcal{T}_W$,

$$\begin{aligned}
\frac{Q_4(t)}{Q_4(t-1)} &= \exp\left\{L\left(\frac{1}{n} - [g(z(t)) - g(z(t-1))]\right)\right\} \\
&= \exp\left\{L\left(\frac{1}{n} - \frac{dg}{dz}(z(t) - z(t-1)) + O((z(t) - z(t-1))^2)\right)\right\} \\
&= \exp\left\{L\left(\frac{1}{n} - \frac{dg}{dz}(z(t) - z(t-1)) + O\left(\frac{(\log n)^2}{v^2}\right)\right)\right\} \\
&= 1 + L\left(\frac{1}{n} - \frac{dg}{dz}(z(t) - z(t-1)) + O\left(\frac{(\log n)^2}{v^2}\right)\right) + O\left(\frac{L^2(\log n)^2}{v^2}\right),
\end{aligned} \tag{71}$$

with dg/dz computed at $z = z(t-1)$. Here the estimate (71) follows from $v = \Omega(nz^2) = \Omega(n^{1-2a})$,

$$\begin{aligned}
\frac{dg}{dz} &= -\frac{z}{c(1+c\beta)} \\
&= O(1),
\end{aligned} \tag{72}$$

and

$$|z(t) - z(t-1)| = O(\log n / v(t-1)), \tag{73}$$

as $|\mathbf{v}(t) - \mathbf{v}(t-1)| \leq \log n$.

Hence, analogously to (63), for $t - 1 < \mathcal{T}_W$,

$$\begin{aligned}
&\mathbf{E}[Q_4(t) | \{\mathbf{v}(s)\}_{s<t}] \\
\leq\ & Q_4(t-1)\left[1 + L\mathbf{E}\left(\frac{1}{n} - \frac{dg}{dz}(z(t) - z(t-1)) | \mathbf{v}(t-1)\right) + O\left(\frac{L^2(\log n)^2}{v^2}\right)\right].
\end{aligned} \tag{74}$$

We proceed to find an estimate for the expectation in the RHS of (74). Write $\mathbf{v} = \mathbf{v}(t-1), \mathbf{v}' = \mathbf{v}(t), z = z(t-1)$ and $z' = z(t)$ and

$$\frac{zf'(z)}{f(z)} = \frac{2m - v_1}{v}, \quad \frac{z'f'(z')}{f(z')} = \frac{2m' - v_1'}{v'}.$$

It follows from (38) and (73) that

$$\frac{z'f'(z')}{f(z')} - \frac{zf'(z)}{f(z)} = \frac{1}{z}\mathbf{Var}(Z(z))(z' - z) + O\left(\frac{(\log n)^2}{v^2}\right).$$

On the other hand

$$\frac{2m' - v_1'}{v'} - \frac{2m - v_1}{v} = \frac{2(m' - m)}{v} - \frac{v_1' - v_1}{v} - (v' - v)\frac{2m - v_1}{v^2} + O\left(\frac{(\log n)^2}{v^2}\right).$$

Taking expectations conditioned on $\mathbf{v}$, using Lemma 6 and then simplifying, we obtain

$$\mathbf{E}(z' - z|\mathbf{v})\frac{1}{z}\mathbf{Var}[Z(z)] = -\frac{1}{2m}\mathbf{Var}[Z(z)] - \frac{vz^2e^z}{(2m)^2 f(z)}\mathbf{Var}[Z(z)] + O\left(\frac{(\log n)^2}{v^2}\right).$$

Thus

$$\mathbf{E}(z' - z|\mathbf{v}) = -\frac{z}{2m}\left(1 + \frac{vz^2e^z}{2mf(z)}\right) + O\left(\frac{(\log n)^2}{v^2}\right). \tag{75}$$

Now along the trajectory (35) (see (47))

$$\frac{dz}{dt} = -\frac{c(1+c\beta)}{nz}$$

and so using (72) we expect to find that

$$\frac{dg}{dz}\mathbf{E}(z(t) - z(t-1)) \mid \mathbf{v}(t-1)) \approx \frac{1}{n}.$$

Indeed, by (75),

$$\frac{dg}{dz}\mathbf{E}[z(t) - z(t-1)|\mathbf{v}(t-1)] = \frac{1}{n} + O\left(\frac{x}{n} + \frac{(\log n)^2}{v^2}\right), \tag{76}$$

provided that $|J_i(\mathbf{v}(t)) - J_i(\mathbf{v}(0))| \leq x$, $i = 1, 2, 3$. We let $x = n^{-\tilde{\alpha}}$ for some $\tilde{\alpha} \in (\alpha', \alpha_0)$. Applying (76) in (74), we obtain: for $t - 1 < \mathcal{T}_W$,

$$\begin{aligned}
\mathbf{E}[Q_4(t)|\{\mathbf{v}(s)\}_{s<t}] &\leq Q_4(t-1)\left[1 + O\left(\frac{L}{n^{1+\tilde{\alpha}}} + \frac{L^2(\log n)^2}{n^{2(1-2a)}}\right)\right] \\
&= Q_4(t-1)\left[1 + O\left(\frac{1}{n^{\omega}}\right)\right];
\end{aligned}$$

here

$$\omega \stackrel{def}{=} 1 + \min\{\tilde{\alpha} - \alpha', 1 - 4a - 2\alpha' + o(1)\} > 1.$$

The last estimate certainly holds for $t - 1 \geq \mathcal{T}_W$ because $Q_4(t) = Q_4(t-1) = 0$ in this case. Introduce the stopping time

$$\mathcal{T}_{W^*} = \begin{cases} \min\{t < \mathcal{T}_W : \max_{1\leq i\leq 3} |J_i(\mathbf{v}(t)) - J_i(\mathbf{v}(0))| > n^{-\tilde{\alpha}} & \text{if such } t \text{ exists,} \\ \mathcal{T}_W, & \text{otherwise.} \end{cases}.$$

Define

$$\mathcal{T}'' = \begin{cases} \min\{t < \mathcal{T}_{W^*} : |J_4(\mathbf{v}(t), t)| > n^{-\alpha}/2\}, & \text{if such } t \text{ exists,} \\ \mathcal{T}_{W^*}, & \text{otherwise.} \end{cases}$$

The above estimate for $\mathbf{E}[Q_4(t) \mid \{\mathbf{v}(s)\}_{s<t}]$ means that $\{(1+n^{-\omega})^{-t}Q_4(t)\}$ is a supermartingale for $t \leq \mathcal{T}_W$. Then (by the optional stopping theorem)

$$\begin{aligned}
\mathbf{E}[Q_4(\mathcal{T}'')] &\leq (1 + n^{-\omega})^n \mathbf{E}[Q_4(0)] \\
&= (1 + n^{-\omega})^n = O(1),
\end{aligned}$$

whence

$$\mathbf{Pr}(\mathcal{T}'' < \mathcal{T}_{W^*}) = O\left(e^{-n^{\alpha'-\alpha}/2}\right).$$

Therefore **qs**, for $t < \mathcal{T}_W$, the process $\{\mathbf{v}(t)\}$ satisfies

$$|J_i(\mathbf{v}(t)) - J_i(\mathbf{v}(0))| \leq n^{-\tilde{\alpha}}, \quad i = 1,2,3, \tag{77}$$

and

$$|J_4(\mathbf{v}(t), t)| = \left| g(z(t)) - \frac{t}{n} \right| < n^{-\alpha}/2. \tag{78}$$

(To be sure, we have proved only the upper bound for $g(z(t)) - t/n$, but the lower bound is handled in exactly the same way.) It follows then, as in the proof of Lemma 11, that

$$\left| g(z(\mathcal{T}_W)) - \frac{\mathcal{T}_W}{n} \right| \leq n^{-\alpha}.$$

□

5.1 The subcritical case: $c < e$

We set $a = 0.1, \alpha = 0.2 - \epsilon$, ($\epsilon < .1$, say), in Lemma 11. Then $z(\mathcal{T}_W - 1) \geq n^{-.1}$, and **whp** , for $t \leq \mathcal{T}_W$, $\mathbf{v}(t)$ satisfies:

$$\begin{aligned}
\left| \frac{m}{nz^2} - \frac{1}{2c} \right| &\leq n^{-.2+\epsilon}. & (79)\\
\left| \frac{v}{np(z)\beta(z)} - 1 \right| &\leq n^{-.2+\epsilon}. & (80)\\
\left| \frac{v_1}{n(z^2 - zc\beta(z)(1-e^{-z}))} - \frac{1}{c} \right| &\leq n^{-.2+\epsilon}, & (81)
\end{aligned}$$

and

$$|\mathbf{v}(t) - \mathbf{v}(t-1)| \leq \log n. \tag{82}$$

This implies that $v(\mathcal{T}_W) > nz^2/\log n$ and so on exit $z(\mathcal{T}_W) < n^{-.1}$ **whp**. Assume that (79)–(82) hold. We see from (79), (82), and $z(\mathcal{T}_W - 1) \geq n^{-.1}$ that $\mathcal{T}_W \neq n$. It follows from (80) that $v(\mathcal{T}_W - 1) = \Omega(n^{.8})$ and then applying $z(\mathcal{T}_W) < n^{-.1}$ and (82) we see that

$$\begin{aligned}
z(\mathcal{T}_W) &= n^{-.1} + O(n^{-.5}), & (83)\\
\beta(\mathcal{T}_W) &= \frac{\gamma}{c} + O(z), & (84)\\
m(\mathcal{T}_W) &= \frac{nz^2}{2c} + O(n^{.9}z^2), & (85)\\
v(\mathcal{T}_W) &= \frac{n\gamma z^2}{2c} + O(n^{.9}z^2), & (86)\\
v_1(\mathcal{T}_W) &= \frac{n(1-\gamma)z^2}{c} + O(n^{.9}z^2), & (87)\\
\mathcal{T}_W &= ng(z(\mathcal{T}_W)) + O(n^{.8+\epsilon}). & (88)
\end{aligned}$$

here, and in the immediate sequel, $z = z(\mathcal{T}_W), v = v(\mathcal{T}_W)$ etc..

¿From Corollary 2 we know that $\mathbf{z}(\mathcal{T}_W)$, conditioned on $\mathbf{v}(\mathcal{T}_W)$, is distributed uniformly on $Z_{\mathbf{v}(\mathcal{T}_W)}$. It follows then from (7) that **qs**

$$\begin{aligned} \sum_{j:X_j\geq 2}(X_j)_2 &= \frac{vz^2e^z}{f(z)}+O(v^{1/2}\log n) \\ &= 2v+O(n^{.9}z^2), \qquad (89) \\ \sum_{j:X_j\geq 3}(X_j)_2 &= \frac{vz^2(e^z-1)}{f(z)}+O(v^{1/2}\log n) \\ &= O(n^{.9}z^2). \qquad (90) \end{aligned}$$

We use these formulas to show that **whp** $G_{\mathbf{z}}$ has no *heavy* cycles i.e. cycles containing vertices of degree 3 or more. Indeed let C_r be the number of heavy cycles of size r in $G_{\mathbf{z}}$. According to (89) and (90), it suffices to show that $\lim_{n\to\infty}\sum_{r\geq 3}\mathbf{E}(C_r|A)=0$ uniformly for the equivalence classes A such that the degree sequence $\mathbf{d}=\mathbf{d}(A)$ satisfies

$$\sum_{j:d(j)\geq 2}(d(j))_2 = 2v+O(n^{.9}z^2), \qquad (91)$$

$$\sum_{j:d(j)\geq 3}(d(j))_2 = O(n^{.9}z^2). \qquad (92)$$

Using (2)

$$\mathbf{E}(C_r|A)=\frac{\binom{m}{r}\frac{1}{2}(r-1)!2^r r!}{(2m)_{2r}}\cdot\sum_B\prod_{s\in B}(d(s))_2, \qquad (93)$$

where the sum is over all $B\in[n]$ such that $|B|=r$ and $\max_{s\in B}d(s)\geq 3$.

Let us explain the numerator. We choose a set I of r indices such that, for every $i\in I$, the location $\{2i-1,2i\}$ is not starred, which can be done in $\binom{m}{r}$ ways. Next we choose a cyclic ordering of the set B, in $(r-1)!/2$ ways. Finally we assign the edges of the cycle to the r chosen locations $\{2i-1,2i\}_{i\in I}$, in $2^r r!$ ways.

The sum is bounded as follows:

$$\begin{aligned} \sum_B\prod_{s\in B}(d(s))_2 &\leq \sum_{j:d(j)\geq 3}(d(j))_2\cdot\sum_{B:|B|=r-1}\prod_{s\in B}(d(s))_2 \\ &\leq \sum_{j:d(j)\geq 3}(d(j))_2\cdot\left(\sum_{s:d(s)\geq 2}(d(s))_2\right)^{r-1}\Bigg/(r-1)! \\ &= O(n^{.9}z^2(v+O(n^{.9}z^2))^{r-1}2^{r-1}/(r-1)! \end{aligned}$$

by (89) and (90).

So, combining this inequality with (93), we obtain

$$\mathbf{E}(C_r|A)=O\left(\frac{4^r(m)_r}{(2m)_{2r}}(v+O(n^{.9}z^2))^{r-1}n^{.9}z^2\right).$$

By (85) and (86), v and m are both of order nz^2 and

$$\frac{v}{m} = \gamma + O(n^{-.1}).$$

Since $\gamma < 1$, we easily obtain

$$\sum_{r\geq 3} \mathbf{E}(C_r|A) = O(n^{-.1}(v/m)^3) = O(n^{-.1}),$$

so that

$$\mathbf{Pr}(G_{\mathbf{z}} \text{ contains a heavy cycle}|A) = o(1),$$

for a likely class A. We can now remove the conditioning on the class A to obtain

$$\mathbf{Pr}(G_{\mathbf{z}} \text{ contains a heavy cycle}) = o(1).$$

So **whp** at this point the graph consists of isolated trees and cycles. KSGREEDY has not made any mistakes and cannot do so from now on.

Now let us consider (isolated) cycles without heavy vertices. Let c_r denote the number of such cycles of size r, $r \geq 3$. Then,

$$\begin{aligned} \mathbf{E}\left(\sum_{r=3}^{n} c_r \;\middle|\; A\right) &= \sum_{r=3}^{n} \binom{v_2}{r} \frac{(r-1)!}{2} \frac{2^r}{(2m)_r} \\ &\leq \sum_{r=3}^{n} \frac{1}{r} \left(\frac{v_2}{m}\right)^r \\ &\leq \log n. \end{aligned} \tag{94}$$

So **whp** there are fewer than $(\log n)^2$ isolated cycles.

We want to show, in addition, that almost all of m edges belong to isolated paths without multiple edges. Let D denote the total number of those edges. Then

$$\mathbf{E}(D|A) = \binom{v_1}{2} \sum_{r\geq 2} (r-1) \frac{(v_2)_{r-2} \binom{m}{r-1} 2^{r-1} (r-1)!}{(2m)_{2(r-1)}} 2^{r-2}, \tag{95}$$

where $v_2 = |\{j : d(j) = 2\}|$. (To construct a path of length r, we (a) choose two endvertices, in $\binom{v_1}{2}$ ways; (b) select and order $r-2$ intermediate vertices of degree 2, in $(v_2)_{r-2}$ ways; (c) select $r-1$ non-starred locations $\{2i-1, 2i\}$, in $\binom{m}{r-1}$ ways; and (d) assign each of $r-1$ edges of the path to one of the selected $r-1$ locations, in $(r-1)!$ ways. Needless to say, the factor $2^{r-2}/(2m)_{2(r-1)}$ comes from (2).

By (7) we see that for a likely class A

$$v_2 = v + O(n^{.9} z^2).$$

After simple computations based on (83)-(87), the above sum for $\mathbf{E}(D \mid A)$ simplifies then to

$$\frac{nz^2}{2c} + O(n^{.9} z^2) = m - O(n^{.9} z^2).$$

So we obtain

$$\mathbf{Pr}(m - D \geq n^{.91} z^2 | A) = o(1),$$

and unconditioning

$$\mathbf{Pr}(m - D \geq n^{.91} z^2) = o(1).$$

Now in general $G_{\mathbf{z}}$ is a multigraph and we are interested in simple graphs. But applying Lemma 1 to the input graph $G_{\mathbf{x}}$ we see also that

$$\mathbf{Pr}(G_{\mathbf{z}} \text{ contains a heavy cycle or } m - D \geq n^{.91} z^2 \mid G_{\mathbf{z}} \text{ is simple}) = o(1).$$

Thus, for the algorithm applied to the random graph, **whp** at time $\mathcal{T}_W$ the graph consists of a set of vertex disjoint cycles $\mathcal{C}$ plus a forest of trees that are almost all paths, with all other trees and cycles containing fewer than $n^{.91}$ edges. By the end of Phase 1 the only remaining vertices will be those of $\mathcal{C}$. Denote by $G_{\mathbf{z}}^*$ the subgraph of $G_{\mathbf{z}}$ which is a union of those cycles and paths. We know that **whp**

$$\begin{aligned} m^* : &= m(G_{\mathbf{z}}^*) = \frac{nz^2}{2c} + O(n^{.9} z^2), \\ v^* : &= v(G_{\mathbf{z}}^*) = \frac{n\gamma z^2}{2c} + O(n^{.9} z^2), \\ v_1^* : &= v_1(G_{\mathbf{z}}^*) = \frac{n(1-\gamma) z^2}{c} + O(n^{.9} z^2), \end{aligned}$$

and that $L^{(n)}$, the total length of the cycles, satisfies $L^{(n)} \leq n^{.91} z^2$. Denote the latter event B. Also denote $(m^*, v^*, v_1^*) = \mathbf{v}^*$. Observe that, conditioned on $\mathbf{v}^*$ and the event B, the graph $G_{\mathbf{z}}^*$ is distributed uniformly on the set of all graphs G, of the structure in question, such that $\mathbf{v}(G) = \mathbf{v}^*$ and $L(G) \leq n^{.91} z^2$. We may and shall assume that the vertices of degree 1 and 2 are specified, and moreover that v_1 vertices of degree 1 are paired in a fixed way as the endvertices of $v_1/2$ paths in the graph G.

Let $Y^{(n)} = (Y_k^{(n)}, k \geq 3)$ where $Y_k^{(n)} = Y_k(G_{\mathbf{z}}^*), k \geq 3$, denotes the number of cycles in $G_{\mathbf{z}}^*$ of length k, and $Y^{(n)} = \{Y_k^{(n)}\}_{k \geq 3}$. Given a (finitary) sequence $\mathbf{j} = \{j_k\}_{k \geq 3}$ of nonnegative integers such that $r := \sum_k k j_k \leq (\log n)^2$, introduce $N(\mathbf{v}^*, \mathbf{j})$ the total number of those graphs G such that $Y_k(G) = j_k$, $k \geq 3$. Then

$$N(\mathbf{v}^*, \mathbf{j}) = \binom{v^*}{r} \left\{ r! \prod_k \left(\frac{1}{2k}\right)^{j_k} \Big/ j_k! \right\} \cdot (v^* - r)! \binom{v^* - r + v_1^*/2 - 1}{v_1^*/2 - 1}.$$

Here is why. $\binom{v^*}{r}$ is the number of ways to choose r vertices for the cycles. The second factor is the number of ways to partition an r-element set into j_3 (undirected) cycles of length 3,..., j_k cycles of length k, etc . The last two factors account for the number of ways to use the remaining $v^* - r$ vertices of degree 2 to build $v_1^*/2$ paths. (The last binomial coefficient is the number of nonnegative integer solutions of $\sum_{s=1}^{v_1^*/2} \ell_s = v^* - r$.)

Since $v^* \approx v$, $m^* \approx m$, $v^* + v_1^*/2 = m^*$ and $r^2 = o(v^*), r = o(v_1^*)$, the formula for $N(\mathbf{v}^*, \mathbf{j})$ becomes

$$N(\mathbf{v}^*, \mathbf{j}) = (1 + o(1)) \frac{(v + v_1/2 - 1)!}{(v_1/2 - 1)!} \prod_k \left(\frac{\gamma^k}{2k}\right)^{j_k} \Big/ j_k!.$$

This easily implies that for every fixed $\mathbf{j}$, conditioned on $\mathbf{v}^*$ and the event B, whence unconditionally, we have

$$\mathbf{Pr}(Y^{(n)} = \mathbf{j}) \to \prod_k e^{-\gamma^k/2k} (\gamma^k/2k)^{j_k} / j_k!,$$

that is the components of $Y^{(n)}$ are asymptotically independent, with $Y_k^{(n)}$ close in distribution to Poisson Y_k with parameter $\gamma^k/2k$. By the Borel-Cantelli lemma, almost surely $Y = \{Y_k\}_{k\geq 3} \in \mathbf{J}$, the *countable* set of all finitary sequences $\mathbf{j}$. Therefore we have

$$\lim_{n\to\infty} \mathbf{Pr}(Y^{(n)} \in B) = \mathbf{Pr}(Y \in B), \quad \forall B \in \mathbf{J}.$$

Consequently, the total number of cycles, $\sum_{k\geq 3} Y_k^{(n)}$ converges —in distribution—to $\sum_{k\geq 3} Y_k$. In particular,

$$\begin{aligned}\mathbf{Pr}(Y_k^{(n)} = 0, k \geq 3) &\to \exp\left(-\sum_{k\geq 3} \frac{\gamma^k}{2k}\right) &(96)\\ &= (1-\gamma)^{1/2} \exp(\gamma/2 + \gamma^2/4). &(97)\end{aligned}$$

The total length of the cycles, $\sum_{k\geq 3} kY_k^{(n)}$ converges to $L = \sum_{k\geq 3} kY_k$, with

$$\mathbf{E}(L) = \frac{\gamma^3}{2(1-\gamma)}.$$

This proves Theorem 2.

Note. The limiting distribution of $Y^{(n)}$ is the same as for the random graph $G_\nu = G(\nu, \gamma\nu/2)$, $\nu \to \infty$, ([3]). However the latter graph **whp** contains plenty of tree components that are not paths.

Now consider the size of the matching μ_1 produced. From Lemma 12, (86) and Lemma 9 we see that **whp**

$$\begin{aligned}\mu_1 &= \mathcal{T}_W + O(v(\mathcal{T}_W))\\ &= ng(z(\mathcal{T}_W)) + O(n^{.8+\epsilon}) + O(nz^2(\mathcal{T}_W))\\ &= n[g(0) + O(z^2(\mathcal{T}_W))] + O(n^{.8+\epsilon}) = ng(0) + O(n^{.8+e}),\\ &= n\left(1 - \frac{\gamma}{c} - \frac{\gamma^2}{2c}\right) + O(n^{.8+\epsilon}),\end{aligned}$$

since when $z = 0$, $\beta = \gamma/c$ and $\log c/\gamma = \gamma$. This proves Theorem 4 for the case $c < e$.

5.2 The supercritical case: $c > e$

This time we take $a = 1/6$ and $\alpha < 1/6 - \epsilon$, $\epsilon < 0.01$. We see that **whp** (79) – (82) hold with $n^{-1/6+\epsilon}$ on the right, and so we assume this for the remainder of the section. Then $\mathcal{T}_W \neq n$ by (79) and, since $v = \Omega(n)$ here,

$$z(\mathcal{T}_W) = z^* + n^{-1/6} + O(n^{-.99}), \tag{98}$$

where z^* is the root of $h(z) = 0$ (see (48)). $v(\mathcal{T}_W) < nz(\mathcal{T}_W)^2/\log n$ is again ruled out by (79) – (82). Using (98) and Lemma 12, we assert that **whp**

$$\begin{aligned}\mathcal{T}_W &= ng(z(\mathcal{T}_W)) + O(n^{5/6+\epsilon}) = ng(z^*) + O(n|z(\mathcal{T}_W) - z^*| + n^{5/6+\epsilon})\\ &= ng(z^*) + O(n^{5/6+\epsilon}).\end{aligned}$$

Now (81) (see (61)) then implies that

$$v_1 = O(n^{5/6+\epsilon}). \tag{99}$$

We complete the analysis of this section by showing that **whp** Phase 1 terminates in at most $n^{.75}$ further steps. The proof combines previous ideas and the tail inequalities for martingales with bounded differences – see for example Bollobás [4] or McDiarmid [16].

Let $\mathcal{T}^*$ be the first time $t > \mathcal{T}_W$ when either (i) $v_1(t) = 0$, or (ii) $t > \mathcal{T}_W + n^{6/7}$, or (iii) $t \geq \mathcal{T}_L$.

But then **qs**

$$\{v_1(\mathcal{T}^*) = 0\} \text{ or } \{\mathcal{T}^* > \mathcal{T}_W + n^{6/7}\}.$$

Furthermore, for $\mathcal{T}_W \leq t < \mathcal{T}^*$, we deduce from Lemma 11 that

$$m(t) = \frac{n(z^*)^2}{2c} + O(n^{5/6+\epsilon}), \tag{100}$$

$$v(t) = np(z^*)\beta(z^*) + O(n^{5/6+\epsilon}), \tag{101}$$

$$v_1(t) = O(n^{5/6+\epsilon}), \tag{102}$$

and from $t < \mathcal{T}_L$ that

$$z(t) = z^* + O(n^{-.1}). \tag{103}$$

Introduce

$$\tilde{v}_1(t) = v_1(\mathcal{T}_W) + \sum_{\tau=\mathcal{T}_W}^{t-1} \Delta\tilde{v}_1(\tau),$$

$$\Delta\tilde{v}_1(\tau) = [v_1(\tau+1) - v_1(\tau)]1_{\{|v_1(\tau+1)-v_1(\tau)| \leq \log n\}}.$$

Clearly **qs**

$$\tilde{v}_1(t) = v_1(t),\ t \in [\mathcal{T}_W, \mathcal{T}^*].$$

Applying Lemma 6 and (100)-(103), we see that, for $\tau \in [\mathcal{T}_W, \mathcal{T}^*)$,

$$\begin{aligned}\mathbf{E}(v_1(\tau+1) - v_1(\tau) \mid \mathbf{v}(\tau)) &= -1 + \frac{n^2 p(z^*)^2\beta(z^*)^2(z^*)^4 e^{z^*}}{n^2(z^*)^4 c^{-2} f(z^*)^2} + o(1) \\ &= -1 + \beta(z^*)^2 c^2 e^{-z^*} + o(1) \\ &= -1 + \frac{e^{z^*}(z^*)^2}{(e^{z^*}-1)^2} + o(1).\end{aligned} \tag{104}$$

(From $h(z^*) = 0$ it follows that $c\beta(z^*) = z^*(1 - e^{-z^*})^{-1}$.) Consequently, using (60) and assuming $\mathcal{T}^* < \mathcal{T}_L$, for τ in question we have

$$\mathbf{E}(\Delta\tilde{v}_1(\tau)) \leq -d,$$

where $d > 0$ is an absolute constant. We recall the classic inequality (see for example [16]) stating that if X is a random variable such that $X \in [-1, 1]$ a.s. and $\mathbf{E}X \leq 0$ then

$$\mathbf{E}(e^{\lambda X}) \leq e^{\lambda^2/2}. \tag{105}$$

We apply this now with $X = (\Delta\tilde{v}_1(\tau) + d)/\log n$ and λ replaced by $\lambda \log n$ to get

$$\mathbf{E}(e^{\lambda\tilde{v}_1(\tau+1)} \mid \{\mathbf{v}(j)\}_{\mathcal{T}_W \leq j \leq \tau}) \leq \rho e^{\lambda\tilde{v}_1(\tau)}, \tag{106}$$

where

$$\rho = \exp\{-\lambda d + \lambda^2 \log^2 n\}.$$

We deduce that the sequence $(Z_t = e^{\lambda \tilde{v}_1(t)}/\rho^{t-\mathcal{T}_W})_{t\geq \mathcal{T}_W}$ is a supermartingale. The factor ρ attains its minimum value of $\exp[-d^2/(2\log n)^2]$ for $\lambda = d/(2(\log n)^2)$, which we now impose. By the optional stopping theorem we obtain

$$\mathbf{E}\left(\frac{e^{\lambda \tilde{v}_1(\mathcal{T}^*)}}{\rho^{\mathcal{T}^*-\mathcal{T}_W}}\middle| \mathbf{v}(\mathcal{T}_W)\right) \leq e^{\lambda \tilde{v}_1(\mathcal{T}_W)} = e^{\lambda v_1(\mathcal{T}_W)}.$$

So, given (102),

$$\mathbf{E}(Z_{\mathcal{T}^*} \mid \mathcal{T}^* > \mathcal{T}_W + n^{6/7}, \mathbf{v}(\mathcal{T}_W))\mathbf{Pr}(\mathcal{T}^* > \mathcal{T}_W + n^{6/7} \mid \mathbf{v}(\mathcal{T}_W)) \leq e^{o(n^{5/6+\epsilon})}. \tag{107}$$

Now

$$\mathbf{E}(Z_{\mathcal{T}^*} \mid \mathcal{T}^* > \mathcal{T}_W + n^{6/7}, \mathbf{v}(\mathcal{T}_W)) \geq \exp\{d^2 n^{6/7}/(2\log n)^2\}. \tag{108}$$

So from (107) and (108) with ϵ small, we deduce that

$$\mathbf{Pr}(\mathcal{T}^* > \mathcal{T}_W + n^{6/7} \mid \mathbf{v}(\mathcal{T}_W)) = o(1)$$

and consequently

$$\mathbf{Pr}(v_1(\mathcal{T}^*) > 0 \mid \mathbf{v}(\mathcal{T}_W)) = o(1).$$

That is, **whp** at time $\mathcal{T}^* \leq \mathcal{T}_W + n^{6/7+\epsilon}$, v_1, the current number of pendant vertices, becomes zero, which signals the end of Phase 1.

We can now prove Theorem 4, assuming the truth of Theorem 3. It follows from (99) and Fact 1 that **whp** the size μ_1 of the matching produced by Phase 1 satisfies

$$\mu_1 = \mathcal{T}_W + O(n^{.75}) = ng(z^*) + n^{5/6+\epsilon}.$$

According to Theorem 3 **whp** the size μ_2 of the matching produced in Phase 2 satisfies

$$\begin{aligned} \mu_2 &= v(\mathcal{T}^*)/2 + O(n^{1/5+\epsilon}) = v(\mathcal{T}_W)/2 + O(n^{6/7+\epsilon}) \\ &= np(z^*)\beta(z*)/2 + O(n^{6/7+\epsilon}). \end{aligned}$$

Now if $G(\mathcal{T}_W)$ denotes the graph remaining at time $\mathcal{T}_W$ then

$$\begin{aligned} \mu(G(n,M)) &= \mu_1 + \mu(G(\mathcal{T}_W)) \\ &\leq \mu_1 + v(\mathcal{T}_W)/2. \end{aligned} \tag{109}$$

Equation (109) is well known and follows from the fact that if u is a vertex of degree one in a graph H and v is adjacent to u in H, then $\mu(H) = 1 + \mu(H \setminus \{u,v\})$. Thus **whp**

$$\mu(G(n,M)) = n(g(z^*) + p(z^*)\beta(z^*)/2) + O(n^{6/7+\epsilon}).$$

Now use the substitutions $z^* = \gamma^* - \gamma_*$, $\beta(z^*) = \gamma^*/c$, $\log \beta(z^*) = z^* - c\beta(z^*) = -\gamma_*$ and $e^{z^*} = \gamma^*/\gamma_*$ (see (48) and (49)) to obtain the expression given in Theorem 4.

6 Analysis of Phase 2

For simplicity we split the ensuing analysis into three stages. First of all let

$$W_1 = \left\{ \mathbf{v} : \ z \geq n^{-1/100}, \, v \geq Anz^2, \, v_1 \leq m, \, v_1 \leq n^{1/5} (\log n)^6 \right\}.$$

The positive constant A will be revealed later – see (128), but it will be small enough so that (101) implies $\mathbf{v}(\mathcal{T}^*) \in W_1$.

Then let

$$\mathcal{T}_1 = \begin{cases} \min\{t \leq \mathcal{T}_L : \mathbf{v}(t) \notin W_1\} & \text{if such } \tau \text{ exist,} \\ n & \text{otherwise.} \end{cases}$$

Let $\mathcal{T}_{6,7} = \min\{t : \ vz < (\log n)^2\}$. The conclusions of Lemmas 6 and 7 are valid for $t < \mathcal{T}_{6,7}$.

Let $Z_t = v_0(t+1) - v_0(t) - 2$ denote the number of unmatched vertices which are created at time t. Then by Lemmas 6 and 7, if $t < \mathcal{T}_{6,7}$,

$$\mathbf{E}(Z_t \mid \mathbf{v}) = O\left(\frac{v_1 + 1}{m}\right). \tag{110}$$

Consequently,

$$\mathbf{E}\left(\sum_{t=\mathcal{T}^*}^{\mathcal{T}_{6,7}} Z_t\right) = O\left(\mathbf{E}\left(\sum_{t=\mathcal{T}^*}^{\mathcal{T}_{6,7}} \frac{v_1+1}{m}\right)\right). \tag{111}$$

We will later define a stopping time $\mathcal{T}_3$ such that

$$\mathbf{Pr}(\exists \mathcal{T}^* \leq t \leq \mathcal{T}_3 : \ v_1(t) \geq 2n^{1/5}(\log n)^9) = o(1/n). \tag{112}$$

$$\mathbf{E}(\text{number of isolated vertices created after } \mathcal{T}_3) = O(n^{1/5}(\log n)^{12}). \tag{113}$$

(111) and (112) imply that

$$\begin{aligned} \mathbf{E}\left(\sum_{t=\mathcal{T}^*}^{\mathcal{T}_3} Z_t\right) &= O\left(n^{1/5}(\log n)^9 \sum_{t=\mathcal{T}^*}^{\mathcal{T}_{6,7}} \frac{1}{m}\right) \\ &= O(n^{1/5}(\log n)^{10}). \end{aligned} \tag{114}$$

The upper bound of Theorem 3 then follows from this and (113).

We first prove below that

$$\mathbf{Pr}(\exists \mathcal{T}^* \leq t \leq \mathcal{T}_1 : \ v_1(t) \geq 800 n^{1/50} (\log n)^3) \leq n^{-2}. \tag{115}$$

We prove (115) by proving the following lemma. Let $\mathcal{E}(t_1, t_2)$ denote the event

$$\{v_1(t_1) = 0, v_1(t) > 0 \text{ for } t_1 < t \leq t_2 \text{ and } v_1(t_2) > 800 n^{1/50} (\log n)^3\}.$$

In the proofs of next two lemmas we will use

$$\alpha = n^{-1/50}/200.$$

Lemma 13

$$\mathbf{Pr}(\exists \mathcal{T}^* \le t_1 < t_2 \le \mathcal{T}_1 :\ \mathcal{E}(t_1,t_2)) = O(n^{-4}). \tag{116}$$

Proof Fix $t_1 \in [\mathcal{T}^*, \mathcal{T}_1]$ and $t_2 > t_1$ where $v_1(t_1) = 0$. Let $X_t = v_1(t+1) - v_1(t)$ for $t_1 \le t \le \min\{t_2, \mathcal{T}_1\}$. Define

$$\tau_1 = \begin{cases} \min\{\tau \le \min\{t_2,\mathcal{T}_1\} :\ v_1(\tau) = 0\} & \text{if such } \tau \text{ exist,} \\ \min\{t_2, \mathcal{T}_1\} & \text{otherwise.} \end{cases}$$

For $t_1 \le t \le \tau_1$ we let $Y_t = X_t$ and for $\tau_1 \le t \le t_2$ we let $Y_t = -\alpha$. Note that $|Y_t| \le \log n$ for $t_1 \le t \le t_2$. ¿From Corollary 3 and $\mathbf{v}(t) \in W_1$ for $t_1 \le t \le \tau_1$ we see that $\mathbf{E}(Y_t \mid \mathbf{v}(\tau), \tau \le t) = -\alpha_t$ where $\alpha_t \ge \alpha$ for $t_1 \le t \le t_2$. The occurrence of $\mathcal{E}(t_1,t_2) \cap \{t_2 \le \mathcal{T}_1\}$ implies that $\tau_1 = t_2$ and

$$\sum_{t=t_1}^{t_2} Y_t > 800 n^{1/50} (\log n)^3. \tag{117}$$

By (105) applied to $(Y_t + \alpha)/(\log n + \alpha)$, $\Lambda_t = e^{\lambda Y_t}$ satisfies

$$\mathbf{E}(\Lambda_t \mid \mathbf{v}(\tau), \tau \le t) \le e^{\lambda^2 (\log n)^2 - \lambda\alpha} \qquad \forall \lambda \ge 0.$$

Hence, for $L > 0$ and $\lambda = (L/\alpha + \alpha T)/(2(\log n)^2)$, $(T = t_2 - t_1)$,

$$\begin{aligned} \mathbf{Pr}\left(\sum_{t=t_1}^{t_2} Y_t \ge L/\alpha \mid \mathbf{v}(\tau), \tau \le t_1\right) &\le e^{-\lambda L/\alpha} \mathbf{E}\left(\prod_{t=t_1}^{t_2} \Lambda_t \mid \mathbf{v}(t), t \le t_1\right) & (118) \\ &= e^{-\lambda L/\alpha} \prod_{t=t_1}^{t_2} \mathbf{E}(\Lambda_t \mid \mathbf{v}(t'), t' \le t) \\ &\le \exp\{-\lambda L/\alpha \,-\, \lambda\alpha T + \lambda^2 T (\log n)^2\} & (119) \\ &= \exp\left\{-\frac{1}{4T(\log n)^2}\left(\frac{L}{\alpha} + \alpha T\right)^2\right\} \\ &\le \exp\left\{-\frac{L}{(\log n)^2}\right\}. \end{aligned}$$

Putting $L = 4(\log n)^3$ proves the Lemma. □

Similarly, if $T = \lceil 16(\log n)^3/\alpha^2 \rceil$ then

Lemma 14

$$\mathbf{Pr}(\exists \mathcal{T}^* \le t_1 \le \mathcal{T}_1 - T :\ v_1(t_1) = 0,\, v_1(t) > 0 \textit{ for } t_1 < t \le t + T) = O(n^{-4}).$$

Proof Putting $L = 0$ and $\lambda = \alpha/(2(\log n)^2)$ in (118) and (119) we obtain

$$\begin{aligned} \mathbf{Pr}\left(\sum_{t=t_1}^{t_1+T} Y_t \ge 0 \;\middle|\; \mathbf{v}(\tau), \tau \le t_1\right) &\le e^{-\alpha^2 T/(4(\log n)^2)} \\ &= O(n^{-4}). \end{aligned}$$

□

Our next task is to find the likely shape of $\mathbf{v}$ when $\mathbf{v}$ first leaves W_1. We follow the ideas of Lemma 11. Let $M = 2m - v_1$. From Lemma 6, if $v_1 > 0$, $\mathbf{v} \in W_1$ and $\mathcal{T}^* \leq t \leq \mathcal{T}_1$, then

$$\begin{aligned}
\mathbf{E}(M' - M \mid \mathbf{v}) &= -1 - \frac{vz^2e^z}{mf} - \frac{v^2z^4e^z}{4m^2f^2} + \frac{v_1}{2m} + \frac{v_1vz^2e^z}{4m^2f} + O\left(\frac{(\log n)^2}{vz}\right) \\
&= -1 - \frac{vz^2e^z}{mf} - \frac{v^2z^4e^z}{4m^2f^2} + O\left(\frac{v_1}{m} + \frac{(\log n)^2}{vz}\right) \qquad (120)\\
\mathbf{E}(v' - v \mid \mathbf{v}) &= -1 - \frac{v^2z^4e^z}{4m^2f^2} + \frac{v_1}{2m} + O\left(\frac{(\log n)^2}{vz}\right) \\
&= -1 - \frac{v^2z^4e^z}{4m^2f^2} + O\left(\frac{v_1}{m} + \frac{(\log n)^2}{vz}\right). \qquad (121)
\end{aligned}$$

Similary, from Lemma 7, when $v_1 = 0$,

$$\begin{aligned}
\mathbf{E}(M' - M \mid \mathbf{v}) &= -2 - \frac{2vz^2e^z}{mf} - \frac{v^2z^4e^z}{2m^2f^2} + O\left(\frac{(\log n)^2}{vz}\right) \qquad (122)\\
\mathbf{E}(v' - v \mid \mathbf{v}) &= -2 - \frac{v^2z^4e^z}{2m^2f^2} + O\left(\frac{(\log n)^2}{vz}\right). \qquad (123)
\end{aligned}$$

The important thing to observe here is that $\mathbf{E}(M' - M \mid \mathbf{v})/\mathbf{E}(v' - v \mid \mathbf{v})$ is the same in both cases, up to error terms. We are therefore left to consider the differential equation,

$$\frac{dM}{dv} = \frac{1 + \frac{vz^2e^z}{mf} + \frac{v^2z^4e^z}{4m^2f^2}}{1 + \frac{v^2z^4e^z}{4m^2f^2}}.$$

The solution of this was obtained in (59) – see (58) ($M = 2m$ (up to error v_1) accounts for a factor 2 here):

$$M = \frac{M^*(e^z - 1)z}{z^*(e^{z^*} - 1)} \exp\left\{-\int_z^{z^*} \frac{\xi e^\xi}{e^\xi(1+\xi) - 1} d\xi\right\}, \qquad (124)$$

where $M^* = M(\mathcal{T}^*) = 2m^*$ etc.. Then (up to a v_1 error term)

$$\begin{aligned}
v &= \frac{2mf(z)}{z(e^z - 1)} \\
&= \frac{2m^*f(z)}{z^*(e^{z^*} - 1)} \exp\left\{-\int_z^{z^*} \frac{\xi e^\xi}{e^\xi(1+\xi) - 1} d\xi\right\}. \qquad (125)
\end{aligned}$$

So we define

$$J_5 = \frac{v}{nf(z)} \exp\left\{\int_z^{z^*} \frac{\xi e^\xi}{e^\xi(1+\xi) - 1} d\xi\right\}$$

and

$$J_6 = \frac{m}{nz(e^z - 1)} \exp\left\{\int_z^{z^*} \frac{\xi e^\xi}{e^\xi(1+\xi) - 1} d\xi\right\}.$$

Lemma 15 *Assume* $\mathbf{v}(\mathcal{T}^*) \in W_1$. *Then*

$$\mathbf{Pr}\left(\max_{\mathcal{T}^* \leq \tau \leq \mathcal{T}_1} |J_i(\mathbf{v}(\tau)) - J_i(\mathbf{v}(\mathcal{T}^*))| > n^{-1/6}\right) = o(1), \qquad i = 5, 6.$$

Proof We follow the proof of Lemma 11. Observe first that equations (64) and (65) can be extended to include $i = 5, 6$. Now fix $i = 5$ or 6 and let $J(t) = J_i(\mathbf{v}(t))$. Let now $L = n^{1/3}$ and define $Q(t) = \exp\{L(J(t) - J(\mathcal{T}_1^*))\}$ for $\mathcal{T}^* \leq t < \mathcal{T}_1$. Let $Q(t) = 0$ for $t \geq \mathcal{T}_1$. Equations (66) and (67) are still valid, as is (68). In place of (69) we obtain

$$\nabla J(t)^* \mathbf{E}[\mathbf{v}(t) - \mathbf{v}(t-1) | \mathbf{v}(t-1)] = O\left(\frac{1}{vz}\left(\frac{v_1}{m} + \frac{(\log n)^2}{vz}\right)\right)$$

since we use (120) – (123) in place of Lemma 6.

So

$$\begin{aligned} \mathbf{E}(Q(t) \mid \{\mathbf{v}(s)\}_{s<t}) &= Q(t-1)\left(1 + O\left(L^2\left(\frac{(\log n)^2 + v_1 z}{(vz)^2}\right)\right)\right) && (126)\\ &= Q(t-1)(1 + O(n^{-21/20})). && (127) \end{aligned}$$

The proof of the lemma is completed as in Lemma 11. □

At time $\mathcal{T}_1$ either (i) $z \leq n^{-1/100}$, (ii) $v < Anz^2$, (iii) $v_1 > \min\{m, n^{1/5}(\log n)^6\}$ or (iv) $\mathcal{T}_1 \geq \mathcal{T}_L$. (iv) is unlikely and (125) shows that $v(\mathcal{T}_1 - 1) \approx Cnz^2$ where

$$C = \frac{m^*}{nz^*(e^{z^*} - 1)} \exp\left\{-\int_0^{z^*} \frac{\xi e^\xi}{e^\xi(1+\xi) - 1} d\xi\right\}.$$

This rules out (ii) and (iii) if we take

$$A = \min\left\{\frac{C}{2}, \frac{v(\mathcal{T}^*)}{2z(\mathcal{T}^*)^2}\right\}, \quad (128)$$

and so we can assume that at time $\mathcal{T}_1$,

$$\begin{aligned} z &\approx n^{-1/100} && (129)\\ v &\approx Cnz^2 && (130)\\ m &\approx Cnz^2 && (131) \end{aligned}$$

$$z(\mathcal{T}_1) - z(\mathcal{T}_1 - 1) = O((\log n)/v) \quad (132)$$

is the justification for (129), (now assuming $\mathcal{T}_1 < \mathcal{T}_L$). $m \approx v$ comes from $z = o(1)$ and

$$2 \leq \frac{2m - v_1}{v} = \frac{z(e^z - 1)}{f(z)} = 2\left(1 + \frac{z}{6} + O(z^2)\right). \quad (133)$$

Now let $\mathcal{T}_1' = \max\{t \leq \mathcal{T}_1 : v_1(t) = 0\}$. It follows from Lemma 14 that **whp**

$$\mathcal{T}_1 - \mathcal{T}_1' = O(n^{1/25}(\log n)^3).$$

This and Lemma 13 implies that **whp** (129) – (131) also hold at time $\mathcal{T}_1'$. Indeed,

$$\begin{aligned} m(\mathcal{T}_1') - m(\mathcal{T}_1) &\leq (\mathcal{T}_1 - \mathcal{T}_1')\log n\\ &= O(n^{1/25}(\log n)^4)\\ &= o(m(\mathcal{T}_1)), \end{aligned}$$

and so $m(\mathcal{T}_1') \approx m(\mathcal{T}_1)$. Similarly, $v(\mathcal{T}_1') \approx v(\mathcal{T}_1)$. Furthermore, applying (132) we see that **whp**

$$\begin{aligned} z(\mathcal{T}_1') - z(\mathcal{T}_1) &= O\left(\frac{(\mathcal{T}_1 - \mathcal{T}_1')\log n}{v}\right) \\ &= O\left(\frac{n^{1/25}(\log n)^4}{n^{97/100}}\right) \\ &= o(z(\mathcal{T}_1)), \end{aligned}$$

and so $z(\mathcal{T}_1') \approx z(\mathcal{T}_1)$ holds at time $\mathcal{T}_1'$.

It will be useful in what follows to know that once z gets "small", it is unlikely to ever grow "large" again. We make this precise with the following lemma:

Lemma 16 *Let* $z_{\alpha,\gamma} = n^{-\alpha}(\log n)^{\gamma}$ *for* $1/100 \le \alpha \le 1/5$ *and* $-100 \le \gamma \le 100$. *Suppose that at time* t_0 *we have*

$$z_{\alpha,\gamma}/2 \le z_0 = z(t_0) \le 2z_{\alpha,\gamma}.$$

Let

$$\tilde{W} = \{\mathbf{v}:\ vz \ge n^{1/5}(\log n)^{12},\ v_1 \le 2n^{1/5}(\log n)^9\}.$$

Then

$$\mathbf{Pr}(\exists t \ge t_0:\ \mathbf{v}(t) \in \tilde{W},\ z(t) \ge \Theta z_0 \mid \mathbf{v}(t_0) \in \tilde{W}) = o(n^{-2}), \tag{134}$$

where Θ *is a sufficiently large absolute constant.*

Proof We prove the lemma by showing that for $t > t_0$,

$$\mathbf{Pr}(z(t) \ge \Theta z_0 \mid \mathbf{v}(\tau) \in \tilde{W},\ z(\tau) < \Theta z_0,\ t_0 \le \tau < t) = o(n^{-3}). \tag{135}$$

Lemma 5 shows that if $t_0 \le t$ and

$$\Delta_\alpha = \lceil 4/\alpha \rceil$$

then

$$\mathbf{Pr}(\Delta(G(t)) \ge \Delta_\alpha \mid \mathbf{v}(t_0) \in \tilde{W}) = o(n^{-3}). \tag{136}$$

We let $G_0 = G(t_0)$,and assume from now on that $\Delta(G_0) \le \Delta_\alpha$. We wish to probabilistically bound $z(t)$ from above for $t \ge t_0$. Since $G(t)$ is a subgraph of G_0 we can do this by bounding $z(K)$ for all *vertex induced* subgraphs K of G_0 (which are suitably large).

Here we have

$$\begin{aligned} \frac{2m(K) - v_1(K)}{v(K)} &= \frac{z(K)(e^{z(K)} - 1)}{e^{z(K)} - 1 - z(K)} \\ &\ge 2\left(1 + \frac{z(K)}{6}\right). \end{aligned} \tag{137}$$

The inequality can best be seen from

$$\frac{z(e^z - 1)}{e^z - 1 - z} - 2 = z\left(\frac{\sum_{k\ge 3}\frac{k-2}{k!}z^{k-3}}{\sum_{k\ge 2}\frac{z^{k-2}}{k!}}\right).$$

So,

$$z(K) \le \frac{3(v_3(K) + 2v_4(K) + \cdots + (\Delta_\alpha - 2)v_{\Delta_\alpha}(K))}{v_2(K) + v_3(K) + v_4(K) + \cdots + v_{\Delta_\alpha}(K)}. \tag{138}$$

If $z(K) > az_0$, then (138) implies that

$$\zeta_a(K) = 3\Delta_\alpha(v(K) - v_2(K)) - az_0v_2(K) > 0. \tag{139}$$

We use this first with $a = \Theta/2$.

We deal first with subgraphs K which are connected. Let K^* denote the vertices of K which are of degree at least 3 in G_0. Let $k = |K^*|$ and let $H = H(K^*)$ be the (multi)-graph with vertex set K^* and an edge joining (x, y) for every path P joining $x \in K$ to $y \in K$ all of whose internal vertices are of degree 2 in G_0. The connectedness of K implies that H is connected.

Let T be a spanning tree of H. Let the edges of T be $\{e_1, e_2, \ldots, e_{k-1}\}$ and let $w(e_i)$ be the number of vertices of degree 2 on the corresponding path joining the end points of e_i in K. If $\zeta_{\Theta/2}(K) > 0$ then, from (139),

$$\sum_{i=1}^{k-1} w(e_i) \leq v_2(K) < \frac{6\Delta_\alpha k}{\Theta z_0}. \tag{140}$$

So now let us consider the following event $\mathcal{A}_1(k)$: there exists a set $K^* = \{u_1, u_2, \ldots, u_k\}$ of vertices of degree at least three in G_0, plus a set L of at most $\Lambda = 6\Delta_\alpha k/(\Theta z_0)$ vertices of degree two in G_0 which together induce a tree. Condition on the equivalence class A (as in Fact 2). Then

$$\mathbf{Pr}(\mathcal{A}_1(k) \mid A) \leq \sum_{K^*} k^{k-2} \sum (v_2(t_0))_X \ (m)_{X+k-1} 2^{2X+k-1} \Delta_\alpha^{2k}/(2m)_{2(X+k-1)}. \tag{141}$$

The second sum is over $x_1 + \cdots + x_{k-1} \leq \Lambda$ and $X = x_1 + x_2 + \cdots + x_{k-1}$.
Explanation: We sum over sets K^* of vertices of degree at least three. For a fixed K^* we choose a labelled tree T on k vertices. We take a canonical ordering $e_1, e_2, \ldots, e_{k-1}$ of the edges of T and then decide on the number of degree 2 vertices $x_1, x_2, \ldots, x_{k-1}$ of the paths corresponding to these edges. We then choose a set L of X vertices and place them on the edges, x_i vertices to edge e_i. This can be done in at most $(v_2(t_0))_X$ ways. This defines the sets K^*, L and the tree T' containing them. We now have to estimate the probability that T' exists in G_0. $(m)_{X+k-1}$ counts the ways of choosing positions in $\mathbf{z}$ for the edges of T'. There are 2^{X+k-1} choices for ordering the corresponding entries in $\mathbf{z}$ and then if d_j, t_j denote the degrees of vertex j in G_0, T, Corollary 1 gives

$$\prod_{j \in K} (d_j)_{t_j}/(2m)_{2(X+k-1)} \leq 2^X \Delta_\alpha^{2k}/(2m)_{2(X+k-1)}$$

for the probability that these edges all exist.

Now (see (7))

$$\textbf{qs} \text{ there are fewer than } vz_0/2 \text{ vertices of degree at least 3 in } G_0. \tag{142}$$

Assume A satisfies this condition. In which case

$$X + k \leq \Lambda + k \leq \frac{4\Delta_\alpha v}{\Theta} \leq \epsilon m$$

where $\epsilon = 4\Delta_\alpha/\Theta$ can be made arbitrarily small. It follows after simple estimations that

$$\begin{aligned} \frac{2^{2X+k-1}(v_2(t_0))_X(m)_{X+k-1}\Delta_\alpha^{2k}}{(2m)_{2(X+k-1)}} &\leq \frac{\Delta_\alpha^{2k}}{2^{k-1}((1-\epsilon)m)^{k-1}} \\ &\leq \frac{\Delta_\alpha^{2k}}{m^{k-1}}. \end{aligned}$$

The number of choices for $x_1, x_2, \ldots, x_{k-1}$ is at most $\binom{\Lambda+k-1}{k-1}$ and the number of choices for K^* is at most $\binom{vz_0/2}{k}$. Hence,

$$\begin{aligned}
\mathbf{Pr}(\mathcal{A}_1(k) \mid A) &\leq \binom{\Lambda+k-1}{k-1}\binom{vz_0/2}{k}\frac{k^{k-2}\Delta_\alpha^{2k}}{m^{k-1}} \\
&\leq \left(\frac{2\Lambda e}{k}\right)^k \left(\frac{vz_0 e}{2k}\right)^k \frac{k^{k-2}\Delta_\alpha^{2k}}{m^{k-1}} \\
&\leq \frac{m}{k^2}\left(\frac{6e^2\Delta_\alpha^3}{\Theta}\right)^k .
\end{aligned}$$

Thus if $\kappa = 4/\log(\Theta/(6e^2\Delta_\alpha^3))$ then we have, after removing the conditioning on A,

$$\sum_{k \geq \kappa \log n} \mathbf{Pr}(\mathcal{A}_1(k)) = o(n^{-3}). \tag{143}$$

So now let us condition on the non-occurrence of $\bigcup_{k\geq\kappa\log n}\mathcal{A}_1(k)$. Let $C_1, C_2, \ldots, C_s$ be the components of $G(t)$ with $|C_1^*| \leq |C_2^*| \leq \cdots \leq |C_r^*| < \kappa \log n \leq |C_{r+1}^*| \leq \cdots \leq |C_u^*|$ which are not paths or cycles. Note that from what we have just shown we know that **whp** $\zeta_{\Theta/2}(C_i) < 0$ for $i > r$. Let $C_{u+1}, C_{u+2}, \ldots C_s$ be the components which are paths or cycles. Now if $z(G(t)) > \Theta z_0$ then

$$\begin{aligned}
0 < \zeta_\Theta(G(t)) &= \sum_{i=1}^{s}\zeta_\Theta(C_i) \\
&\leq 3\Delta_\alpha \sum_{i=1}^{r}|C_i^*| + \sum_{i=r+1}^{s}\zeta_{\Theta/2}(C_i) + \sum_{i=r+1}^{s}(\zeta_\Theta(C_i) - \zeta_{\Theta/2}(C_i)) \\
&\leq 3\Delta_\alpha\kappa r\log n + \sum_{i=r+1}^{s}\zeta_{\Theta/2}(C_i) - \frac{\Theta z_0}{2}\sum_{i=r+1}^{s}v_2(C_i) \\
&\leq 3\Delta_\alpha\kappa r\log n - \frac{\Theta z_0}{2}\sum_{i=r+1}^{s}v_2(C_i). \qquad (144)
\end{aligned}$$

Let Γ denote the length of the longest path in $G(t)$ whose interior vertices are of degree 2 in $G(t)$. We will show later that if $\mathbf{v}(t) \in \tilde{W}$ then either $z(t) \leq \Theta z_0$ (and we are done) or with probability $1 - o(n^{-3})$

$$\Gamma \leq 20(\log n)/(\Theta z_0). \tag{145}$$

Assume that (145) holds. Now $|C_i| \leq \Delta_\alpha \Gamma |C_i^*|$ and hence

$$\sum_{i=r+1}^{s} v_2(C_i) \geq v_2(G(t)) - 20\kappa r\Delta_\alpha(\log n)^2/(\Theta z_0).$$

Substituting in (144) we obtain

$$\begin{aligned}
21\kappa(\log n)^2\Delta_\alpha r &> \Theta z_0 v_2(G(t)) \\
&\geq z(t-1)(v_2(G(t-1)) - \Delta_\alpha) \qquad (146) \\
&\geq n^{1/5}(\log n)^{12}/2,
\end{aligned}$$

since $\mathbf{v}(t-1) \in \tilde{W}$.

Thus

$$r = \Omega(n^{1/5}(\log n)^{10}). \tag{147}$$

Let $r = r' + r''$ where r' is the number of C_i, $i \le r$, which contain a vertex of degree one in $G(t)$ and r'' is the number of C_i, $i \le r$, with minimum degree at least two in $G(t)$. Now

$$r' \le v_1(t) \le v_1(t-1) + \Delta_\alpha \le 3n^{1/5}(\log n)^9 < r/2. \tag{148}$$

Let $\nu = \lceil \log n \rceil$ and suppose $r'' \ge \nu\kappa \log n$. Then there exists $k \le \kappa \log n$ and ν indices $i \le r$ each with $|C_i^*| = k$ and C_i having minimum degree at least two. No C_i is a cycle and so each contains at least 2 cycles. In which case we see that C_i spans at least $|C_i| + 1$ edges in G_0. We let $\mathcal{A}_2(k)$ denote the event: G_0 contains ν disjoint sets of vertices K_i with $|K_i^*| = k$, and each $H(K_i^*)$ spanning at least $k+1$ edges.

Condition on the equivalence class A. Arguing as in (141) we see that

$$\mathbf{Pr}(\mathcal{A}_2(k) \mid A) \le \sum \binom{\binom{k}{2}}{k+1}^{\nu} \sum (v_2(t_0))_X \ (m)_{X+\nu(k+1)} \frac{2^{2X+\nu(k+1)}\Delta_\alpha^{2\nu(k+1)}}{(2m)_{2(X+\nu(k+1))}}. \tag{149}$$

The first sum is over $K_1^*, K_2^*, \ldots, K_\nu^*$. The second sum is over $x_1, \ldots, x_{\nu(k+1)}$ and $X = x_1 + x_2 + \cdots + x_{\nu(k+1)}$.
Explanation: We sum over disjoint sets $K_1^*, K_2^*, \ldots, K_\nu^*$ of vertices of degree at least three. For a fixed $K_1^*, K_2^*, \ldots, K_\nu^*$ we choose fixed graphs $H_1, H_2, \ldots, H_\nu$, each with k vertices and $k+1$ edges. We take a canonical ordering $e_1, e_2, \ldots, e_{\nu(k+1)}$ of the edges of $H_1 \cup H_2 \cup \cdots \cup H_\nu$ and then decide on the number of degree 2 vertices $x_1, x_2, \ldots, x_{\nu(k+1)}$ of the paths corresponding to these edges. The rest is as before. We estimate

$$\frac{(v_2(t_0))_X \ (m)_{X+\nu(k+1)} 2^{2X+\nu(k+1)}\Delta_\alpha^{2\nu(k+1)}}{(2m)_{2(X+\nu(k+1))}} \le \left(\frac{v_2(t_0)}{m}\right)^X \frac{\Delta_\alpha^{2\nu(k+1)}}{2^{\nu(k+1)}((1-\epsilon)m)^{\nu(k+1)}},$$

where $\epsilon = (X + \nu k)/(2m) = o(1)$ using (145) and $X = O(\nu(\log n)^2/z_0) = o(m)$, since $\mathbf{v}(t_0) \in \tilde{W}$.

Now (see (7)) we can assume that $v_2(t_0)/m \le 1 - z_0/4$. Indeed, assume $v_2(t_0)/m > 1 - z_0/4$ and hence that $v \approx m$. Then from (7) at $t = t_0$,

$$\begin{aligned} 1 - \frac{v_2}{m} &\ge \frac{3v_3}{2m} \\ &= \frac{vz^3}{4mf(z)} + O(\sqrt{vz}\log n/m) \\ &\ge \frac{z}{4}, \end{aligned}$$

which is a contradiction.

Thus,

$$\begin{aligned} \mathbf{Pr}(\mathcal{A}_2(k) \mid A) &\le \left(\frac{vz_0e}{2k}\right)^{k\nu} \left(\frac{ke}{2}\right)^{\nu(k+1)} \frac{\Delta_\alpha^{2\nu(k+1)}}{2^{\nu(k+1)}((1-\epsilon)m)^{\nu(k+1)}} \sum_{X=0}^{\infty} \binom{X+\nu(k+1)-1}{\nu(k+1)-1} \left(1 - \frac{z_0}{4}\right)^X \\ &= \left(\left(\frac{vz_0e}{2k}\right)^k \left(\frac{ke}{2}\right)^{k+1} \frac{\Delta_\alpha^{2k+2}}{2^{k+1}((1-\epsilon)m)^{k+1}} \left(\frac{4}{z_0}\right)^{k+1}\right)^\nu \end{aligned}$$

$$
\begin{aligned}
&= \left(\left(\frac{ve^2\Delta_\alpha^2}{2(1-\epsilon)m}\right)^k \frac{\Delta_\alpha^2 ke}{(1-\epsilon)mz_0}\right)^\nu \\
&\leq \left(\left(\frac{ve^2\Delta_\alpha^2}{2(1-\epsilon)m}\right)^k \frac{2\Delta_\alpha^2 ke\Theta}{n^{1/5}(\log n)^{12}}\right)^\nu
\end{aligned}
$$

For the last inequality follow (146). Thus after removing the conditioning on A and choosing κ sufficiently small (i.e. Θ is sufficiently large),

$$\sum_{k\leq\kappa\log n} \mathbf{Pr}(\mathcal{A}_2(k)) = O(n^{-K}),$$

for any constant $K>0$. Thus **qs** $r'' \leq \nu\kappa\log n$ which contradicts (147). Thus $z(K)\leq\Theta z_0$ for all K and it only remains to verify (145).

Proof of (145)

Let Λ_k denote the number of induced paths of length k in $G(t)$. Condition on the equivalence class A. Then

$$
\begin{aligned}
\mathbf{E}(\Lambda_k \mid A) &\leq \frac{\binom{v}{2}\binom{v_2}{k-1}(m)_k 2^{2(k+1)}\Delta_\alpha^2}{(2m)_{2k}} \\
&\leq \frac{2v^2\Delta_\alpha^2}{m}\left(\frac{v_2}{m}\right)^{k-1} \\
&\leq \frac{2v^2\Delta_\alpha^2}{m}\left(1-\frac{z(t)}{4}\right)^{k-1}. \qquad (150)
\end{aligned}
$$

Now either $z(t)\leq\Theta z_0$ and we are done already or $z(t)>\Theta z_0$ and the RHS of (150) is $o(n^{-3})$ when $k>20(\log n)/(\Theta z_0)$.

This completes the proof of Lemma 16. □

So now let

$$\mathcal{T}_z = \begin{cases} \min\{t\geq\mathcal{T}_1 : \mathbf{v}(t)\in\tilde{W} \text{ and } \exists\mathcal{T}_1\leq t'\leq t \text{ such that } z(t)>\Theta z(t')\} & \text{if such } t \text{ exist,} \\ n & \text{otherwise.}\end{cases}$$

Let

$$
\begin{aligned}
W_2 = \{\mathbf{v}: \quad & z(t)\geq n^{-1/5}(\log n)^2,\ vz\geq n^{1/5}(\log n)^{12},\ vz^3\geq(\log n)^7 \\
& v_1\leq n^{1/5}(\log n)^9, m\leq 3v,\ Anz^2\leq m, v\leq 2Cnz^2\}.
\end{aligned}
$$

Note that $W_2\subseteq\tilde{W}$. Let

$$\mathcal{T}_2 = \begin{cases} \min\{\mathcal{T}_1\leq t\leq\mathcal{T}_L\wedge\mathcal{T}_z : \mathbf{v}(t)\notin W_2\} & \text{if such } t \text{ exist,} \\ n & \text{otherwise.}\end{cases}$$

We now repeat the analyses that we did for $\mathbf{v}\in W_1$, but we take greater care with estimating the size of v_1. Let

$\mathcal{B}(t)$ be the event $\{m(t+1)=m(t)-2,\ v(t+1)=v(t)-2,\ v_1(t+1)=v_1(t)\}$.

Lemma 17 *Suppose $\mathcal{T}_1 \leq t \leq \mathcal{T}_2$ and $v_1(t) > 0$. Thus* **whp** *$z(t) = O(n^{-1/100})$. Then for $\mathbf{v}(t) \in W_2$*

$$\mathbf{Pr}(\mathcal{B}(t) \mid \mathbf{v}(t)) = 1 - z + o(z).$$

Proof $\mathcal{B}(t)$ occurs when y is of degree 2 and its neighbour other than x is also of degree 2. Then

$$\mathbf{Pr}(\mathcal{B}(t) \mid A(t)) = \frac{2v_2}{2m-1}\frac{2(v_2-1)}{2m-3}.$$

[Recall that $A(t)$ denotes the degree sequence at time t.]

Now $\mathcal{D}(t)$ (see (7)) occurs by assumption in which case

$$\begin{aligned} v_2 &= \frac{vz^2}{2f(z)} + O(v^{1/2}\log n) \\ &= v(1 - z/3 + O(z^2)) + O(v^{1/2}\log n) \\ &= v(1 - z/3 + o(z)). \end{aligned}$$

Thus

$$\mathbf{Pr}(\mathcal{B}(t) \mid A(t), \mathcal{D}(t)) = \left(\frac{v}{m}\right)^2 \left(1 - \frac{2z}{3} + o(z)\right).$$

Now (133) shows

$$\begin{aligned} \frac{m}{v} &= 1 + \frac{z}{6} + O(z^2 + v_1/v) \\ &= 1 + \frac{z}{6} + o(z), \end{aligned}$$

when $\mathbf{v} \in W_2$. Hence,

$$\mathbf{Pr}(\mathcal{B}(t) \mid A(t), \mathcal{D}(t)) = 1 - z + o(z).$$

The lemma follows on removing the conditioning on $A(t), \mathcal{D}(t)$. □

Lemma 18 *Suppose that $\mathcal{T}_1' \leq t \leq \mathcal{T}_2$ and that $v_1(t) = 0$ e.g. $t = \mathcal{T}_1'$. Let $T = T(t) = K(\log n)^3/z(t)^3$, for some large positive constant K. Then*

$$\mathbf{Pr}(t + T \leq \mathcal{T}_2 \text{ and } v_1(\tau) > 0 \text{ for } t < \tau \leq t + T \mid \mathbf{v}(t)) = O(n^{-4}), \tag{151}$$

and

$$\mathbf{Pr}(\exists \tau \in [t, \min\{\mathcal{T}_2, t+T\}]:\ v_1(\tau) \geq n^{1/5}\log n \mid \mathbf{v}(t)) = O(n^{-4}). \tag{152}$$

Proof Fix $t' > t$. Let $b = b(t')$ denote the number of *non-occurrences* of event $\mathcal{B}(\tau)$ in the interval $[t, t']$. Note that if $t' \leq \mathcal{T}_2$ then

$$|m(t) - m(t') - 2(t'-t)| \leq b\log n. \tag{153}$$

$$|v(t) - v(t') - 2(t'-t)| \leq b\log n. \tag{154}$$

Hence,

$$\frac{2m(t') - v_1(t')}{v(t')} - \frac{2m(t) - v_1(t)}{v(t)} = \frac{v_1(t)}{v(t)} - \frac{v_1(t')}{v(t')} + \frac{4(t'-t)(m(t) - v(t)) + O(bv(t)\log n)}{v(t)(v(t) - 2(t'-t) + O(b\log n))}. \tag{155}$$

We observe next that

$$\begin{aligned} v(t') &\geq v(t) - O\left(\frac{(\log n)^4}{z(t)^3}\right) \\ &= (1 - o(1))v(t), \end{aligned}$$

since $t \leq \mathcal{T}_2$. Thus

$$\begin{aligned} \left|\frac{v_1(t')}{v(t')}\right|, \left|\frac{v_1(t)}{v(t)}\right| &\leq \frac{2n^{1/5}(\log n)^9 z(t)}{v(t)z(t)} \\ &\leq \frac{2z(t)}{(\log n)^3} \end{aligned} \tag{156}$$

since $t \leq t' \leq \mathcal{T}_2$. Similarly,

$$2(t'-t) + O(b \log n) = O\left(\frac{(\log n)^4}{z(t)^3}\right) = O\left(\frac{v(t)}{(\log n)^3}\right).$$

Thus the absolute value of the RHS of (155) – denoted $|(155)|$ – is at most

$$O\left(\frac{z(t)}{(\log n)^3}\right) + \frac{5(t'-t)(m(t) - v(t))}{v(t)^2} + O\left(\frac{b \log n}{v(t)}\right).$$

Now $z(t) \leq \Theta n^{-1/100}$, (156), (133) and $\mathbf{v}(t) \in W_2$ imply that

$$m(t) - v(t) \leq v(t)z(t).$$

So,

$$\begin{aligned} |(155)| &\leq O\left(\frac{z(t)}{(\log n)^3}\right) + \frac{5(t'-t)z(t)}{v(t)} + O\left(\frac{b \log n}{v(t)}\right) \\ &\leq O\left(\frac{z(t)}{(\log n)^3}\right) + O\left(\frac{z(t)(\log n)^3}{v(t)z(t)^3}\right) + O\left(\frac{b \log n}{v(t)}\right) \\ &\leq O\left(\frac{z(t)}{(\log n)^3}\right) + O\left(\frac{b \log n}{v(t)}\right). \end{aligned} \tag{157}$$

If $(t'-t)z(t)^2 \leq (\log n)^2$ then

$$\frac{b \log n}{v(t)} \leq \frac{(t'-t)\log n}{v(t)} \leq \frac{(\log n)^3 z(t)}{v(t)z(t)^3} \leq \frac{z(t)}{(\log n)^4}, \tag{158}$$

and then (157) gives

$$|(155)| = O\left(\frac{z(t)}{(\log n)^3}\right), \tag{159}$$

which is what we are after. For $(t'-t)z(t)^2 \geq (\log n)^2$ we prove that (158) holds **qs**.

Let $X_\tau, \tau \in [t, \mathcal{T}_2]$ be the indicator random variable for the event $\mathcal{B}(\tau)$ and let $X_\tau = 0$ for $\tau \in [\mathcal{T}_2 + 1, t + T]$. Let $S_\tau = X_t + \cdots + X_\tau$. It follows from Lemma 17 that if $\mathcal{T}_2 \leq \mathcal{T}_z$ then

$$\mathbf{E}(X_\tau \mid \mathbf{v}(\tau'), \, \tau' \leq \tau) \leq 2\Theta z(t).$$

So, see (105), for any $\lambda > 0$,

$$\mathbf{E}(e^{\lambda X_\tau} \mid \mathbf{v}(\tau'), \tau' \leq \tau) \leq e^{\lambda^2 + 2\Theta z(t)\lambda},$$

and hence

$$\mathbf{E}(e^{\lambda b-(t'-t)(\lambda^2+2\Theta\lambda z)}) \leq 1.$$

Then by the Markov inequality, with $\lambda = \Theta z/2$,

$$\begin{aligned}\mathbf{Pr}(b \geq 3\Theta z(t'-t)) &\leq e^{(t'-t)(\lambda^2-\lambda\Theta z)} \\ &= e^{-(t'-t)\Theta^2 z^2/4} \\ &\leq e^{-\Theta^2(\log n)^2/4},\end{aligned} \tag{160}$$

and so (159) holds **qs**. In which case

$$2\left(1+\frac{z(t')}{6}+O(z(t')^2)\right)-2\left(1+\frac{z(t)}{6}+O(z(t)^2)\right)=O\left(\frac{z(t)}{(\log n)^2}\right)$$

and

$$z(\tau) = (1+o(1))z(t) \qquad \text{for } \tau \in [t, \min\{t+T, \mathcal{T}_2\}]. \tag{161}$$

So we introduce yet another stopping time

$$\mathcal{T}_\zeta = \begin{cases} \min\{\tau \geq t : z(\tau) \leq z(t)/2\} & \text{if such } \tau \text{ exist,} \\ n & \text{otherwise.} \end{cases}$$

Next let $Y_\tau = v_1(\tau+1) - v_1(\tau)$ for $t \leq \tau \leq \min\{\mathcal{T}_2, \mathcal{T}_\zeta, t+T\}$. For $\min\{\mathcal{T}_2, \mathcal{T}_\zeta, t+T\} < \tau \leq t+T$ we let $Y_\tau = 0$ with probability $1-2z(t)$ and equal $-z(t)/400$ with probability $2z(t)$. Then using Corollary 3 and Lemma 17 we see that the random variables (Y_τ) satisfy

- $|Y_\tau| \leq \log n$.
- $\mathbf{Pr}(Y_\tau \neq 0 \mid \{\mathbf{v}(\sigma)\}_{\sigma<\tau}) \leq 2z(t)$.
- $\mathbf{E}(Y_\tau \mid \{\mathbf{v}(\sigma)\}_{\sigma\leq\tau},\, Y_\tau \neq 0) \leq -z(t)/400$.

So, for $\delta = z(t)/400$ and $\lambda = \delta/(2(\log n)^2)$,

$$\begin{aligned}\mathbf{E}(e^{\lambda(Y_\tau+\delta)} \mid \{\mathbf{v}(\sigma)\}_{\sigma<\tau}) &\leq (1-2z)e^{\lambda\delta} + 2ze^{\lambda^2(\log n)^2} \\ &= e^{\lambda\delta}(1+2z(e^{\lambda^2(\log n)^2-\lambda\delta}-1)) \\ &= e^{\lambda\delta}(1+2z(e^{-\delta^2/(4(\log n)^2)}-1)) \\ &\leq e^{\lambda\delta}(1-z\delta^2/(3(\log n)^2)) \\ &\leq e^{\lambda\delta - z\delta^2/(3(\log n)^2)}.\end{aligned}$$

Now in order that $t+T \leq \mathcal{T}_2$ and $v_1(\tau) > 0$ for $t < \tau \leq t+T$, we must have

$$\sum_{\tau=t}^{t+T}(Y_\tau+\delta) \geq T\delta.$$

But, then using the Markov inequality

$$\begin{aligned}\mathbf{Pr}\left(\sum_{\tau=t}^{t+T}(Y_\tau+\delta) \geq T\delta \,\middle|\, \mathbf{v}(t)\right) &\leq \exp\left\{-\frac{Tz\delta^2}{3(\log n)^2}\right\} \\ &= O(n^{-4}).\end{aligned}$$

This proves (151). To prove (152) we let $Z_t = 0$ and $Z_\tau = Y_t + \cdots + Y_{\tau-1}$ for $t < \tau \leq \min\{\mathcal{T}_2, \mathcal{T}_\zeta, t+T\}$, where $Y_t, Y_{t+1}, \ldots,$ are as defined above. The above analysis shows that the sequence

$$S_\tau = \exp\{\lambda Z_\tau + (\tau - t)z\delta^2/(3(\log n)^2)\}$$

is a supermartingale. By the maximum inequality, see for example Chung [6] Theorem 9.4.1 (2), for any $\gamma > 0$, we have,

$$\gamma \mathbf{Pr}(\max_{t \leq \tau \leq \min\{\mathcal{T}_2, \mathcal{T}_\zeta, t+T\}} S_\tau \geq \gamma) \leq \mathbf{E}(S_t) = 1.$$

Putting $\gamma = n^5$ we see that

$$\mathbf{Pr}(\exists \tau : \ \lambda Z_\tau \geq 5 \log n + Tz\delta^2/(3(\log n)^2)) \leq n^{-5}.$$

Equation (152) follows as $v_1(\tau) = Z_\tau$ under the assumption $v_1(t) = 0$ and $v_1(t') > 0$ for $t < t' \leq \tau$. □

We must now check that (130), (131) are still valid when $\mathbf{v}$ first exits from W_2. With J_5, J_6 as defined prior to Lemma 11,

Lemma 19

$$\max_{\mathcal{T}_1' \leq t \leq \mathcal{T}_2} |J_i(\mathbf{v}(t)) - J_i(\mathbf{v}(0))| \leq n^{-1/40}, \qquad \mathbf{qs}, \ i = 5, 6.$$

Proof We follow the proof of Lemma 15 with $L = n^{1/20}$ and obtain (see (126)) that for $t \geq \mathcal{T}_1'$

$$\mathbf{E}[Q(t) \mid \{\mathbf{v}(s)\}_{s<t}] = Q(t-1)\left(1 + O\left(L^2\left(\frac{(\log n)^2 + v_1 z}{(vz)^2}\right)\right)\right)$$

where v_1, v, z are evaluated at $t-1$. Hence,

$$\begin{aligned}
\mathbf{E}(Q(\mathcal{T}_2)) &\leq \mathbf{E}\left(\prod_{t=\mathcal{T}_1'}^{\mathcal{T}_2}\left(1 + O\left(\frac{L^2(v_1 z + (\log n)^2)}{v^2 z^2}\right)\right)\right) && (162)\\
&\leq \mathbf{E}\left(\exp\left\{\sum_{t=\mathcal{T}_1'}^{\mathcal{T}_2} O\left(\frac{L^2(v_1 z + (\log n)^2)}{m^2 z^2}\right)\right\}\right) && (163)\\
&\leq \mathbf{E}\left(\exp\left\{\sum_{t=\mathcal{T}_1'}^{\mathcal{T}_2} O\left(\frac{n^{1/2}L^2 v_1}{m^{5/2}} + \frac{L^2(\log n)^2 n}{m^3}\right)\right\}\right) && (164)\\
&\leq \mathbf{E}\left(\exp\left\{O\left(\frac{n^{1/2}L^2 v_1}{m(\mathcal{T}_2)^{3/2}} + \frac{L^2(\log n)^2 n}{m(\mathcal{T}_2)^2}\right)\right\}\right) && (165)\\
&= O(1) && (166)
\end{aligned}$$

We use $m \leq 3v$ for $\mathbf{v} \in W_2$ to go from (162) to (163). We use $Anz^2 \leq m \leq 2Cnz^2$ for $\mathbf{v} \in W_2$ to go from (163) to (164). We use the fact that $m(\mathcal{T}_2) \approx m(\mathcal{T}_2 - 1) = \Omega(n^{3/5-o(1)})$ and $v_1 \leq n^{1/5+o(1)}$ to go from (165) to (166). The proof can then be completed as in Lemma 11. □

At time $\mathcal{T}_2$ either (i) $z < n^{-1/5}(\log n)^2$, or (ii) $vz < n^{1/5}(\log n)^{12}$, or (iii) $vz^3 < (\log n)^7$, or (iv) $v_1 > n^{1/5}(\log n)^9$, or (v) $m > 3v$, or (vi) $m \notin [Anz^2, 2Cnz^2]$ or $\mathcal{T}_L \wedge \mathcal{T}_z < n$ (unlikely).

Now Lemma 19 shows that **whp** (130) and (131) hold at time $\mathcal{T}_2 - 1$. In which case (assuming $\mathcal{T}_2 < \mathcal{T}_L$) $v(\mathcal{T}_2) \geq v(\mathcal{T}_2 - 1) - \log n \approx Cnz(\mathcal{T}_2 - 1)^2 = \Omega(n^{3/5}(\log n)^4)$. Similarly for $m(\mathcal{T}_2)$. Also, $z(\mathcal{T}_2) = z(\mathcal{T}_2 - 1) + O(\log n/v) \approx z(\mathcal{T}_2 - 1) = \Omega(n^{-1/5}(\log n)^2$. Hence **whp** (130) and (131) hold at time $\mathcal{T}_2$. Furthermore, $v(\mathcal{T}_2)z(\mathcal{T}_2)^3 = \Omega((\log n)^{10})$ and then (iii) will not hold on exit. (130) and (131) rule out (v),(vi) and (152) rules out (iv). So we are left with two possible exit cases having a significant probability.

Case 1: $vz < n^{1/5}(\log n)^{12}$.

The number of isolated vertices that are created from $\mathcal{T}_2$ onwards is bounded by the sum of (i) $\sum_{k\geq 3} v_k(\mathcal{T}_2)$, (ii) the number κ_1 of components of $G(\mathcal{T}_2)$ which are paths, and (iii) the number κ_2 of components of $G(\mathcal{T}_2)$ which are cycles. It follows from (7) that **whp** $\sum_{k\geq 3} v_k(\mathcal{T}_2) \approx v(\mathcal{T}_2)z(\mathcal{T}_2)/3 = O(n^{1/5}(\log n)^{12})$. Also, $\kappa_1 \leq v_1(\mathcal{T}_2) = O(n^{1/5}(\log n)^9)$. Finally, if $v_2 = v_2(\mathcal{T}_2)$, then

$$\begin{aligned}
\mathbf{E}(\kappa_2 \mid v_2, \mathbf{v}(\mathcal{T}_2)) &= \sum_{k=3}^{v_2} \binom{v_2}{k} \frac{(k-1)!}{2} \frac{2^k}{(2m)_k} \\
&\leq \sum_{k=3}^{n} \frac{1}{k} \left(\frac{v_2}{m}\right)^k \\
&\leq \log n.
\end{aligned}$$

The upper bound holds after removing the conditioning on v_2. Thus in this case, the expected number of isolated vertices created from $\mathcal{T}_2$ onwards is $O(n^{1/5}(\log n)^{12})$. This completes the proof of (112) and (113) for this case.

Case 2: $z < n^{-1/5}(\log n)^2$.

In this case we exit W_2 with

$$\begin{aligned}
z &\approx n^{-1/5}(\log n)^2 \\
v &\approx Cn^{3/5}(\log n)^4 \\
m &\approx Cn^{3/5}(\log n)^4 \\
v_1 &\leq n^{1/5}(\log n)^9.
\end{aligned} \tag{167}$$

Let

$$W_3 = \{\mathbf{v} :\ z \leq \Theta z(\mathcal{T}_2),\, vz \geq n^{1/5}(\log n)^{12},\, v_1 \leq 2n^{1/5}(\log n)^9\} \subseteq \tilde{W},$$

where Θ is as defined in Lemma 16.

Then let

$$\mathcal{T}_3 = \min \begin{cases} \mathcal{T}_2 < t \leq \mathcal{T}_L & \mathbf{v}(t) \notin W_3 \qquad \text{if such } t \text{ exist.} \\ n & \text{otherwise} \end{cases}$$

We prove

Lemma 20

$$\mathbf{Pr}(\exists \mathcal{T}_2 \leq t \leq \mathcal{T}_3 :\ v_1(t) \geq v_1(\mathcal{T}_2) + n^{1/5}(\log n)^9) = O(n^{-3}).$$

We observe first that $z(\mathcal{T}_2)$ is small enough that Lemma 5 implies

$$\mathbf{Pr}(\Delta(G(\mathcal{T}_2)) \geq 30) = O(n^{-4}). \tag{168}$$

Also, (7) implies that **qs** if $\mathbf{v} \in W_3$ then

$$\begin{aligned} v_2 &= v + O(vz + v^{1/2}\log n) \\ v_3 &= \frac{vz}{3} + O(vz^2 + (vz)^{1/2}\log n) \\ \sum_{k\geq 4} v_k &= O(vz^2 + (v^{1/2}z + 1)\log n)). \end{aligned}$$

For $t \in [\mathcal{T}_2, \mathcal{T}_3]$ let $X_t = v_1(t+1) - v_1(t)$ if $v_1(t) > 0$ and $|v_1(t+1) - v_1(t)| \leq 30$. Otherwise let $X_t = 0$. Let $\delta = n^{-1/5}$. We show that exists a constant $\gamma > 0$ such that

$$\mathbf{E}(e^{\delta X_t} \mid A(t)) \leq \exp\{\gamma(\delta z^2 + \delta^2 z + \delta^4 + \delta v^{-1} + \delta v^{-1/2} z)\log n\}. \tag{169}$$

Here we condition on the degree sequence $A(t)$ of $G(t)$ and assume it satisfies (7). Let $p_k = \mathbf{Pr}(X_t = k \mid A(t))$. Then $p_k = 0$ for $k > 30$ and

$$p_k = \begin{cases} O\left(\frac{v_1}{m}\right) & k \leq -2 \\ \frac{3v_2v_3}{2m^2} + O((z^2 + v^{-1} + v^{-1/2}z)\log n) & k = -1 \\ \frac{v_2^2}{m^2} + O((z^2 + v^{-1} + v^{-1/2}z)\log n) & k = 0 \\ \frac{3v_2v_3}{2m^2} + O((z^2 + v^{-1} + v^{-1/2}z)\log n) & k = 1 \\ O((z^2 + v^{-1} + v^{-1/2}z)\log n) & k \geq 2. \end{cases} \tag{170}$$

Consider for example p_{-1}: $O(v^{-1})$ (see Lemma 6) accounts for the probability that there is a loop or multiple edge within distance two of x. $O(v_1 z/m) = o(z^2 \log n)$ accounts for the case of y being of degree at least 3 and having a degree 1 neighbour other than x. Excluding these cases, we lose one vertex of degree one if y is of degree two and its neighbour (other than x) is of degree at least three. Thus

$$\begin{aligned} p_{-1} &= \frac{2v_2}{2m-1}\frac{3v_3 + 4v_4 + \cdots + 30v_{30}}{2m-3} + O(v^{-1} + z^2\log n) \\ &= \frac{3v_2v_3}{2m^2} + O((z^2 + v^{-1} + v^{-1/2}z)\log n). \end{aligned}$$

The other probabilities are computed similarly.

It follows that the moments

$$\mathbf{E}(X_t^i \mid A(t)) \leq \begin{cases} O((z^2 + v^{-1} + v^{-1/2}z)\log n) & i = 1, 3 \\ O(z + (v^{-1} + v^{-1/2}z)\log n) & i = 2. \end{cases} \tag{171}$$

So,

$$\begin{aligned} \mathbf{E}(e^{\delta X_t} \mid A(t)) &= \sum_{k=-30}^{30} p_k e^{\delta k} \\ &= \sum_{k=-30}^{30} p_k \left(1 + \delta k + \frac{\delta^2 k^2}{2} + \frac{\delta^3 k^3}{6} + O(\delta^4)\right) \\ &= 1 + O((\delta z^2 + \delta^2 z + \delta^4 + \delta v^{-1} + \delta v^{-1/2} z)\log n), \end{aligned}$$

and (169) follows.

Removing the conditioning $A(t)$ does not change the validity of the upper bound in (169). Thus for $\mathbf{v}(t) \in W_3$,

$$\mathbf{E}(e^{\delta X_t} \mid \{\mathbf{v}(s)\}_{t_0 \leq s \leq t}) \leq \exp\{\gamma(\delta z^2 + \delta^2 z + \delta^4 + \delta v^{-1} + \delta v^{-1/2} z)\log n\}.$$

Fix $t_0 \in [\mathcal{T}_2, \mathcal{T}_3]$. We define the stopping time

$$\mathcal{T}' = \min \begin{cases} t_0 \le t \le \mathcal{T}_3 \wedge \mathcal{T}_z : & v_1(t) = 0 \qquad \text{if such } t \text{ exist.} \\ \mathcal{T}_3 & \text{otherwise.} \end{cases}$$

For $t \ge t_0$ define $Y_t = \sum_{\tau=t_0}^{t} X_\tau$ and

$$S_t = \exp\{\delta Y_t - \gamma \log n \sum_{\tau=t_0}^{t} (\delta z(\tau)^2 + \delta^2 z(\tau) + \delta^4 + \delta v^{-1} + \delta v^{-1/2} z(\tau))\}.$$

Then

$$\begin{aligned} \mathbf{E}(S_t \mid \{\mathbf{v}(\tau)\}_{t_0 \le \tau \le t}) &= \\ \mathbf{E}(S_{t-1} \exp\{\delta X_t - \gamma(\delta z^2 + \delta^2 z + \delta^4 + \delta v^{-1} + \delta v^{-1/2} z) \log n\} \mid \{\mathbf{v}(\tau)\}_{t_0 \le \tau \le t}) &= \\ S_{t-1} \mathbf{E}(\exp\{\delta X_t - \gamma(\delta z^2 + \delta^2 z + \delta^4 + \delta v^{-1} + \delta v^{-1/2} z) \log n\} \mid \{\mathbf{v}(\tau)\}_{t_0 \le \tau \le t}) &\le S_{t-1}. \end{aligned}$$

Thus the sequence S_t, $t_0 \le t \le \mathcal{T}'$ is a supermartingale. Applying the maximum inequality, we see that for any $\Lambda > 0$,

$$\mathbf{Pr}(\max_{t_0 \le t \le \mathcal{T}'} S_t \ge \Lambda) \le \frac{1}{\Lambda}.$$

Putting $\Lambda = e^{\delta L}$, $L = n^{1/5} (\log n)^2$ we obtain

$$\mathbf{Pr}(\exists t_0 \le t \le \mathcal{T}' : \; Y_t \ge L + \gamma \log n \sum_{\tau=t_0}^{t} (z(\tau)^2 + \delta z(\tau) + \delta^3 + v(\tau)^{-1} + \delta v(\tau)^{-1/2} z(\tau)) \le e^{-\delta L}.$$

Now observe that $t - t_0 \le v(\mathcal{T}_2) \le (1 + o(1)) C n^{3/5} (\log n)^4$, $z(\tau) \le \Theta z(\mathcal{T}_2)$ and $v(\tau)^{-1} \le z(\tau) n^{-1/5} (\log n)^{-12}$. So

$$\sum_{\tau=t_0}^{t} (z(\tau)^2 + \delta z(\tau) + v(\tau)^{-1} + \delta^3 + \delta v(\tau)^{-1/2} z(\tau)) \le 2C n^{1/5} (\log n)^8.$$

and

$$\mathbf{Pr}(\exists t_0 \le t \le \mathcal{T}' : \; Y_t \ge 3C\Theta^2 \gamma n^{1/5} (\log n)^8) \le n^{-\log n}.$$

Observe next that (168) implies that

$$\mathbf{Pr}(\exists t \ge t_0 : \; Y_t \ne v_1(t) - v_1(t_0) \mid \mathbf{v}(t_0)) = O(n^{-4}).$$

So from time t_0 to $\mathcal{T}'$ we have $v_1(t) - v_1(t_0) = O(n^{1/5} (\log n)^8)$ with probability at least $1 - O(n^{-4})$. Since $v_1(t)$ returns to 0 at most n times in $[\mathcal{T}_2, \mathcal{T}_3]$, the probability that $v_1(t) - v_1(t_0) \ge n^{1/5} (\log n)^9$ for some time in $[\mathcal{T}_2, \mathcal{T}_3]$ is at most $O(n^{-3})$. □

The proof of (112) has now been completed for this case. To finish the proof of the upper bound in Theorem 3 we must verify (113) as well. But we see that **whp** on exit from W_3 we have $vz < n^{1/5} (\log n)^{12}$ and $v_1 \le 2n^{1/5} (\log n)^9$. The number of isolated vertices that are created from now on is bounded as in Case 1, completing the proof of the upper bound.

6.1 Lower Bound

We fix a time $t_0 \in [\mathcal{T}_1', \mathcal{T}_2]$ such that

$$\begin{aligned} z(t_0) &\approx n^{-1/5}(\log n)^5, \\ v(t_0) &\approx m(t_0) \approx Cn^{3/5}(\log n)^{10}. \end{aligned} \tag{172}$$

The existence of t_0 follows from the fact that in $[\mathcal{T}_1', \mathcal{T}_2]$ z drops from $\approx n^{-1/100}$ to $\approx n^{-1/5}(\log n)^2$ in steps bounded by $O(v^{-1}\log n)$.

Next let $T_1 = n^{3/5}/(\log n)^{20} = o(\log n/z(t_0)^3)$. We show that the expected number of isolated vertices created between t_0 and $t_1 = t_0 + T_1$ is $\Omega(n^{1/5}/(\log n)^{75/2})$.

Let $X_t = v_1(t+1) - v_1(t)$. When $v_1(t) > 0$ we decompose X_t as

$$X_t = \delta_t Y_t + \mathrm{Err}_t$$

where

$$\delta_t = \begin{cases} 0 & \text{probability } (v_2/m)^2 - Lz^2\log n \\ 1 & \text{probability } 1 - ((v_2/m)^2 - Lz^2\log n) \end{cases},$$

$$Y_t = \begin{cases} 0 & \delta_t = 0 \text{ or probability } 2Lz^2\log n \text{ when } \delta_t = 1 \\ +1 & \text{probability } 1/2 - Lz^2\log n \text{ when } \delta_t = 1 \\ -1 & \text{probability } 1/2 - Lz^2\log n \text{ when } \delta_t = 1; \end{cases}$$

here L is some large positive constant, large enough that – see (170) –

$$Lz^2\log n \geq 1 - (p_0 + p_1 + p_{-1}).$$

Note that if (172) holds then **whp** $z^2 \gg v^{-1}, v_1/m, z/v^{1/2}$. Note further that **qs** (see (7))

$$\frac{v_2^2}{m^2} = 1 - (1+o(1))z(t), \tag{173}$$

and that (161) implies **whp**

$$z(t) \approx z(t_0) \qquad \text{for } t \in [t_0, t_1].$$

Err_t is simply $X_t - \delta_t Y_t$.

In the case of $v_1(t) = 0$ a simple calculation shows that

$$\mathbf{Pr}(v_1(t+1) = 0 \mid v_1(t) = 0) = O(z^2) = o(1).$$

In this case we take $\mathrm{Err}_t = 0$ and put $\mathbf{Pr}(\delta_t = 0) = \mathbf{Pr}(v_1(t+1) = 0$.

Now $\mathbf{Pr}(\Delta(G(t_0)) \geq 30) = O(n^{-4})$ – see (168) – and we deduce that

$$\mathbf{E}\left(\sum_{t=t_0}^{t_1} \mathrm{Err}_t\right) = O(T_1 z(t_0)^2 \log n).$$

Now,

$$\mathbf{E}\left(\sum_{t=t_0}^{t_1} v_1(t)\right) \geq \mathbf{E}\left(\sum_{t=t_0}^{t_1}\sum_{t'=t_0}^{t-1} X_{t'}\right)$$

$$
\begin{aligned}
&= \mathbf{E}\left(\sum_{t=t_0}^{t_1}\sum_{t'=t_0}^{t-1}\delta_{t'}Y_{t'}\right) + \mathbf{E}\left(\sum_{t=t_0}^{t_1}\sum_{t'=t_0}^{t-1}\mathrm{Err}_{t'}\right)\\
&= \mathbf{E}\left(\sum_{j\in J}Y_{\tau_j}(t_1-\tau_j)\right) + O(T_1^2z(t_0)^2\log n)
\end{aligned}
$$

where $J = \{\tau \in [t_0,t_1]:\ \delta_t = 1\} = \{\tau_1 < \tau_2 < \cdots\}$,

$$
\geq \mathbf{E}\left(\sum_{j=0}^{T_1z(t_0)/3}Y_{\tau_j}(t_1-\tau_j)\right) + O(n^{4/5}(\log n)^{-29}) \tag{174}
$$

We see from (173) that $\mathbf{E}(|J|) \approx T_1z(t_0)$ and we can show as we did with b in (160) that **qs** $|J| \geq T_1z(t_0)/2$. Similarly, $\mathbf{E}(\tau_j - \tau_{j-1}) \approx 1/z(t_0)$ and again it is not hard to show that **qs** $t-1-\tau_j \geq T_1/4$ for all $j \leq T_1z(t_0)/3$. We deduce then from (174) that

$$
\mathbf{E}\left(\sum_{t=t_0}^{t_1}v_1(t)\right) \geq \mathbf{E}\left(\sum_{j=0}^{T_1z(t_0)/3}Y_{\tau_j}\right)T_1/5 + O(n^{4/5}(\log n)^{-29}). \tag{175}
$$

Now $\sum_{j=0}^{r}Y_{\tau_j}$ dominates the distance from the origin of a simple random walk of length $B(r, 1-3Lz(t_0)^2\log n)$ and so

$$
\begin{aligned}
\mathbf{E}\left(\sum_{j=0}^{T_1z(t_0)/3}Y_{\tau_j}\right) &= \Omega((T_1z(t_0))^{1/2})\\
&= \Omega(n^{4/5}(\log n)^{-55/2}).
\end{aligned}
$$

Substituting this in (175) yields

$$
\mathbf{E}\left(\sum_{t=t_0}^{t_1}v_1(t)\right) = \Omega(n^{4/5}(\log n)^{-55/2}).
$$

It follows from Lemma 6 and (172) that where $Z_t = v_0(t+1) - v_0(t) - 2$,

$$
\begin{aligned}
\mathbf{E}\left(\sum_{t=t_0}^{t_1}Z_t\right) &= \Omega\left(\mathbf{E}\left(\sum_{t=t_0}^{t_1}\frac{v_1(t)-1}{m(t)}\right)\right)\\
&= \Omega(n^{1/5}(\log n)^{-75/2}),
\end{aligned}
$$

and this completes the proof of Theorem 3. □

Acknowledgement We are very grateful to the referees whose numerous critical comments helped us to make the paper more readable.

References

[1] J.Aronson, M.E.Dyer, A.M.Frieze and S.Suen, *Randomised greedy matching II*, Random Structures and Algorithms **6**, (1995) 55-74.

[2] J.Aronson, *Analysis of a randomized greedy matching algorithm*, Ph.D. Thesis, Carnegie Mellon University, 1995.

[3] B.Bollobás, Random Graphs, Academic Press, 1984.

[4] B.Bollobás, *Martingales, isoperimetric inequalities and random graphs*, Colloq. Math. Soc. János Bolyai **52** (1987) 113-139.

[5] B.Bollobás and G.Brightwell, *The width of random graph orders*, to appear.

[6] K.L.Chung, A course in probability theory, Academic Press, New York, 1974.

[7] R.Durrett, Probability: Theory and examples, Wadsworth and Brooks/Cole, 1991.

[8] M.E.Dyer and A.M.Frieze, *Randomised greedy matching*, Random Structures and Algorithms **2**, (1991) 29-46.

[9] M.E.Dyer, A.M.Frieze and B.Pittel, *On the average performance of the greedy algorithm for finding a matching in a graph*, Annals of Applied Probability **3**, (1993) 526-552.

[10] J. Edmonds, *Paths, Trees, and Flowers*, Canadian Journal of Mathematics **17**, 1965, pp. 449-467.

[11] A.M.Frieze, *On large matchings and cycles in sparse random graphs*, Discrete Mathematics **59**, (1986) 243-256.

[12] R. M. Karp and M. Sipser, *Maximum matchings in sparse random graphs*, Proceedings of the 22nd Annual IEEE Symposium on Foundations of Computing (1981) 364-375.

[13] D. E. Knuth, R. Motwani and B. Pittel, *Stable husbands*, Random Structures and Algorithms **1**, (1990) 1-14.

[14] B. Korte and D. Hausmann, *An analysis of the greedy algorithm for independence systems* in Algorithmic Aspects of Combinatorics, B. Alspach, P. Hall and D. J. Miller Eds., *Annals of Discrete Mathematics* **2** (1978) 65–74.

[15] T.G.Kurtz, *Solutions of ordinary differential equations as limits of pure Markov jump processes*, Journal of Applied Probability **7** (1970) 49-58.

[16] C.J.H.McDiarmid, *On the method of bounded differences*, Surveys in Combinatorics, 1989, (J.Siemons ed.), Cambridge University Press (1989) 148-188.

[17] S.Micali and V.Vazirani, *An* $O(|V|^{1/2}|E|)$ *algorithm for finding maximum matching in general graphs*, Proceedings of 21st IEEE Symposium on Foundations of Computing (1980).

[18] B.Pittel, *Paths in a random digital tree:limiting distributions*, Advances in Appied. Probabilty **18** (1986), 139-155.

[19] B.Pittel, *On tree census and the giant component in sparse random graphs*, Random Structures and Algorithms **1** (1990), 311-342.

[20] B.Pittel, J.Spencer and N.C.Wormald, *Sudden emergence of a giant* k*-core in a random graph*, J. Comb. Theory B **67** (1996), 111-151.

A Proof of (5) and (6)

To find a sharp estimate for the probabilities in (5) and (6), we have to refine a bit the proof of the local limit theorem, since in our case the variance of Z is not bounded away from zero. However it is enough to consider the case where $v\sigma^2$ tends to infinity. As usual, we start with the inversion formula

$$\begin{aligned}
\mathbf{Pr}\left(\sum_\ell Z_\ell = \tau\right) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\tau x}\mathbf{E}\left(e^{ix\sum_\ell Z_\ell}\right) dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\tau x}\left[\mathbf{E}(e^{ixZ})\right]^v dx, \qquad (176)
\end{aligned}$$

where $\tau = s - k$. Consider first $|x| \geq (vz)^{-5/12}$. Using an inequality (see Pittel [19])

$$|f(\eta)| \leq e^{(\mathrm{Re}\eta - |\eta|)/3} f(|\eta|),$$

we estimate

$$\begin{aligned}
\frac{1}{2\pi}\int_{|x|\geq(vz)^{-5/12}} \left| e^{-i\tau x}\left(\frac{f(e^{ix}z)}{f(z)}\right)^v \right| dx &\leq \frac{1}{2\pi}\int_{|x|\geq(vz)^{-5/12}} e^{vz(\cos x - 1)/3}\, dx \\
&\leq e^{vz[(\cos((vz)^{-5/12}) - 1)/3]} \\
&\leq e^{-(vz)^{1/6}/9}. \qquad (177)
\end{aligned}$$

For $|x| \leq (vz)^{-5/12}$, putting $\eta = ze^{ix}$ and using $vzf'(z)/f(z) = s$, $d/dx = i\eta d/d\eta$ we expand as a Taylor series around $x = 0$ to obtain

$$\begin{aligned}
-i\tau x + v\log\left(\frac{f(e^{ix}z)}{f(z)}\right) &= ikx - \frac{vx^2}{2}\,\mathcal{D}\left(\frac{\eta f'(\eta)}{f(\eta)}\right)\Bigg|_{\eta=z} \\
&\quad - \frac{ivx^3}{3!}\,\mathcal{D}^2\left(\frac{\eta f'(\eta)}{f(\eta)}\right)\Bigg|_{\eta=z} \\
&\quad + O\left[vx^4\,\mathcal{D}^3\left(\frac{\eta f'(\eta)}{f(\eta)}\right)\Bigg|_{\eta=\tilde{\eta}}\right]; \qquad (178)
\end{aligned}$$

here $\tilde{\eta} = ze^{i\tilde{x}}$, with $\tilde{x}$ being between 0 and x, and $\mathcal{D} = \eta(d/d\eta)$. Now, the coefficients of $vx^2/2$, $vx^3/3!$ and vx^4 are $\mathbf{Var}(Z)$, $O(\mathbf{Var}(Z))$, $O(\mathbf{Var}(Z))$ respectively, and $\mathbf{Var}(Z)$ is of order z. (Use (39) and consider the effect of $\mathcal{D}$ on a power of z.) So the second and the third terms in (178) are $o(1)$ uniformly for $|x| \leq (vz)^{-5/12}$. Therefore

$$\frac{1}{2\pi}\int_{|x|\leq(vz)^{-5/12}} = \int_1 + \int_2 + \int_3, \qquad (179)$$

where

$$\begin{aligned}
\int_1 &= \frac{1}{2\pi}\int_{|x|\leq(vz)^{-5/12}} e^{ikx - v\mathbf{Var}(Z)x^2/2}\, dx \\
&= \frac{1}{\sqrt{2\pi v\mathbf{Var}(Z)}} + O\left(\frac{k^2+1}{(zv)^{3/2}}\right), \qquad (180) \\
\int_2 &= O\left(\frac{iv}{3!}\mathcal{D}^2\left(\frac{zf'(z)}{f(z)}\right)\int_{|x|\leq(vz)^{-5/12}} x^3 e^{-v\mathbf{Var}(Z)x^2/2}\, dx\right)
\end{aligned}$$

$$
\begin{aligned}
&= O\left(vz\int_{|x|\geq(vz)^{-5/12}}|x|^3e^{-v\mathbf{Var}(Z)x^2/2}\,dx\right)\\
&= O(e^{-\alpha(vz)^{1/6}}), \qquad (181)
\end{aligned}
$$

($\alpha > 0$ is an absolute constant), and

$$
\begin{aligned}
\int_3 &= O\left[vz\int_{|x|\leq(vz)^{-5/12}}x^4e^{-v\mathbf{Var}(Z)x^2/2}\,dx\right]\\
&= O\left(\frac{1}{(vz)^{3/2}}\right). \qquad (182)
\end{aligned}
$$

Using (176)-(182), we arrive at

$$
\mathbf{Pr}\left(\sum_{\ell} Z_\ell = \tau\right) = \frac{1}{\sqrt{2\pi v\mathbf{Var}(Y)}} \times \left[1 + O\left(\frac{k^2+1}{vz}\right)\right].
$$