

The cover time of random digraphs

Colin Cooper*

Alan Frieze[†]

February 12, 2008

Abstract

We study the cover time of a random walk on the random digraph $D_{n,p}$ when $p = \frac{d \log n}{n}$, $d > 1$. We prove that **whp** the cover time is asymptotic to $d \log \left(\frac{d}{d-1} \right) n \log n$.

1 Introduction

Let $D = (V, E)$ be a strongly connected digraph with $|V| = n$, and $|E| = m$. For $v \in V$ let C_v be the expected time taken for a simple random walk $\mathcal{W}_v = (\mathcal{W}_v(t), t = 0, 1, \dots)$ on D starting at v , to visit every vertex of D . The *cover time* C_D of D is defined as $C_D = \max_{v \in V} C_v$.

For connected undirected graphs, the cover time is well understood, and has been intensely studied. It is an old result of Aleliunas, Karp, Lipton, Lovász and Rackoff [2] that $C_G \leq 2m(n-1)$. It was shown by Feige [8], [9], that for any connected graph G , the cover time satisfies $(1 - o(1))n \log n \leq C_G \leq (1 + o(1))\frac{4}{27}n^3$.¹ As an example of a graph achieving the lower bound, the complete graph K_n has cover time determined by the Coupon Collector problem. The *lollipop* graph consisting of a path of length $n/3$ joined to a clique of size $2n/3$ gives the asymptotic upper bound for the cover time.

In a sequence of papers we have investigated the cover time of various classes of random graph. The main results of these papers can be summarised as follows:

- [4] If $p = d \log n/n$ and $d > 1$ then **whp** $C_{G_{n,p}} \sim d \log \left(\frac{d}{d-1} \right) n \log n$.

*Department of Computer Science, King's College, University of London, London WC2R 2LS, UK

[†]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, USA. Supported in part by NSF grant CCF0502793.

¹Here log refers to natural logarithms, unless explicitly indicated otherwise.

- [7] Let $d > 1$ and let x denote the solution in $(0, 1)$ of $x = 1 - e^{-dx}$. Let X_g be the giant component of $G_{n,p}$, $p = d/n$. Then **whp** $C_{X_g} \sim \frac{dx(2-x)}{4(dx-\log d)}n(\log n)^2$.
- [5] If $G_{n,r}$ denotes a random r -regular graph on vertex set $[n]$ with $r \geq 3$ then **whp** $C_{G_{n,r}} \sim \frac{r-1}{r-2}n \log n$.
- [6] If G_m denotes a *preferential attachment graph* of average degree $2m$ then **whp** $C_{G_m} \sim \frac{2m}{m-1}n \log n$.

In this paper we turn our attention to random directed graphs. For digraphs, the cover time is less well understood, and there are strongly connected digraphs with cover time exponential in n . An example of this is the digraph consisting of a directed cycle $(1, 2, \dots, n, 1)$, and edges $(j, 1)$, from vertices $j = 2, \dots, n-1$. Starting from vertex 1, the expected time for a random walk to reach vertex n is $2^n - 1$.

Let $D_{n,p}$ be the random digraph with vertex set $V = [n]$ where each possible directed edge (i, j) , $i \neq j$ is included with probability p . It is known that if $np = d \log n + \omega$ where $\omega = (d-1) \log n \rightarrow \infty$ then $D_{n,p}$ is strongly connected **whp**. (If ω as defined tends to $-\infty$ then **whp** $D_{n,p}$ is not strongly connected). We discuss the covertime of $D_{n,p}$ for p at or above the strong connectivity threshold.

Theorem 1. *Suppose that $np = d \log n = \log n + \gamma$ where $\gamma = (d-1) \log n \rightarrow \infty$. Then **whp***

$$C_{D_{n,p}} \sim d \log \left(\frac{d}{d-1} \right) n \log n.$$

Note that if $d = d(n) \rightarrow \infty$ with n then we have $C_{D_{n,p}} \sim n \log n$.

Our analysis is based on Lemma 3 below, which is proved in its current state in [7]. In order to apply this lemma we need to have estimates of the stationary distribution $\pi(v)$, $v \in V$ of our walk and the mixing time. For an undirected graph G , the stationary distribution is trivial, we just take $\pi(v) = \deg(v)/2m$ where \deg denotes degree and m is the number of edges in G . For strong digraphs where each vertex v has in-degree equal to out-degree ($\deg^-(v) = \deg^+(v)$), then $\pi(v) = \deg^+(v)/m$. For general digraphs, however, there is no such simple formula, and the main technical task of this paper is to find good estimates for π in the case of $D_{n,p}$. We summarise our result concerning the steady state as follows:

Theorem 2. *Suppose that $np = d \log n = \log n + \gamma$ where $\gamma = (d-1) \log n \rightarrow \infty$. Then **whp***

$$\pi_v \sim \frac{\deg^-(v)}{m} \quad \forall v \in V$$

where $\deg^-(v)$ is the in-degree of v and m is the number of edges of $D_{n,p}$.

Note that if $d = d(n) \rightarrow \infty$ with n then the steady state distribution is asymptotically uniform.

Once we know the steady state distribution, it is basically plain sailing. We can plug into the program from the previously cited papers.

The next section describes Lemma 3. Section 3 deals with estimating the steady state distribution. We then estimate the cover time asymptotically in Section 5.

2 Main Lemma

In this section D denotes a fixed strongly connected digraph with n vertices. A random walk \mathcal{W}_u is started from a vertex u . Let $\mathcal{W}_u(t)$ be the vertex reached at step t , let P be the matrix of transition probabilities of the walk and let $P_u^{(t)}(v) = \mathbf{Pr}(\mathcal{W}_u(t) = v)$. We assume the random walk \mathcal{W}_u on G is ergodic with steady state distribution π .

Let

$$d(t) = \max_{u,x \in V} |P_u^{(t)}(x) - \pi_x|,$$

and let T be such that, for $t \geq T$

$$\max_{u,x \in V} |P_u^{(t)}(x) - \pi_x| \leq n^{-3}. \quad (1)$$

Fix two vertices u, v . Considering the walk \mathcal{W}_v , starting at v , let $r_t = \mathbf{Pr}(\mathcal{W}_v(t) = v)$ be the probability that this walk returns to v at step $t = 0, 1, \dots$. Let

$$R_T(z) = \sum_{j=0}^{T-1} r_j z^j \quad (2)$$

and

$$\lambda = \frac{1}{KT} \quad (3)$$

for some sufficiently large constant K .

An almost identical lemma was first proved in [5]. For $t \geq T$ let $\mathbf{A}_t(v)$ be the event that \mathcal{W}_u does not visit v in steps $T, T+1, \dots, t$.

Lemma 3. *Suppose that*

(a) *For some constant $\theta > 0$, we have*

$$\min_{|z| \leq 1+\lambda} |R_T(z)| \geq \theta.$$

(b) $T^2\pi_v = o(1)$ and $T\pi_v = \Omega(n^{-2})$.

Let

$$p_v = \frac{\pi_v}{R_T(1)(1 + O(T\pi_v))}, \quad (4)$$

where $R_T(1)$ is from (2).

Then for all $t \geq T$,

$$\Pr(\mathbf{A}_t(v)) = \frac{(1 + O(T\pi_v))}{(1 + p_v)^t} + o(e^{-\lambda t/2}). \quad (5)$$

We will first prove Theorem 1 under

Assumption 1: $d - 1 = \Omega(1)$ i.e. d is at least a positive constant strictly larger than one.

3 Estimating the Stationary Distribution

3.1 Two Useful Lemmas

We first give a simple lemma concerning the degree sequence of $D_{n,p}$. It can easily be proven by the use of the first and second moment method. Let \deg^+ denote out-degree and \deg^- denote in-degree.

We deal with two cases. In Case (a) we have $d = \omega(n) \rightarrow \infty$ and in Case (b) we have $d > 1$ a fixed constant. In the following lemma there will be two intervals I_a, I_b . We define events

$$\begin{aligned} \mathcal{D}^+(v) &= \deg^+(v) \in I, \\ \mathcal{D}^-(v) &= \deg^-(v) \in I. \end{aligned}$$

Here $I = I_a$ or I_b depending on the context. For a set of vertices S let

$$\mathcal{D}^+(S) = \bigcap_{v \in S} \mathcal{D}^+(v)$$

and let $\mathcal{D}^-(S)$ be defined analogously.

Lemma 4. *Let $\epsilon_1 = \omega^{-1/3}$, $c_a = 1 - \epsilon_1$, $C_a = 1 + \epsilon_1$, $c_b = d'/2$ and $C_b = d'' + 1$ where $d' < d''$ are the roots of $x \log(e/x) = 1 - \log n/np$. Then,*

(a) *If $np = \omega \log n$ let $I_a = [c_a np, C_a np]$. Then*

$$\Pr \left(\bigcap_{v \in V} \mathcal{D}^+(v) \cap \bigcap_{v \in V} \mathcal{D}^-(v) \right) = \rho_a = 1 - O(n^{-K})$$

where $K > 0$ can be arbitrarily large.

(b) If $np = d \log n$ where $d = O(1)$ let $I_b = [c_b np, C_b np]$. Then

1.

$$\Pr \left(\bigcap_{v \in V} \mathcal{D}^+(v) \cap \bigcap_{v \in V} \mathcal{D}^-(v) \right) = \rho_b = 1 - O(n^{-\psi})$$

where $\psi = \psi(d) > 0$.

2. For $k \in I_b$ there are $\leq 2(nep/k)^k n^{1-d}$ vertices v with $\deg^-(v) = k$.

3. Let $k^* = (d-1) \log n$. There are $\sim (nep/k^*)^{k^*} n^{1-d} = n^{\gamma_d + o(1)}$ vertices v with $\deg^-(v) = k^*$, where $\gamma_d = (d-1) \log(d/(d-1))$.

Furthermore,

$$\Pr(\exists v \in V : \deg^+(v) > 10d \log n \text{ or } \deg^-(v) > 10d \log n) \leq n^{-10}. \quad (6)$$

Lemma 5. Let B_I denote the binomial random variable $\text{Bin}(n-1, p) |_{\text{Bin}(n-1, p) \in I}$ ² where $I = [cnp, Cnp]$, where $c < 1 < C$ are constants such that ρ , the probability that $\text{Bin}(n-1, p) \in I = 1 - o(1/n)$. For $\xi = X, Y, Z$ let $W_0^\xi = 1$ and let

$$W_t^X = \sum_{i=1}^{B_0-\theta} \frac{A_i^X}{B_0}, \quad W_t^Y = \sum_{i=1}^{B_0-\theta} \frac{A_i^Y}{B_i}, \quad W_t^Z = \sum_{i=1}^{B_0-\theta} \frac{A_i^Z}{B_i} + \sum_{j=1}^{\theta} \frac{A_j^Z}{cnp}.$$

where the $B_i, i \geq 0$ are independent copies of B_I , the $A_i, i \geq 1$ are independent copies of W_{t-1}^ξ , and θ is a fixed non-negative integer. Then for $|\lambda| \leq M$, $M > 1$, $Mt = o(np)$, sufficiently large n , and $\xi = X, Y, Z$,

$$\mathbf{E}(e^{\lambda W_t^\xi}) \leq \exp \left\{ \lambda + \frac{L_\xi(5+\theta)|\lambda|t}{cnp} \right\}, \quad (7)$$

where $L_X = 1, L_Y = L_Z = M$.

Proof The claim is true for $t = 0$ and we proceed by induction on t . For W_t^X we have

$$\begin{aligned} \mathbf{E}(e^{\lambda W_t^X}) &= \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \mathbf{E} \left(\exp \left\{ \frac{\lambda W_{t-1}^X}{k} \right\} \right)^{k-\theta} \\ &\leq \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \exp \left\{ (k-\theta) \left(\frac{\lambda}{k} + \frac{(5+\theta)|\lambda|(t-1)}{k cnp} \right) \right\} \\ &\leq \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \exp \left\{ \lambda + \frac{\theta|\lambda|}{cnp} + \frac{(5+\theta)(t-1)|\lambda|}{cnp} \right\} \\ &\leq \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \exp \left\{ \lambda + \frac{(5+\theta)t|\lambda|}{cnp} \right\} \\ &= \exp \left\{ \lambda + \frac{(5+\theta)t|\lambda|}{cnp} \right\} \end{aligned}$$

²We will use the notation $Z |_{\mathcal{E}}$ to denote random variable Z conditioned on the occurrence of the event \mathcal{E}

completing the inductive proof of (7) for W_t^X .

For W_t^Y we have

$$\mathbf{E}(e^{\lambda W_t^Y}) = \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \prod_{i=1}^{k-\theta} \mathbf{E}(e^{\lambda A_i/B_i}), \quad (8)$$

where

$$\mathbf{E}(e^{\lambda A_i/B_i}) = \rho^{-1} \sum_{l \in I} \binom{n-1}{l} p^l q^{n-1-l} \mathbf{E}(e^{\lambda A_i/l}).$$

We estimate

$$\begin{aligned} \rho^{-1} \sum_{l \in I} l^{-1} \binom{n-1}{l} p^l q^{n-1-l} &\leq \rho^{-1} \left(1 + \frac{1}{cnp}\right) \sum_{l \in I} (l+1)^{-1} \binom{n-1}{l} p^l q^{n-1-l} \\ &= \left(1 + \frac{1}{cnp}\right) \frac{1}{np} \frac{\mathbf{Pr}(\text{Bin}(n, p) \in I)}{\mathbf{Pr}(\text{Bin}(n-1, p) \in I)} \\ &\leq \left(\frac{1}{np} + \frac{1}{c(np)^2}\right) \left(1 + p \frac{\mathbf{Pr}(\text{Bin}(n, p) = cnp - 1)}{\mathbf{Pr}(\text{Bin}(n-1, p) \in I)}\right) \\ &\leq \frac{1}{np} + \frac{1}{c(np)^2} + \frac{1}{n^2}. \end{aligned}$$

Applying induction, and using $e^x \leq 1 + x + x^2$ for $|x| \leq 1$ we see that,

$$\begin{aligned} \mathbf{E}(e^{\lambda A_i/B_i}) &\leq \rho^{-1} \sum_{l \in I} \binom{n-1}{l} p^l q^{n-1-l} \exp \left\{ \frac{\lambda}{l} + \frac{M(5+\theta)|\lambda|(t-1)}{lcnp} \right\} \\ &\leq \rho^{-1} \sum_{l \in I} \binom{n-1}{l} p^l q^{n-1-l} \left(1 + \frac{1}{l} \left(\lambda + \frac{M(5+\theta)(t-1)|\lambda|}{cnp}\right) + \frac{2\lambda^2}{l^2}\right) \\ &\leq 1 + \left(\lambda + \frac{M(5+\theta)(t-1)|\lambda|}{cnp} + \frac{2\lambda^2}{cnp}\right) \left(\frac{1}{np} + \frac{1}{c(np)^2} + \frac{1}{n^2}\right) \\ &\leq 1 + \frac{\lambda}{np} + \frac{|\lambda|M((5+\theta)t - (2+\theta))}{c(np)^2}. \end{aligned} \quad (9)$$

Plugging (9) into (8) we get

$$\begin{aligned} \mathbf{E}(e^{\lambda W_t^Y}) &\leq \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \left(1 + \frac{\lambda}{np} + \frac{|\lambda|M((5+\theta)t - (2+\theta))}{c(np)^2}\right)^{k-\theta} \\ &\leq \rho^{-1} \left(1 + \frac{\lambda}{np}\right)^{-\theta} \left(1 + \frac{\lambda}{n} + \frac{|\lambda|M((5+\theta)t - (2+\theta))}{cn^2p}\right)^n \\ &\leq \exp \left\{ \lambda + \frac{M(5+\theta)|\lambda|t}{cnp} - \frac{|\lambda|M}{cnp} \left((2+\theta) - \frac{\theta c(1+o(1/np))}{M(1-|\lambda|/np)} \right) \right\} \\ &\leq \exp \left\{ \lambda + \frac{M(5+\theta)|\lambda|t}{cnp} \right\} \end{aligned}$$

which is (7).

For W_t^Z we have

$$\mathbf{E}(e^{\lambda W_t^Z}) = \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \prod_{i=1}^{k-\theta} \mathbf{E}(e^{\lambda A_i/B_i}) \prod_{j=1}^{\theta} \mathbf{E}(e^{\lambda A_j/cnp}), \quad (10)$$

where, as in (9) we have

$$\mathbf{E}(e^{\lambda A_i/B_i}) \leq 1 + \frac{\lambda}{np} + \frac{|\lambda|M((5+\theta)t - (2+\theta))}{c(np)^2}$$

and by induction

$$\mathbf{E}(e^{\lambda A_j/cnp}) \leq \exp \left\{ \frac{\lambda}{cnp} + \frac{M(5+\theta)|\lambda|t}{(cnp)^2} \right\}.$$

Plugging into (10) we get

$$\begin{aligned} \mathbf{E}(e^{\lambda W_t^Z}) &\leq \\ &\exp \left\{ \frac{\lambda\theta}{cnp} + \frac{\theta M(5+\theta)|\lambda|t}{(cnp)^2} \right\} \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \left(1 + \frac{\lambda}{np} + \frac{|\lambda|M((5+\theta)t - (2+\theta))}{c(np)^2} \right)^{k-\theta} \\ &\leq \exp \left\{ \frac{\lambda\theta}{cnp} + \frac{\theta M(5+\theta)|\lambda|t}{(cnp)^2} \right\} \left(1 + \frac{\lambda}{np} \right)^{-\theta} \rho^{-1} \left(1 + \frac{\lambda}{n} + \frac{|\lambda|M((5+\theta)t - (2+\theta))}{cn^2p} \right)^n \\ &\leq \exp \left\{ \lambda + \frac{M(5+\theta)|\lambda|t}{cnp} - \frac{|\lambda|M}{cnp} \left((2+\theta) - \frac{2\theta(1+o(1/np))}{M(1-|\lambda|/np)} \right) \right\} \\ &\leq \exp \left\{ \lambda + \frac{M(5+\theta)|\lambda|t}{cnp} \right\} \end{aligned}$$

which is (7). □

We will apply Lemma 5 as follows. It follows from (7) that $\mathbf{E}(e^{MW_t}) \leq e^{M+\gamma M}$ and $\mathbf{E}(e^{-MW_t}) \leq e^{-M+\gamma M}$. So, by the Markov Inequality,

$$\Pr(W_t \leq 1 - A) \leq e^{-M(A-\gamma)} \quad (11)$$

$$\Pr(W_t \geq 1 + B) \leq e^{-M(B-\gamma)}. \quad (12)$$

In the following we approximate the stationary distribution $\boldsymbol{\pi}$ as follows: Iterating the equation $\boldsymbol{\pi} = \boldsymbol{\pi}P$, k times gives $\boldsymbol{\pi} = \boldsymbol{\pi}P^k$. For fixed y this gives $\pi_y = \sum_{x \in V} \pi_x P_x^{(k)}(y)$. By bounding $P_x^{(k)}(y)$ from above and below by values independent of x , i.e. $P_x^{(k)}(y) \sim \theta_y$ we obtain $\pi_y \sim \theta_y$.

3.2 Lower Bounds on the stationary distribution

To bound $P_x^{(k)}(y)$ from below, we consider random walks between x and y consisting of simple directed (x, y) -paths of length k . There are three cases;

Case I: $np = d \log n$, where $d \geq 1$ constant or $d = \omega \rightarrow \infty$ and $\omega \leq n^{3/10}$.

Case II: $n^{3/10} \leq \omega \leq n^{3/5}$.

Case III: $\omega \geq n^{3/5}$.

Most of the work is involved in the proof of the first case.

Lemma 6. *There exists an integer $s = O(\log_{np} n)$ such that **whp** for all $x, y \in V$,*

$$P_x^{(s)}(y) \geq (1 - o(1)) \frac{\deg^-(y)}{m} \text{ for all } v \in V,$$

where m is the number of edges in $D_{n,p}$.

Proof

Case I: Let

$$\ell = \left\lfloor \frac{2}{3} \log_{np} n \right\rfloor. \quad (13)$$

Fix $x \in V$ and using Breadth First Search construct the sets $X_0 = \{x\}, X_1, \dots, X_\ell$ where $X_{i+1} = N^+(X_i) \setminus (X_0 \cup \dots \cup X_i)$ for $0 \leq i < \ell$. Here $N^+(S)$ is the set of out-neighbours of set S .

Let $X = \bigcup_{i=0}^{\ell} X_i$ and let T_X denote the BFS tree constructed in this manner. If $w \in X_{i+1}$ is the out-neighbour of more than one vertex of X_i , we only keep the *first* edge into w for T_X .

Given X and y , ($y = x$ is allowed) we define $Y_0 = \{y\}, Y_1, \dots, Y_\ell$ where $Y_{i+1} = N^-(Y_i) \setminus (X \cup Y_0 \cup \dots \cup Y_i)$ for $0 \leq i < \ell$. If $w \in Y_{i-1}$ is the in-neighbour of more than one vertex of Y_i , we only keep the first edge into w for T_Y . Let $Y = \bigcup_{i=0}^{\ell} Y_i$ and let T_Y denote the BFS tree constructed in this manner.

Observe that

$$\mathcal{D}_1^+ = \mathcal{D}^+ \left(\bigcup_{i=0}^{\ell-1} X_i \right) \text{ implies } |X| \leq (Cnp)^\ell \leq n^{2/3+o(1)}. \quad (14)$$

$$\mathcal{D}_1^- = \mathcal{D}^- \left(\bigcup_{i=0}^{\ell-1} Y_i \right) \text{ implies } |Y| \leq (Cnp)^\ell \leq n^{2/3+o(1)}. \quad (15)$$

For $u \in X_i$ let P_u denote the path from x to u in T_X and

$$\alpha_{i,u} = \prod_{\substack{w \in P_u \\ w \neq u}} \frac{1}{\deg^+(w)} \leq \Pr(\mathcal{W}_x(i) = u)$$

and for $v \in Y_i$ let Q_v denote the path from v to y in T_Y and

$$\beta_{i,v} = \prod_{\substack{w \in Q_v \\ w \neq y}} \frac{1}{\deg^+(w)} \leq \Pr(\mathcal{W}_v(i) = y). \quad (16)$$

Given $D_{n,p}$ we have

$$P_x^{(2\ell+1)}(y) \geq Z = Z(x, y) = \sum_{\substack{u \in X_\ell \\ v \in Y_\ell}} \alpha_{\ell,u} \beta_{\ell,v} \frac{1_{uv}}{\deg^+(u)} \quad (17)$$

where 1_{uv} is the indicator for the existence of the edge (u, v) and we take $\frac{1_{uv}}{\deg^+(u)} = 0$ if $\deg^+(u) = 0$.

Let $\mathcal{C} = \mathcal{C}(x, y)$ denote $X_i, Y_i, 0 \leq i \leq \ell$ and the collection of edge sets $(X_{i-1} : X_i), 1 \leq i \leq \ell$ in T_X (resp. $(Y_i : Y_{i-1}), 1 \leq i \leq \ell$ in T_Y). Now,

$$\mathbf{E} \left(\frac{1_{uv}}{\deg^+(u)} \middle| \mathcal{C} \right) = \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \frac{k}{n-1} \frac{1}{k} \sim \frac{1}{n-1}.$$

(We are using the convention $\frac{0}{0} = 0$ for this expression.)

We therefore have,

$$\mathbf{E}(Z \mid \mathcal{C}) = \frac{(1 + o(1))}{n} \left(\sum_{u \in X_\ell} \alpha_{\ell,u} \right) \left(\sum_{v \in Y_\ell} \beta_{\ell,v} \right). \quad (18)$$

We will show next that for some $\epsilon_X, \epsilon_Y = o(1)$,

$$\Pr \left(\sum_{u \in X_\ell} \alpha_{\ell,u} < 1 - \epsilon_X \middle| \mathcal{D}_1^+, \mathcal{D}_1^- \right) = o(n^{-2}), \quad (19)$$

$$\Pr \left(\sum_{v \in Y_\ell} \beta_{\ell,v} < (1 - \epsilon_Y) \frac{\deg^-(y)}{m} \middle| \mathcal{D}_1^+, \mathcal{D}_1^- \right) = o(n^{-2}). \quad (20)$$

Let

$$\epsilon_X = \frac{1}{\log \log \log n}.$$

Now for each $v \in X_i$ and $j > i$ there is a unique path $v = v_i, v_{i+1}, \dots, v_j$ from v to X_j in T_X . For such a path, let

$$\gamma_{i,j;v,v_j} = \prod_{k=i}^{j-1} \frac{1}{\deg^+(v_k)}.$$

Now consider the equation

$$\sum_{u \in X_\ell} \alpha_{\ell,u} = \sum_{u \in X_\ell} \gamma_{0,\ell;x,u} = \sum_{w \in N_D^+(x)} \sum_{u \in X_{\ell,w}} \frac{\gamma_{1,\ell;w,u}}{\deg^+(x)} \quad (21)$$

where $X_{\ell,w} = \{u \in X_\ell : \text{the path from } x \text{ to } u \text{ in } T_X \text{ goes through } w \in X_1\}$ and $N_D^\pm(x)$ is the set of in/out-neighbours of x in T_X .

This leads us to claim that given \mathcal{D}_1^+ ,

$$\sum_{u \in X_\ell} \gamma_{0,\ell;x,u} \text{ (and hence } \sum_{u \in X_\ell} \alpha_{\ell,u}) \text{ dominates the random variable } \widehat{W}_\ell^X \text{ defined below.} \quad (22)$$

Let $\nu = n - n^{2/3+o(1)}$ (see (14)). Let $\widehat{W}_0^X = 1$ and let

$$\widehat{W}_t^X = \sum_{i=1}^{B_0 - B'_0} \frac{A_i}{B_0}$$

where (i) B_0 has the distribution B_I , (ii) B'_0 is the number of successes for B_0 in the first $n^{2/3+o(1)}$ trials ($o(1)$ as in (14)) and (iii) the A_i are independent copies of \widehat{W}_{t-1}^X .

We can use induction to verify (22): Going back to (21), given \mathcal{D}_1^+ , $\deg^+(x)$ is distributed as B_0 and $\gamma_{\ell,1;w,u}$ dominates $\widehat{W}_{\ell-1}^X$. We sum from 1 to $B_0 - B'_0$ in order to account for the out-neighbours of $u \in X_{\ell-t}$, $1 \leq t < \ell$ that are in $X_0 \cup X_1 \cup \dots \cup X_{\ell-t}$. We rely here on the tree structure and the fact that our BFS construction of T_X only checks edges of the form (a, b) where b has not yet been placed in X to give us the claimed independence.

Note that

$$\Pr(B'_0 \geq 100) \leq \rho^{-1} (n^{2/3+o(1)})^{100} p^{100} \leq n^{-20}. \quad (23)$$

In generating \widehat{W}_ℓ^X we will **qs**³ sample B'_0 at most $o(n^2)$ times and so if W_t^X is defined as in Lemma 5 with $\theta = 100$ then

$$\Pr(W_\ell^X \geq \widehat{W}_\ell^X) = o(n^{-2}).$$

Thus for any value σ ,

$$\Pr\left(\sum_{u \in X_\ell} \alpha_{\ell,u} \geq \sigma \mid \mathcal{D}_1^+\right) \geq \Pr(W_\ell^X \geq \sigma) - o(n^{-2}). \quad (24)$$

³An event \mathcal{E}_n occurs *quite surely* (**qs**) if $\Pr(\neg \mathcal{E}_n) = O(n^{-K})$ for any constant $K > 0$.

Using (11) with $M = 10/(\epsilon_X c)$ we have

$$\begin{aligned}
& \Pr\left(\sum_{u \in X_\ell} \alpha_{\ell,u} \leq 1 - \epsilon_X/2 \mid \mathcal{D}_1^+\right) \\
& \leq \Pr(W_\ell^X \leq 1 - \epsilon_X/2) + o(n^{-2}) \\
& \leq \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \Pr\left(\sum_{i=1}^{k-100} A_i \leq k(1 - \epsilon_X/2)\right) + o(n^{-2}) \\
& \leq \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} \exp\left\{-Mk \left(\left(1 - \frac{105(\ell-1)}{cnp}\right)(1 - 100/k) - (1 - \epsilon_X/2)\right)\right\} + o(n^{-2}) \\
& \leq \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} e^{-M\epsilon_X k/3} + o(n^{-2}) \\
& \leq n^{-3} \rho^{-1} \sum_{k \in I} \binom{n-1}{k} p^k q^{n-1-k} + o(n^{-2}) \\
& = o(n^{-2}).
\end{aligned}$$

This completes the proof of (19). We can add the extra conditioning on \mathcal{D}^- by writing

$$\Pr\left(\sum_{u \in X_\ell} \alpha_{\ell,u} < 1 - \epsilon_X \mid \mathcal{D}_1^+, \mathcal{D}_1^-\right) \leq (1 + o(1)) \Pr\left(\sum_{u \in X_\ell} \alpha_{\ell,u} < 1 - \epsilon_X \mid \mathcal{D}_1^+\right).$$

The above inequality holds because of the first of the following simple inequalities: They will only be used for cases where $\Pr(B), \Pr(C) = 1 - o(1)$.

$$\Pr(A \mid B, C) \leq \Pr(A \mid B) \frac{\Pr(B)}{\Pr(BC)}. \quad (25)$$

$$\Pr(A \mid B) \leq \Pr(A \mid B, C) + \Pr(\neg C \mid B). \quad (26)$$

We next consider the proof of (20). Now for each $v \in Y_i$ and $j < i$ with unique path $v = v_i, v_{i-1}, \dots, v_j$ from v to Y_j in T_Y , let

$$\gamma_{i,j;v,v_j} = \prod_{k=j}^{i-1} \frac{1}{\deg^+(v_k)}.$$

Consider the equation

$$\sum_{v \in Y_\ell} \beta_{\ell,v} = \sum_{v \in Y_\ell} \gamma_{\ell,0;v,y} = \sum_{w \in Y_1} \sum_{v \in Y_{\ell,w}} \frac{\gamma_{\ell,1;v,w}}{\deg^+(w)} \quad (27)$$

where $Y_{\ell,w} = \{v \in Y_\ell : \text{the path from } v \text{ to } y \text{ in } T_Y \text{ goes through } w \in Y_1\}$.

This leads us to claim that given $\mathcal{D}_1^-, \mathcal{D}_1^+$,

$$\sum_{v \in Y_\ell} \gamma_{\ell,0;v,y} \text{ (and hence } \sum_{v \in Y_\ell} \beta_{\ell,v} \text{) dominates the random variable } \widehat{W}_\ell^Y \quad (28)$$

which is defined as follows: Let $\widehat{W}_0^Y = 1$ and for $1 \leq t \leq \ell$,

$$\widehat{W}_t^Y = \sum_{i=1}^{B_0 - B'_0} \frac{A_i}{B_i}, \quad (29)$$

where (i) B_0 has the distribution B_I , (ii) B'_0 is the number of successes for B_0 in the first $n^{2/3+o(1)}$ trials (see (14),(15)), (iii) $B_i, i = 1, 2, \dots, B_0 - B'_0$ are independent with distribution B_I , and (iii) the A_i are independent copies of \widehat{W}_{t-1}^Y .

We can use induction to verify (28): Going back to (27) we see that given $\mathcal{D}_1^- \cap \mathcal{D}_1^+$ we have $|Y_1|$ dominating $B_0 - B'_0$. Then $\deg^+(w)$ is distributed as B_I and $\gamma_{\ell,1;v,w}$ dominates $\widehat{W}_{\ell-1}^Y$. B'_0 accounts for in-neighbours of $u \in Y_{\ell-t}$ that are in $X \cup Y_0 \cup \dots \cup Y_{\ell-t}$. To justify the independence of the B_i , given $\mathcal{D}_1^- \cap \mathcal{D}_1^+$, we observe that given $Y_0, Y_1, \dots, Y_{\ell-t}, \mathcal{D}_1^-, \mathcal{D}_1^+$, the out-degrees of the vertices in $Y_{\ell-t-1}$ are independent and distributed as B_I .

We can replace B'_0 by 100 as we did for W^X . In the case where $np = d \log n$, d constant, we make the following adjustment. Let $\zeta = 1/\log \log n$ and let a vertex y be *normal* if at most $\zeta_0 = \lceil 4/(\zeta^3 d) \rceil$ of its in-neighbours have out-degrees which are not in the range $[(1 - \zeta)np, (1 + \zeta)np]$. Now

$$\Pr(\exists \text{ a vertex which is not normal} \mid \mathcal{D}_1^+, \mathcal{D}_1^-) \leq n \sum_{s=c_b np}^{C_b np} \binom{s}{\zeta_0} (2e^{-\zeta^2 np/3})^{\zeta_0} = O(n^{-\log \log n}). \quad (30)$$

Let $\mathcal{N}(y)$ be the event that every vertex is normal.

We continue by proving, that

$$\Pr \left(\sum_{v \in Y_\ell} \beta_{\ell,v} \leq (1 - \epsilon_Y) \frac{\deg^-(y)}{np} \mid \mathcal{D}_1^-, \mathcal{D}_1^+ \right) = o(n^{-2}), \quad (31)$$

where ϵ_Y is defined below.

Note that $\deg^-(y) - |Y_\ell|$ is dominated by $\text{Bin}(n^{2/3+o(1)}, p)$ and so with probability $1 - O(n^{-3})$ we have $\deg^-(y) \leq |Y_\ell| + 100$.

For $np = d \log n$, d constant, let $\epsilon_Y = 2\zeta$ and $M = 10/(\epsilon_Y c)$. For $np = \omega \log n$, $\omega \rightarrow \infty$ let

$\epsilon_Y = 1/\omega^{1/3}$ and $M = 1$. Then, with $d_y^- = \deg^-(y)$ and $d_0 = cnp - \zeta_0 - 100$,

$$\begin{aligned}
& \Pr \left(\sum_{v \in Y_\ell} \beta_{\ell,v} \leq (1 - \epsilon_Y) \frac{d_y^-}{np} \middle| \mathcal{D}_1^-, \mathcal{D}_1^+ \right) \\
& \leq \Pr \left(\sum_{v \in Y_\ell} \beta_{\ell,v} \leq (1 - \epsilon_Y) \frac{d_y^-}{np} \middle| \mathcal{D}_1^-, \mathcal{D}_1^+, \mathcal{N}(y) \right) + \Pr(\neg \mathcal{N}(y) \mid \mathcal{D}_1^-, \mathcal{D}_1^+) \quad \text{see (26)} \\
& \leq \Pr \left(\sum_{i=1}^{d_y^- - \zeta_0 - 100} \frac{A_i}{(1 + \zeta)np} \leq (1 - \epsilon_Y) \frac{d_y^-}{np} \right) + o(n^{-2}) \\
& \leq \Pr \left(\sum_{i=1}^{d_y^- - \zeta_0 - 100} A_i \leq (1 - \epsilon_Y/2)d_y^- \right) + o(n^{-2}) \\
& \leq \exp \left\{ -Md_0 \left(\left(1 - \frac{105M(\ell - 1)}{cnp} \right) (1 - \zeta_0/d_0) - (1 - \epsilon_Y/2) \right) \right\} + o(n^{-2}) \\
& \leq 2e^{-M\epsilon_Y d_0/3} \\
& = o(n^{-2}).
\end{aligned}$$

This completes the proof of (20)

We now consider the concentration of $Z \mid_{\mathcal{C}}$.

For $u \in X_\ell$, and $|Y_\ell| = n^{2/3+o(1)}$, from (23) we have,

$$\Pr(|N^+(u) \cap Y_\ell| \geq 100 \mid \mathcal{D}_1^-, \mathcal{D}_1^+) \leq 2n^{-20}.$$

We write $Z = \sum_{u \in X_\ell} Z_u$ where

$$Z_u = \frac{\alpha_{\ell,u}}{\deg^+(u)} \sum_{v \in Y_\ell} \beta_{\ell,v} 1_{uv}.$$

Conditional on $\mathcal{D}_1^-, \mathcal{D}_1^+$ and $|N^+(u) \cap Y_\ell| \leq 100$ we have $Z_u \leq 100/(cnp)^{2\ell+1}$. Let $\widehat{Z}_u = (cnp)^{2\ell+1} Z_u/100$, then for $u \in X_\ell$ the \widehat{Z}_u are independent random variables, and $0 \leq \widehat{Z}_u \leq 1$. Let $\widehat{Z} = \sum_{u \in X_\ell} \widehat{Z}_u$ and note that

$$\mathbf{E}(\widehat{Z}) \geq \frac{(cnp)^{2\ell+1}}{100n} (1 - \epsilon_X)(1 - \epsilon_Y) - 2n^{-18} \geq n^{1/3-o(1)}.$$

By Hoeffding's inequality we see that,

$$\Pr(|\widehat{Z} - \mathbf{E}(\widehat{Z})| \geq 4(np\mathbf{E}(\widehat{Z}))^{1/2}) = o(n^{-4})$$

and hence that

$$\Pr(|Z - \mathbf{E}(Z)| \geq 400(np\mathbf{E}(\widehat{Z}))^{1/2}/(cnp)^{2\ell+1}) = o(n^{-4}).$$

Now $\frac{400\sqrt{np}\mathbf{E}(\widehat{Z})}{(cnp)^{2\ell+1}} = O\left(\frac{1}{n^{7/6+o(1)}}\right)$ and so this implies that

$$\Pr\left(|Z - \mathbf{E}(Z)| = O\left(\frac{1}{n^{7/6+o(1)}}\right) \mid \mathcal{D}_1^-, \mathcal{D}_1^+\right) = 1 - o(n^{-4}). \quad (32)$$

It then follows from (18), (19) and (20) that

$$\Pr\left(Z \leq (1 - o(1))\frac{\deg^-(y)}{m} \mid \mathcal{D}_1^-, \mathcal{D}_1^+\right) = o(n^{-2}). \quad (33)$$

Now let \mathcal{D}^+ be the intersection of the events \mathcal{D}_1^+ as x varies and let \mathcal{D}^- be the intersection of the events \mathcal{D}_1^- as y varies. We note that $\Pr(\mathcal{D}^+, \mathcal{D}^-) = 1 - o(1)$. Furthermore, for any pair x, y ,

$$\begin{aligned} \Pr\left(Z \leq (1 - o(1))\frac{\deg^-(y)}{m} \mid \mathcal{D}^-, \mathcal{D}^+\right) &\leq \Pr\left(Z \leq (1 - o(1))\frac{\deg^-(y)}{m} \mid \mathcal{D}_1^-, \mathcal{D}_1^+\right) \frac{\Pr(\mathcal{D}_1^-, \mathcal{D}_1^+)}{\Pr(\mathcal{D}^-, \mathcal{D}^+)} \\ &= o(n^{-2}). \end{aligned}$$

So,

$$\Pr\left(P_x^{(2\ell+1)}(y) \geq (1 - o(1))\frac{\deg^-(y)}{m}, \forall x, y \mid \mathcal{D}^-, \mathcal{D}^+\right) = 1 - o(1)$$

and

$$\Pr\left(P_x^{(2\ell+1)}(y) \geq (1 - o(1))\frac{\deg^-(y)}{m}, \forall x, y\right) = 1 - o(1). \quad (34)$$

Case II: Fix $x, y \in V$. x will **qs** have $\sim np$ out-neighbours X_+ and y will **qs** have $\sim np$ in-neighbours Y_- and **qs** $|X_+ \cap Y_-| \sim np^2$. Given this, the Chernoff bound shows that **qs** there are at least $(1 - o(1))n^2p^3$ paths of length three joining x to y . Therefore, for any $x, y \in V$,

$$P_x^{(3)}(y) \geq (1 - o(1))\frac{n^2p^3}{(np)^3} = \frac{1 - o(1)}{n}$$

and we can proceed as we did for the previous case from (34).

Case III: $\omega \geq n^{3/5}$. We use a similar argument to the previous case. We now use the fact that **qs** there are at least $\sim np^2$ paths of length two from x to y .

This completes the proof of Lemma 6. □

Lemma 7.

$$\pi_v \geq (1 - o(1))\frac{\deg^-(v)}{m} \text{ for all } v \in V,$$

Proof It follows from Lemma 6 that **whp**, that for any $y \in V$ we have

$$\pi_y = \sum_{x \in V} \pi_x P_x^{(s)}(y) \geq (1 - o(1)) \frac{\deg^-(y)}{m} \sum_{x \in V} \pi_x = (1 - o(1)) \frac{\deg^-(y)}{m}. \quad (35)$$

□

At this point we have proved that the expression in Theorem 2 is a lower bound for the steady state.

3.3 Upper Bounds on the stationary distribution

Lemma 8. *There exists an integer $s = O(\log_{np} n)$ such that **whp** for all $x, y \in V$,*

$$P_x^{(s)}(y) \leq (1 + o(1)) \frac{\deg^-(y)}{m} \text{ for all } v \in V,$$

where m is the number of edges in $D_{n,p}$.

The proof of Lemma 8 will also be split into (two) cases and each requires a sequence of lemmas.

Case I: $np \leq n^\delta$ where $0 < \delta \ll \eta \ll 1$ are positive constants.

We will use the following values: Here $\Lambda = \log_{np} n$.

$$\begin{aligned} \ell_0 &= (1 + \eta)\Lambda, & \ell_1 &= (1 - 10\eta)\Lambda, & \ell_2 &= 11\eta\Lambda, \\ \ell_3 &= (1 - \eta/10)\Lambda, & \ell_4 &= \eta\Lambda/20, & \ell_5 &= 9\eta\Lambda/10. \end{aligned}$$

Given a fixed x, y we will grow T_X from x to a depth ℓ_1 and T_Y into y to a depth ℓ_2 . We re-define $\mathcal{D}_1^+, \mathcal{D}_1^-$ to fit these new specifications.

We begin with a lemma that will help simplify our calculations.

Lemma 9. *Suppose that $np \leq n^\delta$ where $\delta \ll 1$. Then **whp** $S \subseteq V, |S| \leq s_0 = \frac{1}{2} \log_{np} n$ implies that S contains at most $|S|$ edges.*

Proof The expected number of sets S with more than $|S|$ edges can be bounded by

$$\sum_{s=3}^{s_0} \binom{n}{s} \binom{s^2}{s+1} p^{s+1} \leq \sum_{s=3}^{s_0} (e^2 np)^s sep = o(1).$$

□

Fix x, y and let X_i , $0 \leq i \leq \ell_3$ be the set of vertices that are reachable from x by a walk of length i . These sets are slightly larger than the X_i of Lemma 6 in that we allow them to overlap. Let

$$X^* = X^*(x) = \bigcup_{i=0}^{\ell_3} X_i.$$

For $1 \leq i \leq \ell_1$ let

$$\tilde{X}_i = \left\{ a \in X_i \setminus \bigcup_{k=0}^{i-1} X_k : \exists b \in X_j \text{ and } j \leq i \text{ such that } (a, b) \text{ is an edge} \right\}$$

and let $\tilde{X} = \bigcup_{i=\ell_4}^{\ell_1} \tilde{X}_i$. Let

$$\begin{aligned} \mathcal{L}_a &= \{ \exists z \in X^* : z \text{ has } > 100/\eta \text{ in-neighbours in } X^* \} \\ \mathcal{L}_b &= \left\{ \exists 0 \leq i \leq \ell_1 : |\tilde{X}_i| > n^{2\delta-10\eta} (Cnp)^i + O(\log n) \right\} \end{aligned}$$

Lemma 10.

(a)

$$\Pr(\neg \mathcal{L}_a \mid \mathcal{D}_1^+) = O(n^{-9}).$$

(b)

$$\Pr(\neg \mathcal{L}_b \mid \mathcal{D}_1^+) = O(n^{-9}).$$

Proof

(a) Let ζ be the number of in-neighbours of z in $\bigcup_{i=1}^{\ell_3} X_i$. $\zeta \mid_{\mathcal{D}_1^+}$ is stochastically dominated by $1 + \text{Bin}((Cnp)^{\ell_3}, p) \mid_{\mathcal{D}_1^+}$. Hence if $r + 1 = 100/\eta$,

$$\Pr(\zeta \geq r + 1 \mid \mathcal{D}_1^+) \leq (1 + o(1))(Cnp)^{r\ell_3} p^r \leq 2n^{r(\delta-\eta/10)} \leq n^{-9}.$$

Part (a) of the lemma follows.

(b) Given \mathcal{D}_1^+ and $\ell_1 \leq \ell_3$ we see that $|\tilde{X}_i|$ is dominated by $\text{Bin}(n^{1-10\eta+o(1)}(Cnp)^i, p) \mid_{\mathcal{D}_1^+}$ and the result follows from Chernoff bounds. \square

We say that a vertex $z \in X_i$, $i \leq \ell_4$ is *special* if it has two in-neighbours in $\bigcup_{j<i} X_j$ or if it has x as an out-neighbour. It follows from Lemma 9 that **whp** there can be at most one special vertex for a given x . Denote it by ξ_x if it exists and put $\xi_x = n + 1$ if x does not have a special vertex.

For $u \in X_{\ell_1}$ we let

$$\alpha_{u, \ell_1} = \Pr(\mathcal{W}_x(\ell_1) = u \text{ and } \mathcal{W}_x \text{ does not use } \xi_x),$$

where of course

$$\sum_{u \in X} \alpha_{u, \ell_1} \leq 1. \quad (36)$$

Now define $Y_0 = \{y\}, Y_1, \dots, Y_{\ell_2}$ where $Y_{i+1} = N^-(Y_i) \setminus (Y_0 \cup \dots \cup Y_i)$ for $0 \leq i < \ell_2$. (This time we allow $Y_i \cap X \neq \emptyset$).

Now define the BFS tree T_Y as we did in the proof of Lemma 6 (i.e. by only keeping the first edge placing a vertex into a Y_i). Then define the $\beta_{i,v}$ as we did in (16). We will estimate

$$P_x^{\ell_0+1}(y) \leq Z_x^{\ell_0+1}(y) + Q_x^{\ell_0+1}(y) + R_x^{\ell_0+1}(y) + S_x^{\ell_0+1}(y) \quad (37)$$

where

- $$Z = Z_x^{\ell_0+1}(y) = \sum_{\substack{u \in X_{\ell_1} \\ v \in Y_{\ell_2} \setminus X}} \alpha_{\ell_1, u} \beta_{\ell_2, v} \frac{1_{uv}}{\deg^+(u)} \quad (38)$$

is the probability that $\mathcal{W}_x(\ell_0 + 1) = y$ and the walk does not go through ξ_x and the $(\ell_1 + 1)$ th edge (u, v) is such that $v \in Y_{\ell_2} \setminus X$ and the last ℓ_2 vertices of the walk use edges of the tree T_Y .

- $Q_x^{\ell_0+1}(y)$ is the probability that $\mathcal{W}_x(\ell_0 + 1) = y$ and the $(\ell_1 + 1)$ th edge (u, v) is such that $v \in Y_{\ell_2} \cap X$ and the last ℓ_2 vertices of the walk use edges of the tree T_Y .
- $R_x^{\ell_0+1}(y)$ is the probability that $\mathcal{W}_x(\ell_0 + 1) = y$ and the last ℓ_2 vertices of the walk use an edge which is not part of the tree T_Y .
- $S_x^{\ell_0+1}(y)$ is the probability that $\mathcal{W}_x(\ell_0 + 1) = y$ and the walk goes through ξ_x .

We first deal with Z , as defined in (38). For this we bound $\sum_{v \in Y_{\ell_2}} \beta_{\ell_2, v}$ stochastically from above by $W_{\ell_2}^Z$ (see Lemma 5). We will condition on $\mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{N}(y), \mathcal{L}_a$. The first summation in the definition of W_t^Z accounts for the edges into $v \in Y_{\ell_2-t} \setminus X$. The second summation comes from vertices in $Y_{\ell_2-t} \cap X$. For such a vertex $v \in X$, the construction of X has exposed some edges into v . But by \mathcal{L}_a implies that this is bounded by $100/\eta$. Thus we take $\theta = 100/\eta$ in the definition of W_t^Z . The denominator cnp for the terms of the second summation is a lower bound on the out-degree of in-neighbours of v in X . We have to use a lower bound, since the out-neighbours of these in-neighbours have been exposed.

Observe now that with $\epsilon_Z = 1/\log \log n$ and $\zeta = \epsilon_Z/3$ and $M = 10/(\epsilon_Z c)$,

$$\begin{aligned} & \Pr \left(\sum_{v \in Y_{\ell_2}} \beta_{\ell_2, v} \geq (1 + \epsilon_Z) \frac{s}{np} \middle| \mathcal{D}_1^+, \mathcal{D}_1^- \right), \quad s = \deg^-(y) \\ & \leq \Pr \left(\sum_{v \in Y_{\ell_2}} \beta_{\ell_2, v} \geq (1 + \epsilon_Z) \frac{s}{np} \middle| \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{N}(y), \mathcal{L}_a \right) + \\ & \quad \Pr(\neg \mathcal{N}(y) \mid \mathcal{D}_1^+, \mathcal{D}_1^-) + \Pr(\neg \mathcal{L}_a \mid \mathcal{D}_1^+) \frac{\Pr(\mathcal{D}_1^+)}{\Pr(\mathcal{D}_1^+, \mathcal{D}_1^-)} \end{aligned} \quad (39)$$

$$\leq 2\Pr \left(\sum_{i=1}^{s-\zeta_0-\theta} \frac{A_i}{(1-\zeta)np} + \sum_{i=s-\zeta_0-\theta+1}^s \frac{A_i}{cnp} \geq (1 + \epsilon_Z) \frac{s}{np} \right) + o(n^{-2}) \quad (40)$$

$$\leq 2\mathbf{E} \left(\exp \left\{ M \left(\left(\frac{s-\zeta_0-\theta}{1-\zeta} + \frac{\zeta_0+\theta}{c} \right) \left(1 + \frac{M(5+\theta)\ell_2}{cnp} \right) - (1 + \epsilon_Z)s \right) \right\} \right) + o(n^{-2}) \quad (41)$$

$$\begin{aligned} & \leq 2e^{-M\epsilon_Z cnp/2} + o(n^{-2}) \\ & = o(n^{-2}). \end{aligned} \quad (42)$$

Note that in going from (40) to (41) we are applying Lemma 5 to each individual A_i and then using their independence to deal with the sum.

Using the Hoeffding inequality, we see as in (32), (33) that

$$\Pr \left(Z \geq (1 + o(1)) \frac{d^-(y)}{m} \middle| \mathcal{D}_1^+, \mathcal{D}_1^- \right) = o(n^{-2}). \quad (43)$$

We deal next with the expression $Q_x^{\ell_0+1}(y)$.

We claim that

$$\Pr \left(Q_x^{\ell_0+1}(y) \geq \left(\frac{100}{\eta} \right)^{\ell_1} \frac{1}{(cnp)^{\ell_0}} \sum_{l=0}^{\ell_1} \sum_{\kappa=1}^{\ell_1-l+1} |N_\kappa^+(\tilde{X}_l) \cap Y| \middle| \mathcal{D}_1^+, \mathcal{D}_1^- \right) = o(n^{-2}). \quad (44)$$

where $N_\kappa^+(S)$ is the set of vertices that are reachable from S by a walk of length κ .

To see this, suppose that $W = (x = u_{\ell_1}, \dots, u_1, u_0, v, \dots, y)$ is a walk that contributes to $Q_x^{\ell_0+1}(y)$. Here we have $u_j \in X_{\ell_1-j}$ for $0 \leq j \leq \ell_1$ and $v \in X_i \cap Y$ for some $0 \leq i \leq \ell_1$. There may be a choice for the index i such that $v \in X_i$, in which case we choose i as small as possible, i.e. $i = \min \{r : v \in X_r\}$. We show next that $v \in N_\kappa^+(\tilde{X}_l)$ where $0 \leq l \leq \ell_1$ and $l + \kappa \leq \ell_1 + 1$.

Define $k \geq 0$ by $u_j \in X_{i-1-j}$ for $0 \leq j < k$ and either (i) $u_k \notin X_{i-1-k}$ or (ii) $k = i - 1$ and $u_k = x$.

Assume first that k is defined through case (i). It follows that $u_k \notin X_s$ for any $s < i - k - 1$. Otherwise we have $v \in X_{s+k+1}$ and we could have chosen a smaller i . Now define $l = \min\{j : u_k \in X_j\}$. We must have $i - k \leq l \leq \ell_1 - k$ and we also have $u_{k-1} \in X_{i-k}$, where $u_{k-1} = v$ if $k = 0$. The existence of the edge (u_k, u_{k-1}) implies that $u_k \in \tilde{X}_l$ and hence $v \in N_{k+1}^+(\tilde{X}_l)$. We now take $\kappa = k + 1$.

Now suppose that k is defined through case (ii). Since $x \neq u_{k+1} \in X_s$ implies $v \in X_{s+k+2}$ we know that $l = \min\{j : u_{k+1} \in X_j\} > 0$ satisfies $i - 2 - k \leq l \leq \ell_1 - k - 2$. The existence of the edge $(u_{k+1}, x = u_k)$ implies that $u_{k+1} \in \tilde{X}_l$ and hence $v \in N_{k+2}^+(\tilde{X}_l)$. We now take $\kappa = k + 2$.

At this point we have bounded the number of choices for v . This also bounds the number of walks contributing to $Q_x^{\ell_0+1}(y)$ since for such walks, there is a unique path from v to y in T_Y . We use the term $(100/\eta)^{\ell_1}$ to bound the number of walks that can use v . It comes from additionally conditioning on \mathcal{L}_a . This adds $o(n^{-2})$ to our probability estimate, see (39). This term is then an upper bound on the number of walks W going through a fixed v . (There is a unique path from v to y using the edges of T_Y). The term $1/(cnp)^{\ell_0}$ is an upper bound on the probability of this walk being taken. This completes the verification of (44).

We show that

$$\Pr\left(\sum_{i=0}^{\ell_1} \sum_{j=1}^{\ell_1-i+1} |N_j^+(\tilde{X}_i) \cap Y| > n^{\eta/2} \mid \mathcal{D}_1^+, \mathcal{D}_1^-\right) = o(n^{-2}). \quad (45)$$

In which case we have

$$\Pr(Q_x^{\ell_0+1}(y) \geq n^{-1-\eta/2+o(1)} \mid \mathcal{D}_1^+, \mathcal{D}_1^-) = o(n^{-2}). \quad (46)$$

For (45) we let $\Sigma_1 = \sum |N_j^+(\tilde{X}_i) \cap Y|$ and write for any $t \geq 0$,

$$\Pr(\Sigma_1 \geq t \mid \mathcal{D}_1^+, \mathcal{D}_1^-) \leq \Pr(\Sigma_1 \geq t \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a) + \Pr(\neg \mathcal{L}_a \mid \mathcal{D}_1^+) \frac{\Pr(\mathcal{D}_1^+)}{\Pr(\mathcal{D}_1^+, \mathcal{D}_1^-)}. \quad (47)$$

Now let Σ_2 be a bound on the number of vertices in Y that have an in-neighbour in \tilde{X} . Given, \mathcal{L}_a , we can then bound Σ_1 by $(100/\eta)^{\ell_2} \Sigma_2$. Thus,

$$\Pr(\Sigma_1 \geq t \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a) \leq \Pr(\Sigma_2 \geq (100/\eta)^{-\ell_2} t \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a). \quad (48)$$

For a bound we use $\Sigma_2 = \Sigma_3 + \Sigma_4$ where Σ_3 is the number of vertices in Y that have an in-neighbour in $\tilde{X} \cap Y$ and Σ_4 is the number of edges from $\tilde{X} \setminus Y$ to $Y \setminus \tilde{X}$.

Let $\hat{\mathcal{D}}_1^-$ be the event $\{\deg^-(v) \leq Cnp\}$ for $v \in Y$ i.e. drop the lower bound on the $\deg^-(v)$ in order to make a monotone decreasing event. Then given, $\mathcal{L}_a, \mathcal{L}_b, \hat{\mathcal{D}}_1^-$ we see that Σ_3 is stochastically dominated by $(100/\eta)^{\ell_2}$ times

$$\text{Bin}\left(2(Cnp)^{\ell_2}, p \sum_{i=0}^{\ell_1} ((Cnp)^i n^{2\delta-10\eta} + O(\log n))\right). \quad (49)$$

We use $\hat{\mathcal{D}}_1^-$ in place of \mathcal{D}_1^- so that we can appeal to the FKG inequality here. It follows that

$$\Pr(\Sigma_3 \geq n^\delta \mid \hat{\mathcal{D}}_1^-, \mathcal{L}_a, \mathcal{L}_b) \leq n^{-10}, \text{ say.}$$

We next observe that given $\mathcal{L}_b, \mathcal{D}_1^-, \Sigma_4$ is dominated by (49). It follows that

$$\Pr(\Sigma_2 \geq 2n^\delta \mid \hat{\mathcal{D}}_1^-, \mathcal{L}_a, \mathcal{L}_b) \leq n^{-10}.$$

From (25) we have

$$\Pr(\Sigma_2 \geq 2n^\delta \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a, \mathcal{L}_b) \leq n^{-10} \frac{\Pr(\hat{\mathcal{D}}_1^-, \mathcal{L}_a, \mathcal{L}_b)}{\Pr(\mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a, \mathcal{L}_b)} \leq 2n^{-10}.$$

From (26) we deduce that

$$\Pr(\Sigma_2 \geq 2n^\delta \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a) \leq 2n^{-10} + \Pr(-\mathcal{L}_b \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a) = O(n^{-9}).$$

Going back to (47) we see that

$$\Pr(\Sigma_1 \geq n^{\delta+o(1)} \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a) = O(n^{-9}). \quad (50)$$

Equation (45) follows from (46) and (50).

We deal next with $R_x^{\ell_0+1}(y)$. Observe that Lemma 9 implies indirectly that **whp** the set Y contains at most $|Y|$ edges and that there is at most one edge $e = (u_y, v_y)$ that is contained in Y and is not part of T_Y . If Y contained more than $|Y| + 1$ edges then it would contain two distinct cycles C_1, C_2 . In which case, C_1, C_2 and the shortest (semi-)path joining them in Y would violate the condition of the lemma. Here a semi-path is a path if orientation is ignored.

We must have $u_y \in Y_i, v_y \in Y_j$ for some $0 \neq j \geq i - 1$. Otherwise $(u_y, v_y) \in T_Y$.

Suppose first that $j \geq \ell_5$. Then conditional on the occurrence of $\mathcal{L}_a, \mathcal{D}_1^-$ there are certainly less than $(Cnp)^{\frac{2}{5} \log_{np} n} \times (100/\eta)^{\ell_3}$ walks contributing to $R_x^{\ell_0+1}(y)$. The first term in this factor is the number of walks of length at most $\frac{2}{5} \log_{np} n$ that start in X^* and end at v_y and the second factor bounds the number of walks of length at most ℓ_3 that start at x and end at the starting point of such a walk. Thus under these circumstances,

$$R_x^{\ell_0+1}(y) \leq \frac{(Cnp)^{\frac{2}{5} \log_{np} n} \times (100/\eta)^{\ell_3}}{(cnp)^{(1+\eta) \log_{np} n}} \leq n^{-1-\eta/2}. \quad (51)$$

Remark 1. *We can summarise part of what we have proved so far as, that if x, y are such that (i) there are no special vertices and (ii) there are no extra edges within ℓ_5 of y and if $\mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a, \mathcal{L}_b$ occur, then with probability $1 - o(n^{-2})$, $P_x^{\ell_0+1}(y) = (1 + o(1)) \frac{\deg^-(y)}{m}$.*

Suppose then that $\xi_y = v_y$ if v_y exists and is within ℓ_5 of y . Let ℓ_x be the distance from x to ξ_x if ξ_x exists, and $\ell_x = 0$ otherwise. Let ℓ_y be the distance from ξ_y to y if ξ_y exists, and $\ell_y = 0$ otherwise. We have from Remark 1, that

$$\Pr \left(P_{\xi_x}^{\ell_6}(\xi_y) = (1 + o(1)) \frac{\deg^-(\xi_y)}{m} \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a, \mathcal{L}_b \right) = 1 - o(n^{-2}).$$

where $\ell_6 = \ell_0 + 1 - \ell_x - \ell_y$. This is because $\ell_6 \geq (1 + \frac{\eta}{20}) \log_{np} n$ and we can repeat the arguments above with ξ_x, ξ_y replacing x, y and with a smaller value of η . This assumes that the high probability event $\mathcal{L}9$ described in Lemma 9 occurs. In which case there is no edge (u_{ξ_y}, v_{ξ_y}) to deal with.

We write

$$R_x^{\ell_0+1}(y) + S_x^{\ell_0+1}(y) \leq A_1 + A_2 + A_3$$

where

- A_1 is the probability that \mathcal{W}_x goes to y through ξ_x and ξ_y .
- A_2 is the probability that \mathcal{W}_x goes to y through ξ_x and not through ξ_y .
- A_3 is the probability that \mathcal{W}_x goes to y through ξ_y and not through ξ_x .

$A_1 = 0$ if one of ξ_x, ξ_y does not exist etcetera. Also, conditional on $\mathcal{D}_1^+, \mathcal{D}_1^-$, we have

$$\Pr(\mathcal{W}_x(\ell_x) = \xi_x) = O(1/np) \text{ and } \Pr(\mathcal{W}_{\xi_y}(\ell_y) = y) = O(1/np). \quad (52)$$

So

$$\begin{aligned} A_1 &= \Pr(\mathcal{W}_x(\ell_x) = \xi_x) P_{\xi_x}^{\ell_6}(\xi_y) \Pr(\mathcal{W}_{\xi_y}(\ell_y) = y) \quad \text{assuming } \xi_x, \xi_y \text{ exist} \\ &\leq (1 + o(1)) \Pr(\mathcal{W}_x(\ell_x) = \xi_x) \Pr(\mathcal{W}_{\xi_y}(\ell_y) = y) \frac{Cnp}{m} \\ &= o\left(\frac{\deg^-(y)}{m}\right). \\ A_2 &\leq \Pr(\mathcal{W}_x(\ell_x) = \xi_x) \frac{Cnp}{m} \quad \text{assuming } \xi_x \text{ exists} \\ &= o\left(\frac{\deg^-(y)}{m}\right). \\ A_3 &\leq \frac{Cnp}{m} \Pr(\mathcal{W}_{\xi_y}(\ell_y) = \xi_x) \quad \text{assuming } \xi_y \text{ exists} \\ &= o\left(\frac{\deg^-(y)}{m}\right). \end{aligned}$$

Hence,

$$\Pr(R_x^{\ell_0+1}(y) + S_x^{\ell_0+1}(y) = o(1/n) \mid \mathcal{D}_1^+, \mathcal{D}_1^-, \mathcal{L}_a, \mathcal{L}_b) = 1 + o(n^{-2}). \quad (53)$$

Removing conditioning as done previously, it follows from equations (43), (46), (51) and (53) that

$$P_x^{\ell_0+1}(y) \leq (1 + o(1)) \frac{\deg^-(y)}{m} \text{ for all } x, y \in V.$$

This completes the proof for Case I of Lemma 8.

Case II: $np \geq n^\delta$.

We can deal with this case by using a concentration inequality (54) from Kim and Vu [11]: Let $\Upsilon = (W, E)$ be a hypergraph where $e \in E$ implies that $|e| \leq s$. Let

$$Z = \sum_{e \in E} w_e \prod_{i \in e} z_i$$

where the $w_e, e \in E$ are positive reals and the $z_i, i \in W$ are independent random variables taking values in $[0, 1]$. For $A \subseteq W, |A| \leq s$ let

$$Z_A = \sum_{\substack{e \in E \\ e \supseteq A}} w_e \prod_{i \in e \setminus A} z_i.$$

Let $M_A = \mathbf{E}(Z_A)$ and $M_j(Z) = \max_{A, |A| \geq j} M_A$ for $j = 0, 1$. There exist positive constants a and b such that for any $\lambda > 0$,

$$\Pr(|Z - \mathbf{E}(Z)| \geq a\lambda\sqrt{M_0 M_1}) \leq be^{-\lambda}. \quad (54)$$

For us, W will be the set of edges of \vec{K}_n the complete digraph on n vertices. E will be the set of sets of edges in walks of length $s = \lceil 2/\delta \rceil$ between two fixed vertices x and y in \vec{K}_n , and $w_e = 1$. Z will be the number of walks that are in $D_{n,p}$. In which case we have

$$\begin{aligned} (n)_{s-1} p^s &\leq \mathbf{E}(Z) \leq (1 + o(1))n^{s-1} p^s \\ M_0 &= \mathbf{E}(Z) \\ M_1 &\leq (1 + o(1))n^{s-2} p^{s-1} = (1 + o(1))M_0/np. \end{aligned}$$

Applying Theorem 54 with $\lambda = (s + 2) \log n$ we see that for any x, y we have

$$\Pr(Z \geq \mathbf{E}(Z) + O((\log n)^s \mathbf{E}(Z) n^{-\delta/2})) \leq O(n^{-3}).$$

Thus **whp**

$$P_x^s(y) \leq (1 + o(1)) \frac{n^{s-1} p^s}{((1 - \epsilon_1)np)^s} = \frac{1 + o(1)}{n} \quad \forall x, y \in V.$$

This completes the proof of Lemma 6. □

We can now give an upper bound complementing that in Lemma 6.

Lemma 11.

$$\pi_v \leq (1 + o(1)) \frac{\deg^-(v)}{m} \text{ for all } v \in V,$$

Proof It follows from Lemma 8 that **whp**, that for any $y \in V$ we have

$$\pi_y = \sum_{x \in V} \pi_x P_x^{(s)}(y) \leq (1 + o(1)) \frac{\deg^-(y)}{m} \sum_{x \in V} \pi_x = (1 + o(1)) \frac{\deg^-(y)}{m}. \quad (55)$$

□

This completes the proof of Theorem 2. □

3.4 Removing Assumption 1

We will assume now that

$$1 + o(1) \leq d \leq 2.$$

We now need to prove some additional lemmas. Let a vertex be *small* if it has in-degree or out-degree at most $\log n/20$ and *large* otherwise.

Lemma 12.

- (a) **Whp** there are fewer than $n^{1/5}$ small vertices.
- (b) **Whp** every pair of small vertices are at least distance $\ell_{10} = \frac{\log n}{10 \log \log n}$ apart.
- (c) **Whp** there does not exist a vertex v with $\max \{ \deg^+(v), \deg^-(v) \} \leq \log n/20$.
- (d) **Whp**

$$\sum_{z \in N^-(y)} \frac{\deg^-(z)}{\deg^+(z)} = (1 + o(1)) \deg^-(y) \text{ for all small } y.$$

Proof

- (a) The expected number of small vertices is at most

$$n \sum_{k=0}^{\log n/20} \binom{n-1}{k} p^k q^{n-1-k} = O(n^{.1998}).$$

Part (a) now follows from the Markov inequality.

- (b) The expected number of pairs of vertices at distance ℓ_{10} or less is at most

$$n^2 \sum_{k=0}^{\ell_{10}} n^k p^{k+1} \left(2 \sum_{l=0}^{\log n/20} \binom{n-1}{l} p^l (1-p)^{n-1-l} \right)^2 =$$

$$O(n(d \log n)^{\ell_{10}+1} (20ed)^{\log n/10} n^{-2d}) = O(n \cdot n^{1/10+o(1)} \cdot n^{1/2} \cdot n^{-2}) = o(1).$$

- (c) The expected number of vertices with small out- and in-degree is $O(n^{1-2 \times .8002}) = o(1)$.

- (d) We let $\epsilon = \frac{1}{\log \log n}$ and for $k \in [\log n/20]$ we let

$$\lambda_k = \begin{cases} 1 & 1 \leq k \leq \frac{\log n}{(\log \log n)^4} \\ (\log \log n)^4 & \frac{\log n}{(\log \log n)^4} \leq k \leq \frac{\log n}{20} \end{cases}.$$

Then the probability that there exists a vertex of in-degree $k \in [\log n/20]$ with λ_k in-neighbours, one of whose degrees is outside $(1 \pm \epsilon)np$, is bounded by

$$\sum_{k=1}^{\log n/20} n \binom{n-1}{k} p^k q^{n-1-k} \binom{k}{\lambda_k} (2e^{-\epsilon^2 np/3})^{\lambda_k} \leq \sum_{k=1}^{\log n/20} 2n^{1-d} \left(\frac{nep}{k} \cdot \frac{e}{\lambda_k} \cdot n^{-\epsilon^2 d \lambda_k / (3k)} \right)^k = o(1).$$

Now assume that there are fewer than λ_k neighbours of v , one of whose degrees is outside $(1 \pm \epsilon)np$. Then, assuming no neighbours of v are small,

$$\sum_{z \in N^-(y)} \frac{\deg^-(z)}{\deg^+(z)} = \begin{cases} (1 + O(\epsilon))k & 1 \leq k \leq \frac{\log n}{(\log \log n)^4} \\ (1 + O(\epsilon))(k - \lambda_k) + O(\lambda_k) & \frac{\log n}{(\log \log n)^4} \leq k \leq \frac{\log n}{20} \end{cases}$$

This completes the proof of the lemma. \square

Let a cycle of the underlying graph of $D_{n,p}$ be called a *weak cycle*. Let *weak distance* refer to distance in this graph.

Lemma 13. *Whp there does not exist a small vertex that is within weak distance ℓ_{10} of a weak cycle C of length at most ℓ_{10} .*

Proof The probability that such a pair v, C exist is at most

$$\begin{aligned} \sum_{i=3}^{\ell_{10}} (2np)^i \sum_{j=0}^{\ell_{10}} (np)^j \left(2 \sum_{l=0}^{\log n/20} \binom{n-1}{l} p^l (1-p)^{n-1-l} \right) \\ = O(n^{1/10+o(1)} \cdot n^{1/10+o(1)} \cdot n^{-4/5+o(1)}) = o(1). \end{aligned}$$

Here $i = |C|$ and j is the weak distance of v from C . \square

3.4.1 Lower bound on steady state

For this we proceed as in Section 3.2 but initially restrict our analysis to large x, y . Also, we do not include the small vertices when creating the X_i, Y_i . Our previous analysis holds up with $c = 1/20$ in the use of Lemma 5. Furthermore, avoiding the $\leq n^{1/5}$ small vertices is easily incorporated because we have allowed in the proof for the avoidance of $n^{2/3+o(1)}$ vertices from $\bigcup_i X_i$ etc..

In this way, we show that **whp**

$$P_x^{(2\ell+1)}(y) \geq (1 - o(1)) \frac{\deg^-(y)}{m}, \quad \text{for all large } x, y.$$

If x is small, then **whp** it will only have large out-neighbours and so

$$P_x^{(2\ell+2)}(y) = \frac{1}{\deg^+(x)} \sum_{z \in N^+(x)} P_z^{(2\ell+1)}(y) \geq (1 - o(1)) \frac{\deg^-(y)}{m}. \quad (56)$$

A similar argument deals with small y i.e.

$$P_x^{(2\ell+2)}(y) = \sum_{z \in N^-(y)} \frac{P_x^{(2\ell+1)}(z)}{\deg^+(z)} \geq (1 - o(1)) \sum_{z \in N^-(y)} \frac{\deg^-(z)}{m \deg^+(z)} \geq (1 - o(1)) \frac{\deg^-(y)}{m}. \quad (57)$$

and we end with a proof that **whp**

$$P_x^{(2\ell+3)}(y) \geq (1 - o(1)) \frac{\deg^-(y)}{m}, \quad \text{for all } x, y.$$

3.4.2 Upper bound on steady state

For this we proceed as in Section 3.3 but initially try to restrict our analysis to large x, y . It is comforting to know that ξ_x, ξ_y are large if they exist at all. This follows from Lemma 13. This is important because (52) needs to have ξ_x, ξ_y large.

Also, if ξ_y exists then the nearest small vertex to y is at distance more than ℓ_{10} . (58)

We will have to add an extra term $T_x^{\ell_0+1}(y)$ to deal with walks that go through the (at most one) small vertex $z_y \in Y_j$. We must deal with these separately, because of complications with the second summation in the definition of W_t^Z .

Case (a): We deal first with large x and large y for which z_y does not exist. For this we can follow Section 3.3. We will not have trouble when we have to estimate A_1, A_2, A_3 because of (58).

Case (b): We can now extend to the case where $x \in V$ and y is small. Here we use (57) with the inequalities reversed.

Case (c): Now suppose that y is large and z_y exists. We deal with this in a similar way to that in which we dealt with $R_x^{\ell_0+1}(y)$. The case $j \geq \ell_5$ yields the same term as in (51). Note here that ξ_{z_y} does not exist, see Lemma 13.

To deal with $j < \ell_5$ we argue as for (53), relying on Case (b).

Case (d): We finally deal with x small and $y \in V$. Here we use (56) with the inequalities reversed.

4 Upper Bound on Mixing time

We next show that the mixing time T as defined in (1) satisfies

$$T = o(\ell \log n) = o((\log n)^2) \quad (59)$$

where ℓ is given by (13).

Define

$$\bar{d}(t) = \max_{x, x' \in V} |P_x^{(t)} - P_{x'}^{(t)}| \quad (60)$$

to be the maximum over x, x' of the variation distance between $P_x^{(t)}$ and $P_{x'}^{(t)}$. Equation (34) implies that

$$\bar{d}(2\ell + 1) = o(1). \quad (61)$$

Lemma 20 of Chapter 2 of Aldous and Fill [1] proves that

$$\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t) \text{ and } \max_x |P_x^{(t)} - \pi_x| \leq \bar{d}(t)$$

and so (59) follows immediately from (61).

5 The Cover Time

We see immediately from (59) that Condition (b) of Lemma 3 is satisfied.

We will show shortly that if $\mathcal{D}^+ \cap \mathcal{D}^-$ holds then

$$R_T(1) = 1 + o(1) \quad (62)$$

and therefore $\sum_{t=1}^T r_t = o(1)$. Thus if $|z| \leq 1 + \lambda$, then as $\lambda = 1/KT$ we have

$$R_T(z) \geq 1 - \sum_{t=1}^T r_t |z|^t \geq 1 - (1 + \lambda)^T \sum_{t=1}^T r_t = 1 - o(1)$$

and Condition (a) of the Lemma is also satisfied. So for $v \in V$,

$$p_v = (1 + o(1)) \frac{\deg^-(v)}{m}.$$

Proof of (62): We first observe that because the minimum out-degree of $D_{n,p}$ is $\Omega(d \log n)$ we have for any x, y

$$\Pr(\mathcal{W}_v(t) = y \mid \mathcal{W}_v(t-1) = x) = O\left(\frac{1}{d \log n}\right). \quad (63)$$

The expected number of returns to $v \in V$ by \mathcal{W}_v is therefore $O(T/d \log n)$. Thus if $d \geq (\log n)^2$ we are done immediately.

If $d \leq (\log n)^2$ then a simple first moment calculation shows that **whp** for every vertex $v \in V$, there is at most one edge from a vertex in $N^+(v)$ to $\{v\} \cup N^+(v)$ or from a vertex in $N^+(N^+(v))$ to $\{v\} \cup N^+(v) \cup N^+(N^+(v))$. Thus with probability $1 - O(1/(\log n)^2)$, $x = \mathcal{W}_v(3)$ satisfies $\text{dist}(x, v) \geq 3$ and then the probability of a return to v is $O(T/(\log n)^3)$.

5.1 Upper Bound on the Cover Time

For $np = d \log n$, d constant, let $t_0 = (1 + \epsilon)d \log \left(\frac{d}{d-1}\right) n \log n$. For $np = d \log n$ $d = d(n) \rightarrow \infty$ let $t_0 = (1 + \epsilon)n \log n$. Here $\epsilon \rightarrow 0$ sufficiently slowly so that all claimed inequalities below are valid.

Let $T_D(u)$ be the time taken by the random walk \mathcal{W}_u to visit every vertex of D . Let U_t be the number of vertices of D which have not been visited by \mathcal{W}_u at step t . We note the following:

$$C_u = \mathbf{E}(T_D(u)) = \sum_{t>0} \Pr(T_D(u) \geq t), \quad (64)$$

$$\Pr(T_D(u) \geq t) = \Pr(T_D(u) > t - 1) = \Pr(U_{t-1} > 0) \leq \min\{1, \mathbf{E}(U_{t-1})\}. \quad (65)$$

Recall that $\mathbf{A}_v(t)$, $t \geq T$ denotes the event that $\mathcal{W}_u(t)$ did not visit v in the interval $[T, t]$. It follows from (64), (65) that for all $t \geq T$,

$$C_u \leq t + 1 + \sum_{s \geq t} \mathbf{E}(U_s) \leq t + 1 + \sum_v \sum_{s \geq t} \Pr(\mathbf{A}_s(v)). \quad (66)$$

Assume first that $d(n) \rightarrow \infty$. If $v \in V$ and $t \gg T$ then

$$\Pr(\mathbf{A}_s(v)) \leq (1 + o(1)) \exp \left\{ -\frac{(1 - o(1))s}{n} \right\} + O(e^{-\Omega(s/T)}). \quad (67)$$

Plugging (67) into (66) we get

$$\begin{aligned} C_u &\leq t_0 + 1 + 2n \sum_{s \geq t_0} \left(\exp \left\{ -\frac{(1 - o(1))s}{n} \right\} + O(e^{-\Omega(s/T)}) \right) \\ &\leq t_0 + 1 + 3n^2 \exp \left\{ -\frac{(1 - o(1))t_0}{n} \right\} + O(nT e^{-\Omega(t_0/T)}) \\ &= (1 + o(1))t_0. \end{aligned} \quad (68)$$

Now assume that d is constant. For $v \in V$ we have

$$p_v \geq (1 - o(1)) \frac{\text{deg}^-(v)}{m}.$$

Then in place of (68) we find, using the bounds in Lemma 4 (b2),

$$\begin{aligned}
C_u &\leq t_0 + 1 + \frac{2 + o(1)}{n^{d-1}} \sum_{k=cnp}^{Cnp} \left(\frac{nep}{k}\right)^k \sum_{s \geq t_0} \left(\exp \left\{ -\frac{(1 - o(1))ks}{m} \right\} + O(e^{-\Omega(s/T)}) \right) \\
&\leq t_0 + 2 + O(nTe^{-\Omega(t_0/T)}) + \frac{2 + o(1)}{n^{d-1}} \sum_{k=cnp}^{Cnp} \left(\frac{nep}{k}\right)^k \frac{m}{k} e^{-(1-o(1))kt_0/m} \\
&\leq t_0 + 3 + \frac{2 + o(1)}{n^{d-1}} m \sum_{k=cnp}^{Cnp} \left(\frac{nep}{k}\right)^k \left(\frac{d-1}{d}\right)^{(1+\epsilon/2)k} \\
&\leq t_0 + 3 + (2 + o(1))m \sum_{k=cnp}^{Cnp} e^{-\epsilon k/2d} \\
&= (1 + o(1))t_0,
\end{aligned}$$

where we have used the fact that $(nep(d-1))/(kd)^k$ is maximized at $k = np(d-1)/d$.

5.2 Lower Bound on the Cover Time

For $np = d \log n$, d constant, let $t_1 = (1 - \epsilon)d \log \left(\frac{d}{d-1}\right) n \log n$. For $np = d \log n$ $d = d(n) \rightarrow \infty$ let $t_0 = (1 - \epsilon)n \log n$. Here $\epsilon \rightarrow 0$ sufficiently slowly so that all claimed inequalities below are valid.

Case 1: $np \leq n^\delta$ where $0 < \delta \ll \eta$ is a positive constant.

Let $V^* = \{v : \deg^-(v) = k^*\}$ where $k^* = (d-1) \log n$, (see Lemma 4(b3)). The maximum degree in D is $O(np)$ and so V^* contains a sub-set V^{**} of size $\geq n^{\gamma d/2}$ such that $v, w \in V^{**}$ implies $\text{dist}(v, w) > \ell_2$.

Now choose $u \notin V^{**}$ and let V^\dagger denote the set of vertices in V^{**} that have not been visited by \mathcal{W}_u by time t_1 .

Then

$$\begin{aligned}
\mathbf{E}(|V^\dagger|) &\geq n^{\gamma d/2} \left(\exp \left\{ -\frac{(1 + o(1))k^*t_1}{m} \right\} - o(e^{-t_1/T}) \right) \\
&\geq n^{\gamma d/3}.
\end{aligned}$$

As in previous papers we can use the Chebyshev inequality in order to show that $V^\dagger \neq \emptyset$ **whp**, completing the proof of Theorem 1.

We have to show that if $v, w \in V^{**}$ then

$$\Pr(\mathbf{A}_v(t_1) \cap \mathbf{A}_w(t_1)) \leq (1 + o(1))\Pr(\mathbf{A}_v(t_1))\Pr(\mathbf{A}_w(t_1)). \quad (69)$$

To prove this, we identify vertices v, w into a “supernode” σ to obtain a digraph D_σ with $n - 1$ vertices. In this digraph σ has in-degree $\deg^-(v) + \deg^-(w) = 2k^*$. The arguments we used in Section 3 remain valid in D_σ . In particular,

$$\pi_\sigma \sim (1 - o(1)) \frac{2k^*}{m}.$$

Using the suffice \mathbf{Pr}_σ to denote probabilities related to random walks in D_σ , it follows that

$$\begin{aligned} \mathbf{Pr}_\sigma(\mathbf{A}_\sigma(t_1)) &\leq \exp\left\{-\frac{(1 + o(1))2k^*t_1}{m}\right\} - o(e^{-t_1/T}) \\ &= (1 + o(1))\mathbf{Pr}(\mathbf{A}_v(t_1))\mathbf{Pr}(\mathbf{A}_w(t_1)). \end{aligned}$$

But, using rapid mixing in D_σ ,

$$\begin{aligned} \mathbf{Pr}_\sigma(\mathbf{A}_\sigma(t_1)) &= \sum_{x \neq \sigma} P_{\sigma,u}^T(x) \mathbf{Pr}_\sigma(\mathcal{W}_x(t) \neq \sigma, T \leq t \leq t_1) \\ &= \sum_{x \neq \sigma} \left((1 + o(1)) \frac{\deg^-(x)}{m} + O(n^{-3}) \right) \mathbf{Pr}_\sigma(\mathcal{W}_x(t) \neq \sigma, T \leq t \leq t_1) \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathbf{Pr}(\mathbf{A}_v(t_1) \cap \mathbf{A}_w(t_1)) &= \sum_{x \neq v,w} P_u^T(x) \mathbf{Pr}(\mathcal{W}_x(t) \neq v, w, T \leq t \leq t_1) \\ &= \sum_{x \neq v,w} \left((1 + o(1)) \frac{\deg^-(x)}{m} + O(n^{-3}) \right) \mathbf{Pr}(\mathcal{W}_x(t) \neq v, w, T \leq t \leq t_1) \end{aligned}$$

But,

$$\mathbf{Pr}_\sigma(\mathcal{W}_x(t) \neq \sigma, T \leq t \leq t_1) = \mathbf{Pr}(\mathcal{W}_x(t) \neq v, w, T \leq t \leq t_1)$$

because random walks from x that do not meet v, w or σ have the same measure in both digraphs.

It follows that

$$\mathbf{Pr}(\mathbf{A}_v(t_1) \cap \mathbf{A}_w(t_1)) - \mathbf{Pr}_\sigma(\mathbf{A}_\sigma(t_1)) \leq O(n^{-2}) + o(1)(\mathbf{Pr}(\mathbf{A}_v(t_1) \cap \mathbf{A}_w(t_1)) + \mathbf{Pr}_\sigma(\mathbf{A}_\sigma(t_1))).$$

This implies that

$$\mathbf{Pr}(\mathbf{A}_v(t_1) \cap \mathbf{A}_w(t_1)) \leq (1 + o(1))\mathbf{Pr}_\sigma(\mathbf{A}_\sigma(t_1)) + O(n^{-2}) \leq (1 + o(1))\mathbf{Pr}(\mathbf{A}_v(t_1))\mathbf{Pr}(\mathbf{A}_w(t_1)) + O(n^{-2}).$$

Thus

$$\mathbf{Var}(|V^\dagger|) = o(\mathbf{E}(|V^\dagger|)^2) + O(|V^*|^2 n^{-2}) = o(\mathbf{E}(|V^\dagger|)^2)$$

and so $V^\dagger \neq \emptyset$ **whp**.

Case 2: $np \geq n^\delta$.

In this range we take $t_1 = (1 - \epsilon)n \log n$ and let V^{**} be the set of vertices of degree $\lfloor np \rfloor$. A simple second moment calculation shows that **whp** we have $|V^{**}| = n^{1/2+o(1)}$. We then choose ϵ so that $\mathbf{E}(|V^\dagger|) \geq n^{1/4}$. It is then only a matter of verifying (69). We define D_σ as in the previous case and now use Theorem 54 to show that **whp** we have $\pi_\sigma \geq (2 - o(1))/n$. The details are as in the previous case.

This completes the proof of Theorem 1. □

References

- [1] D. Aldous and J. Fill, Reversible Markov Chains and Random Walks on Graphs, <http://stat-www.berkeley.edu/pub/users/aldous/RWG/book.html>.
- [2] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász and C. Rackoff, Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems. *Proceedings of the 20th Annual IEEE Symposium on Foundations of Computer Science* (1979) 218-223.
- [3] N. Alon and J. Spencer, The Probabilistic Method, Second Edition, Wiley-Interscience, 2000.
- [4] C. Cooper and A. M. Frieze, The cover time of sparse random graphs, *Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms*, (2003) 140-147.
- [5] C. Cooper and A. M. Frieze, The cover time of random regular graphs, *SIAM Journal on Discrete Mathematics*, 18 (2005) 728-740.
- [6] C. Cooper and A. M. Frieze, The cover time of the preferential attachment graph, to appear in *Journal of Combinatorial Theory Series B*.
- [7] C. Cooper and A. M. Frieze, The cover time of the giant component of a random graph, *Random Structures and Algorithms*, to appear.
- [8] U. Feige, A tight upper bound for the cover time of random walks on graphs, *Random Structures and Algorithms*, 6 (1995) 51-54.
- [9] U. Feige, A tight lower bound for the cover time of random walks on graphs, *Random Structures and Algorithms*, 6 (1995) 433-438.
- [10] W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* 58 (1963) 13-30.
- [11] J. H. Kim and V. Vu, *Concentration of multivariate polynomials and its applications*. *Combinatorica* 20 (2000) 417-434.