MIXING TIME OF PAGERANK SURFERS
ON SPARSE RANDOM DIGRAPHS

PIETRO CAPUTO[♯] AND MATTEO QUATTROPANI[♭]

ABSTRACT. We consider the generalised PageRank walk on a digraph G, with refresh probability α and resampling distribution λ. We analyse convergence to stationarity when G is a large sparse random digraph with given degree sequences, in the limit of vanishing α. We identify three scenarios: when α is much smaller than the inverse of the mixing time of G the relaxation to equilibrium is dominated by the simple random walk and displays a cutoff behaviour; when α is much larger than the inverse of the mixing time of G on the contrary one has pure exponential decay with rate α; when α is comparable to the inverse of the mixing time of G there is a mixed behaviour interpolating between cutoff and exponential decay. This trichotomy is shown to hold uniformly in the starting point and uniformly in the resampling distribution λ.

1. INTRODUCTION AND RESULTS

Given a directed graph $G = (V, E)$ and a parameter $\alpha \in (0, 1)$, the PageRank surf on G with damping factor $1 - \alpha$ is the Markov chain with state space V and transition probabilities given by

$$(1.1) \qquad P_\alpha(x, y) = (1 - \alpha)P(x, y) + \frac{\alpha}{n},$$

where $n = |V|$ is the number of vertices of G, and, writing d_x^+ for the out-degree of vertex x,

$$(1.2) \qquad P(x, y) = \begin{cases} 1/d_x^+ & \text{if } (x, y) \in E \\ 0 & \text{otherwise} \end{cases}$$

denotes the transition matrix of the simple random walk on G. The interpretation is that of a surfer that at each step, with probability $1 - \alpha$ moves to a vertex chosen uniformly at random among the out-neighbours of its current state, and with probability α moves to a uniformly random vertex in V. The surfer reaches eventually a stationary distribution π_α over V, called the PageRank of G. Since its introduction by Brin and Page in the seminal paper [10], PageRank has played a fundamental role in the ranking functions of all major search engines; see e.g. [15, 17]. A common generalization is the so-called customised or generalised PageRank, where the uniform resampling is replaced by an arbitrary probability distribution λ over V, so that (1.1) becomes

$$(1.3) \qquad P_{\alpha,\lambda}(x, y) = (1 - \alpha)P(x, y) + \alpha\lambda(y).$$

The resulting stationary distribution $\pi_{\alpha,\lambda}$, characterised by the equation

$$(1.4) \qquad \pi_{\alpha,\lambda}(y) = \sum_{x \in V} \pi_{\alpha,\lambda}(x) P_{\alpha,\lambda}(x, y),$$

depends in a nontrivial way on the parameter α and the distribution λ. There have been several investigations of the structural properties of $\pi_{\alpha,\lambda}$; see e.g. [18, 2, 9]; we refer in particular to the recent works [11, 16, 21] for cases where the graph G is drawn from the configuration model. Here we focus on the dynamical problem of determining the time needed for the surfer to reach the equilibrium distribution $\pi_{\alpha,\lambda}$, namely we study the mixing time of the Markov chain with transition matrix $P_{\alpha,\lambda}$. In the case $\alpha = 0$, this corresponds to the classical question of determining the mixing time of the simple random walk on the graph G; see e.g. [19]. Even for graphs where the latter is well understood, it is in general not immediate to deduce the influence of the parameter α and of the resampling distribution λ on the speed of convergence to equilibrium.

It is intuitively reasonable to guess that if the parameter α is suitably large compared to the inverse of the mixing time of the graph G, then the time to reach stationarity will be essentially the expected time needed to make the first λ-resampling transition, that is a geometric random variable with parameter α, while if α is suitably small compared to the inverse of the mixing time of the graph G, then one should reach stationarity well before the first λ-resampling, so that the speed of convergence to equilibrium will be essentially that of the simple random walk on G. Moreover, one could expect that when α is neither too small nor too large compared to the inverse of the mixing time of the graph G, then some interpolation between the two opposite behaviours should take place. In this paper we substantiate this intuitive picture for a large class of sparse directed graphs. The results hold uniformly in the initial position and uniformly in the resampling distribution λ.

1.1. **Two models of sparse digraphs.** We shall consider two families of directed graphs. Both are obtained via the so-called configuration model, with the difference that in the first case we fix both in and out degrees, while in the second case we only fix the out degrees. The models are sparse in that the degrees are bounded. We now proceed with the formal definition.

Let V be a set of n vertices. For simplicity we often write $V = [n]$, with $[n] = \{1, \ldots, n\}$. For each n, we are given two finite sequences $\mathbf{d}^+ = (d_x^+)_{x \in [n]}$ and $\mathbf{d}^- = (d_x^-)_{x \in [n]}$ of non negative integers such that

$$(1.5) \qquad\qquad m = \sum_{x \in V} d_x^+ = \sum_{x \in V} d_x^-.$$

The *directed configuration model* DCM($\mathbf{d}^\pm$), is the distribution of the random graph G obtained as follows: 1) equip each node x with d_x^+ tails and d_x^- heads; 2) pick uniformly at random one of the $m!$ bijective maps from the set of all tails into the set of all heads, call it ω; 3) for all $x, y \in V$, add a directed edge (x, y) every time a tail from x is mapped into a head from y through ω. The resulting graph G may have self-loops and multiple edges, however it is classical that by conditioning on the event that there are no multiple edges and no self-loops one obtains a uniformly random simple digraph with in degree sequence $\mathbf{d}^-$ and out degree sequence $\mathbf{d}^+$.

Structural properties of random graphs obtained in this way have been extensively studied in [13]. Here we shall consider the sparse case corresponding to bounded degree sequences. Moreover, in order to avoid non irreducibility issues, we shall assume that all degrees are at least 2. Thus, throughout this work it will always be assumed that

$$(1.6) \qquad\qquad \min_{x \in [n]} d_x^- \wedge d_x^+ \geq 2, \qquad \max_{x \in [n]} d_x^- \vee d_x^+ = O(1).$$

We often use the notation $\Delta = \max_{x \in [n]} d_x^- \vee d_x^+$. Under the first assumption it is known that DCM($\mathbf{d}^\pm$) is strongly connected with high probability; see e.g. [13]. Under the second assumption, it is known that DCM($\mathbf{d}^\pm$) has a uniformly (in n) positive probability of having no self-loops nor multiple edges; see e.g. [12]. In particular, any property that holds with high probability for DCM($\mathbf{d}^\pm$) will also hold with

high probability for a uniformly chosen simple digraph subject to the constraint that in and out degrees be given by $\mathbf{d}^-$ and $\mathbf{d}^+$ respectively. Here and throughout the rest of the paper we say that a property holds *with high probability* (w.h.p. for short) if the probability of the corresponding event converges to 1 as $n \to \infty$. In particular, it follows that w.h.p. there exists a unique stationary distribution π_0 for the simple random walk on G. Several properties of π_0 have been established recently in [6], where it was shown, among other facts, that π_0 can be described in terms of recursive distributional equations determined by the sequences $\mathbf{d}^\pm$.

To define the second model, for each n let $\mathbf{d}^+ = (d_x^+)_{x\in[n]}$ be a finite sequence of non negative integers and define the *out-configuration model* OCM($\mathbf{d}^+$) as the distribution of the random graph G obtained as follows: 1) equip each node x with d_x^+ tails; 2) pick, for every x independently, a uniformly random injective map from the set of tails at x to the set of all vertices V, call it ω_x; 3) for all $x, y \in V$, add a directed edge (x, y) if a tail from x is mapped into y through ω_x. Equivalently, G is the graph whose adjacency matrix is uniformly random in the set of all $n \times n$ matrices with entries 0 or 1 such that every row x sums to d_x^+. Notice that G may have self-loops, but there are no multiple edges in this construction. This is due to the requirement that the maps ω_x be injective. The latter choice is only a matter of convenience, and everything we say below is actually seen to hold as well for the model obtained by dropping that requirement. We write $\omega = (\omega_x)_{x\in[n]}$ for the collection of maps. As before we shall make the assumptions

$$(1.7) \qquad \min_{x\in[n]} d_x^+ \geq 2, \qquad \max_{x\in[n]} d_x^+ = O(1),$$

and use the notation $\Delta = \max_{x\in[n]} d_x^+$. We remark that under the above assumptions there can still be vertices with in-degree zero, and therefore in this case G is not necessarily strongly connected. However, it is still possible to show that w.h.p. there exists a unique stationary distribution π_0 for the simple random walk on G; see e.g. [1, 7] for more details.

In what follows $G = G(\omega)$ denotes a given realization of either the directed configuration model DCM($\mathbf{d}^\pm$) or the out-configuration model OCM($\mathbf{d}^+$) and all the results to be discussed will hold w.h.p. within these two ensembles. For the sake of simplicity we often refer to these as *model 1* and *model 2* respectively.

1.2. **Main results.** Let P denote the transition matrix of the simple random walk on G. When G is a digraph without multiple edges this is given by (1.2). If G has multiple edges, $P(x, y)$ is defined as $m(x, y)/d_x^+$ where $m(x, y)$ denotes the number of directed edges from x to y. For any $\alpha \in (0, 1)$ and any resampling distribution λ, let $P_{\alpha,\lambda}$ denote the PageRank transition matrix defined in (1.3). Notice that as soon as $\alpha > 0$, regardless of the realization of the graph G and of the chosen distribution λ, there exists a unique stationary distribution $\pi_{\alpha,\lambda}$ on V. Indeed, the transition matrix $P_{\alpha,\lambda}$ satisfies the so-called Doeblin condition if $\alpha > 0$; see Proposition 7 below for an explicit expression of $\pi_{\alpha,\lambda}$. Convergence to equilibrium will be quantified using the total variation distance. For two probability measures μ, ν, the latter is defined by

$$(1.8) \qquad \|\mu - \nu\|_{\mathrm{TV}} = \max_E |\mu(E) - \nu(E)|,$$

where the maximum ranges over all possible events in the underlying probability space. Starting at a node x the distribution of the PageRank surfer after t steps is $P_{\alpha,\lambda}^t(x, \cdot)$, and the distance to equilibrium is defined by

$$(1.9) \qquad \mathcal{D}_{\alpha,\lambda}^x(t) = \left\| P_{\alpha,\lambda}^t(x, \cdot) - \pi_{\alpha,\lambda} \right\|_{\mathrm{TV}}.$$

This defines a non-increasing function of $t \in \mathbb{N}$. It is convenient to extend it to a monotone function of $t \in [0, \infty)$, e.g. by considering the integer part of the argument. Finally, for any $\varepsilon \in (0, 1)$, the *ε-mixing time* is defined by

$$(1.10) \qquad T_{\alpha, \lambda}(\varepsilon) = \inf \left\{ t \geq 0 : \max_{x \in V} \mathcal{D}^x_{\alpha, \lambda}(t) \leq \varepsilon \right\}.$$

Both $\mathcal{D}^x_{\alpha, \lambda}(t)$ and $T_{\alpha, \lambda}(\varepsilon)$ are functions of the underlying graph G, and are therefore random variables. When $\alpha = 0$, we write $\mathcal{D}^x_0(t)$ and $T_0(\varepsilon)$ for the corresponding quantities. The behaviour of the distance $\mathcal{D}^x_0(t)$ and of the mixing time $T_0(\varepsilon)$ has been thoroughly investigated in [6] for model 1 and in [7] for model 2. Let us briefly recall the main conclusions of these works. In order to simplify the exposition, we shall adopt the following unified notation. Let us define the *in-degree distribution*

$$(1.11) \qquad \mu_{\mathrm{in}}(x) = \frac{1}{n} \times \begin{cases} d_x^- / \langle d \rangle & \text{model 1} \\ 1 & \text{model 2} \end{cases}$$

where we use the notation

$$\langle d \rangle = \frac{1}{n} \sum_{x \in V} d_x^- = \frac{m}{n}$$

for the average degree. Note that for model 2 the distribution μ_{in} represents the average in-degrees rather than the actual in-degrees. Next, let the *entropy H* and the associated *entropic time T_{ENT}* be defined by

$$(1.12) \qquad H = \sum_{x \in V} \mu_{\mathrm{in}}(x) \log d_x^+, \qquad T_{\mathrm{ENT}} = \frac{\log n}{H}.$$

Note that under our assumptions on $\mathbf{d}^{\pm}$ the deterministic quantities H, T_{ENT} satisfy $H = \Theta(1)$ and $T_{\mathrm{ENT}} = \Theta(\log n)$. The main results of [6, 7] state that, uniformly in the starting point $x \in V$, the rescaled function $\mathcal{D}^x_0(s\, T_{\mathrm{ENT}})$, $s > 0$, converges in probability as $n \to \infty$ to the step function

$$(1.13) \qquad \vartheta(s) = \begin{cases} 1 & \text{if } s < 1 \\ 0 & \text{if } s > 1. \end{cases}$$

More precisely, we may combine [6, Theorem 1] and [7, Theorem 1] to obtain the following statement.

Theorem 1 (Uniform cutoff at the entropic time [6, 7])**.** *Let G be a random graph from either the directed configuration model DCM($\mathbf{d}^{\pm}$) or the out-configuration model OCM($\mathbf{d}^{+}$). For each $s > 0, s \neq 1$ one has:*

$$(1.14) \qquad \max_{x \in [n]} |\mathcal{D}^x_0(s\, T_{\mathrm{ENT}}) - \vartheta(s)| \xrightarrow{\mathbb{P}} 0.$$

In (1.14) we use the notation $\xrightarrow{\mathbb{P}}$ for convergence in probability as $n \to \infty$. In terms of mixing times, (1.14) implies in particular that for any $\varepsilon \in (0, 1)$:

$$(1.15) \qquad \frac{T_0(\varepsilon)}{T_{\mathrm{ENT}}} \xrightarrow{\mathbb{P}} 1,$$

The fact that the distance to equilibrium approaches a step function, or equivalently that the ε-mixing time is to leading order insensitive to the value of $\varepsilon \in (0, 1)$, is commonly referred to as a *cutoff phenomenon*; see e.g. [14, 19] for a review. We also refer to [20, 4, 5] for similar results in the case of undirected graphs. We stress that a fundamental difference between the case of undirected graphs and the case of directed graphs considered here is that the underlying stationary distribution π_0 is not known explicitly in the directed case.

We now formulate our main results. To obtain explicit asymptotic statements we shall assume that $\alpha = \alpha(n) \in (0, 1)$ is a sequence such that $\alpha \to 0$ and such that the limit

$$(1.16) \qquad \gamma = \lim_{n \to \infty} \alpha \, T_{\mathrm{ENT}} \in [0, \infty]$$

exists, with possibly $\gamma = 0$ or $\gamma = \infty$. We call $\mathcal{S}_n$ the set of all probability measures on $[n]$.

Theorem 2. *Let G be a random graph from either the directed configuration model DCM($\mathbf{d}^\pm$) or the out-configuration model OCM($\mathbf{d}^+$). Let $\alpha = \alpha(n) \in (0, 1)$ be parameters as in (1.16). Then, according to the value of γ there are three scenarios:*

(1) If $\gamma = 0$ then for all $s > 0, s \neq 1$:

$$(1.17) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \left| \mathcal{D}_{\alpha,\lambda}^x(s \, T_{\mathrm{ENT}}) - \vartheta(s) \right| \overset{\mathbb{P}}{\longrightarrow} 0.$$

(2) If $\gamma \in (0, \infty)$ then for all $s > 0, s \neq \gamma$:

$$(1.18) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \left| \mathcal{D}_{\alpha,\lambda}^x(s/\alpha) - e^{-s} \vartheta(s/\gamma) \right| \overset{\mathbb{P}}{\longrightarrow} 0.$$

(3) If $\gamma = \infty$ then for all $s > 0$:

$$(1.19) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \left| \mathcal{D}_{\alpha,\lambda}^x(s/\alpha) - e^{-s} \right| \overset{\mathbb{P}}{\longrightarrow} 0.$$

In terms of mixing times, Theorem 2 implies the following statements.

Corollary 3. *In the setting of Theorem 2, the following holds uniformly with respect to λ:*

(1) If $\gamma = 0$ then for all $\varepsilon \in (0, 1)$

$$(1.20) \qquad \frac{T_{\alpha,\lambda}(\varepsilon)}{T_{\mathrm{ENT}}} \overset{\mathbb{P}}{\longrightarrow} 1,$$

(2) If $\gamma \in (0, \infty)$:

$$(1.21) \qquad \frac{T_{\alpha,\lambda}(\varepsilon)}{T_{\mathrm{ENT}}} \overset{\mathbb{P}}{\longrightarrow} \begin{cases} 1 & \text{if } \varepsilon \in (0, e^{-\gamma}) \\ \frac{1}{\gamma} \log(1/\varepsilon) & \text{if } \varepsilon \in [e^{-\gamma}, 1). \end{cases}$$

(3) If $\gamma = \infty$ then for all $\varepsilon \in (0, 1)$:

$$(1.22) \qquad \alpha \, T_{\alpha,\lambda}(\varepsilon) \overset{\mathbb{P}}{\longrightarrow} \log(1/\varepsilon).$$

The trichotomy displayed in Theorem 2 and Corollary 3 reflects the competition between two distinct mechanisms of relaxation to equilibrium: the simple random walk dominates in the first scenario, while the λ-resampling dominates in the third; the intermediate scenario interpolates between the two extremes; see Figure 1.

Remarkably, essentially the same trichotomy was uncovered recently by [3] in a model of random walk on dynamically evolving undirected graphs. In that case, the role of the resampling is played by the underlying reshuffling of the graph edges. It is interesting to observe that, in contrast with the undirected case considered in [3], in our setting the two competing processes may well have very distinct goals, and the overall stationary distribution $\pi_{\alpha,\lambda}$ is the result of a nontrivial balance.

To give some guidelines, below we illustrate the main ideas involved in the proof.

The starting point is the observation that the distance to stationarity $\mathcal{D}_{\alpha,\lambda}^x(t)$ satisfies the following general identity at all times t, for all choices of the parameter α and distribution λ:

$$(1.23) \qquad \left\| P_{\alpha,\lambda}^t(x, \cdot) - \pi_{\alpha,\lambda} \right\|_{\mathrm{TV}} = (1 - \alpha)^t \left\| P^t(x, \cdot) - \pi_{\alpha,\lambda} P^t \right\|_{\mathrm{TV}}.$$
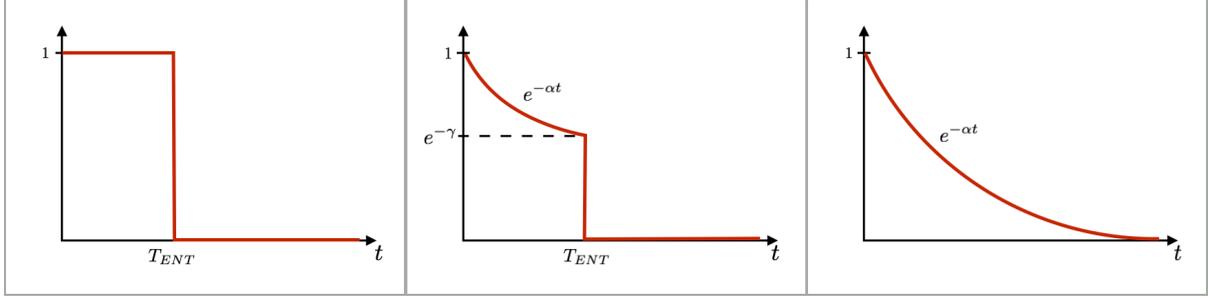
Figure 1

Here we use the notation $\mu P^t(y) = \sum_{x \in V} \mu(x) P^t(x, y)$ for the distribution at time t of the simple random walk started at a random vertex distributed according to some distribution μ. The relation (1.23) follows from a simple coupling argument; see Proposition 8 below. Moreover, the stationary distribution admits the power series expansion

$$\tag{1.24} \pi_{\alpha,\lambda} = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k \lambda P^k \,,$$

see Proposition 7 below. A particularly simple special case is when the resampling distribution λ equals the stationary distribution π_0. Indeed, in this case the stationary distribution is the result of a trivial balance and $\pi_{\alpha,\lambda} = \pi_0$, so that (1.23) becomes

$$\tag{1.25} \mathcal{D}^x_{\alpha,\pi_0}(t) = (1-\alpha)^t \mathcal{D}^x_0(t) \,.$$

Therefore, when $\lambda = \pi_0$ the results in Theorem 2 are an immediate consequence of Theorem 1. Moreover, this shows that the trichotomy in Theorem 2 follows from Theorem 1 whenever the distribution $\lambda \in \mathcal{S}_n$ is such that

$$\tag{1.26} \|\pi_{\alpha,\lambda} - \pi_0\|_{\mathrm{TV}} \xrightarrow{\mathbb{P}} 0,$$

since in this case $\pi_{\alpha,\lambda} P^t$ is well approximated by π_0, and the the three claims in Theorem 2 would follow from (1.23). As we shall see, the approximation (1.26) is rather straightforward in the first scenario. Indeed, if $\alpha T_{\mathrm{ENT}} \to 0$ then the simple random walk has enough time to reach equilibrium between successive resampling events and (1.26) holds uniformly in $\lambda \in \mathcal{S}_n$, see Proposition 16 below. The second and third scenarios require a different approach since one cannot expect (1.26) to hold for all $\lambda \in \mathcal{S}_n$. There is however a special class of distributions, that we refer to as *widespread*, which does satisfy (1.26) in all three scenarios.

Definition 4 (Widespread measure). *A sequence of probability measures $\lambda = \lambda_n$ on $[n]$ is* widespread *if*

  (i) There exists $\delta > 0$ such that

$$\tag{1.27} |\lambda|_\infty = \max_{x \in [n]} \lambda(x) = O(n^{-1/2-\delta}).$$

  (ii) Bounded ℓ_2-distance from the uniform distribution:

$$\tag{1.28} \frac{1}{n} \sum_{j \in [n]} (1 - n\lambda(j))^2 = O(1).$$

Note that there is no requirement on the minimum of $\lambda(x)$, so that large portions of the set of vertices are allowed to receive zero mass. An important property of widespread measures is that, if we start with such a distribution λ, then the time needed to reach stationarity for the simple random walk is much smaller than the entropic time T_{ENT}. More precisely we shall establish the following facts.

Lemma 5. *Let G be a random graph from either the directed configuration model $DCM(\mathbf{d}^{\pm})$ or the out-configuration model $OCM(\mathbf{d}^+)$. If $\lambda = \lambda_n$ is widespread, then for any sequence $t = t(n) \to \infty$,*

$$(1.29) \qquad \left\| \lambda P^t - \pi_0 \right\|_{TV} \xrightarrow{\mathbb{P}} 0.$$

Moreover, in all three scenarios (1.26) *holds for every widespread distribution λ.*

The result in Lemma 5 illustrates well the mechanism behind the trichotomy in the case of widespread measures λ, but it is far from explaining the general phenomenon described in Theorem 2. For instance, if $\lambda = \delta_z$ is a Dirac mass at a vertex z, then $\lambda P^t = P^t(z, \cdot)$ and therefore (1.29) must fail for all $t = sT_{\mathrm{ENT}}$, with $s \in (0, 1)$ fixed, since by Theorem 1 we know that in this case

$$(1.30) \qquad \min_{z \in [n]} \left\| P^t(z, \cdot) - \pi_0 \right\|_{TV} \xrightarrow{\mathbb{P}} 1.$$

Moreover, the stationary distribution π_{α, δ_z} can be very far from π_0 in both scenarios 2 and 3. In particular, using our analysis in Section 3 one can check that in scenario 3,

$$(1.31) \qquad \min_{z \in [n]} \left\| \pi_{\alpha, \delta_z} - \pi_0 \right\|_{TV} \xrightarrow{\mathbb{P}} 1.$$

While we believe the result in Lemma 5 to be of interest in its own, the proof of Theorem 2 will be based on a different approach.

The first observation is that the identity (1.23) together with the result of Theorem 1 is already sufficient to establish all the upper bounds on the distance $\mathcal{D}_{\alpha, \lambda}^x(t)$ required in the proof of Theorem 2, see Section 4 for the details. On the other hand, some extra work is needed for the proof of the lower bounds on $\mathcal{D}_{\alpha, \lambda}^x(t)$. A key technical point for establishing the desired lower bounds will be the following fact concerning scenarios 2 and 3.

Lemma 6. *Let G be a random graph from either the directed configuration model $DCM(\mathbf{d}^{\pm})$ or the out-configuration model $OCM(\mathbf{d}^+)$. For fixed $\gamma > 0$, including $\gamma = \infty$, and $s \in (0, \gamma)$, for any sequence $\alpha \to 0$, satisfying $\alpha T_{\mathrm{ENT}} \to \gamma$, and $t = s/\alpha$:*

$$(1.32) \qquad \min_{\lambda \in \mathcal{S}_n} \min_{x \in [n]} \left\| P^t(x, \cdot) - \pi_{\alpha, \lambda} P^t \right\|_{TV} \xrightarrow{\mathbb{P}} 1.$$

Essentially, (1.32) says that the t-step evolution of the random walk starting at any given vertex x is singular with respect to the evolution starting at the page rank distribution, as soon as $t \leq (1 - \varepsilon)T_{\mathrm{ENT}}$ for some fixed $\varepsilon > 0$. The uniformity in $x \in [n]$ and $\lambda \in \mathcal{S}_n$ in Lemma 6 is a delicate matter. We shall see that for general $\lambda \in \mathcal{S}_n$, if $\alpha T_{\mathrm{ENT}} \to \gamma > 0$, then $\pi_{\alpha, \lambda} P^t$ is a nontrivial mixture of π_0 and another measure μ_λ, see Lemma 13 below for the precise version of this statement. Depending on the nature of $\lambda \in \mathcal{S}_n$, the measure μ_λ can be either supported on a small subset of $[n]$, e.g. if $\lambda = \delta_z$ for some z, or very spread out, e.g. if λ is widespread as in Definition 4. We shall however show that structural features of the random random graph G and the fact that $\alpha \to 0$ imply that the measure μ_λ cannot concentrate any mass on the support of the distribution $P^t(x, \cdot)$ and thus μ_λ and $P^t(x, \cdot)$ are approximately singular. We refer to Section 6 for the derivation of this anti-concentration phenomenon. Since π_0 and $P^t(x, \cdot)$ are approximately singular for $t \leq (1 - \varepsilon)T_{\mathrm{ENT}}$ as in (1.30), this will be sufficient to prove Lemma 6.

The rest of the paper is arranged as follows: the next section establishes the basic identities (1.23) and (1.24) and some more preliminary material; Section 3 contains our main technical estimates and the proof of Lemma 6; Section 4 shows how to derive the main results from Lemma 6 and the facts established in Section 2. The discussion of widespread measures and the proof of Lemma 5 form an independent piece of work and are given in Section 5.

2. PRELIMINARIES

Here we collect some simple general facts about the PageRank surf. The statements in this section do not depend on the graph G where the original walk takes place. Therefore, we fix an arbitrary digraph G with vertex set $V = [n]$, and let P be the transition matrix in (1.2). If $d_x^+ = 0$ for some x we may define $P(x, x) = 1$ and $P(x, y) = 0$ for all $y \in V \setminus \{x\}$.

2.1. The stationary distribution $\pi_{\alpha,\lambda}$.

Proposition 7. *For any $\alpha \in (0, 1)$, any probability vector λ, let $P_{\alpha,\lambda}$ be defined by (1.3). There exists a unique probability vector $\pi_{\alpha,\lambda}$ satisfying $\pi_{\alpha,\lambda} P_{\alpha,\lambda} = \pi_{\alpha,\lambda}$. Moreover, $\pi_{\alpha,\lambda}$ is given by*

$$(2.1) \qquad \pi_{\alpha,\lambda} = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k \lambda P^k.$$

Proof. The equation $\pi_{\alpha,\lambda} P_{\alpha,\lambda} = \pi_{\alpha,\lambda}$ is equivalent to

$$\pi_{\alpha,\lambda}(\mathbf{1} - (1-\alpha)P) = \alpha\lambda.$$

Since P is a stochastic matrix, the matrix $\mathbf{1} - (1-\alpha)P$ is strictly diagonally dominant, and therefore invertible. Then (2.1) follows by expanding the expression $\pi_{\alpha,\lambda} = \alpha\lambda(\mathbf{1} - (1-\alpha)P)^{-1}$. $\qquad\square$

In particular, (2.1) and the triangle inequality imply that for any other probability vector μ:

$$(2.2) \qquad \|\pi_{\alpha,\lambda} - \mu\|_{\text{TV}} \le \alpha \sum_{k=0}^{\infty} (1-\alpha)^k \|\lambda P^k - \mu\|_{\text{TV}}.$$

2.2. Walk vs. teleport.
A trajectory of the PageRank surf can be sampled as follows. At each time unit independently, we flip a α-biased coin: if heads (with probability α) then the surfer is teleported to a new vertex, chosen according to λ; if tails (with probability $1 - \alpha$) then the surfer walks one step according to the transition matrix P. The probability associated to this construction will be denoted by $\mathbb{P}$. If τ_α denotes the first time the surfer is teleported, then for all $t \in \mathbb{N}$:

$$(2.3) \qquad \mathbb{P}(\tau_\alpha > t) = (1-\alpha)^t.$$

Proposition 8. *For any $\alpha \in (0, 1)$, any probability vector λ, and all $t \in \mathbb{N}$, $x \in [n]$:*

$$(2.4) \qquad \|P_{\alpha,\lambda}^t(x, \cdot) - \pi_{\alpha,\lambda}\|_{TV} = (1-\alpha)^t \|P^t(x, \cdot) - \pi_{\alpha,\lambda} P^t\|_{TV}.$$

Proof. We use the construction introduced above, and write X_t^x for the position of the surfer at time t with initial vertex x. By using the same sample of the teleporting distribution λ we couple two trajectories X_t^x, X_t^z in such a way that $X_t^x = X_t^z$, for all $t \ge \tau_\alpha$. Therefore, letting $\mathbb{E}$ denote the expectation with respect to this coupling:

$$
\begin{aligned}
P_{\alpha,\lambda}^t(x, y) - P_{\alpha,\lambda}^t(z, y) &= \mathbb{E}\left[\mathbf{1}(X_t^x = y) - \mathbf{1}(X_t^z = y)\right] \\
(2.5) \qquad &= \mathbb{E}\left[\mathbf{1}(X_t^x = y) - \mathbf{1}(X_t^z = y); \tau_\alpha > t\right].
\end{aligned}
$$

Moreover,

$$(2.6) \qquad \mathbb{E}\left[\mathbf{1}(X_t^x = y); \tau_\alpha > t\right] = \mathbb{P}(\tau_\alpha > t)\mathbb{P}(X_t = y | X_0 = x, \tau_\alpha > t) = \mathbb{P}(\tau_\alpha > t)P^t(x, y).$$

Therefore,

$$(2.7) \qquad P_{\alpha,\lambda}^t(x, y) - P_{\alpha,\lambda}^t(z, y) = \mathbb{P}(\tau_\alpha > t)(P^t(x, y) - P^t(z, y)).$$

Multiplying by $\pi_{\alpha,\lambda}(z)$, summing over z, and using (2.3) one obtains

$$(2.8) \qquad P_{\alpha,\lambda}^t(x, y) - \pi_{\alpha,\lambda}(y) = (1 - \alpha)^t \left(P^t(x, y) - [\pi_{\alpha,\lambda}P^t](y)\right).$$

It follows that

$$\|P_{\alpha,\lambda}^t(x, \cdot) - \pi_{\alpha,\lambda}\|_{\mathrm{TV}} = \frac{1}{2}\sum_{y \in V} |P_{\alpha,\lambda}^t(x, y) - \pi_{\alpha,\lambda}(y)|$$

$$= (1 - \alpha)^t \frac{1}{2}\sum_{y \in V} \left|P^t(x, y) - [\pi_{\alpha,\lambda}P^t](y)\right|$$

$$(2.9) \qquad\qquad = (1 - \alpha)^t \|P^t(x, \cdot) - \pi_{\alpha,\lambda}P^t\|_{\mathrm{TV}}.$$

$\square$

Since the total variation distance is always bounded above by 1, Proposition 8 implies the upper bound

$$(2.10) \qquad \mathcal{D}_{\alpha,\lambda}^x(t) = \|P_{\alpha,\lambda}^t(x, \cdot) - \pi_{\alpha,\lambda}\|_{\mathrm{TV}} \leq (1 - \alpha)^t.$$

The latter, in turn, gives the following upper bound on the mixing time.

Corollary 9. *For any $\alpha \in (0, 1)$, any probability vector λ, and all $\varepsilon \in (0, 1)$, the ε-mixing time (1.10) satisfies*

$$(2.11) \qquad T_{\alpha,\lambda}(\varepsilon) \leq \frac{1}{\alpha}\log(1/\varepsilon).$$

A further immediate consequence of Proposition 8 is that if λ is stationary for P, then the distance to equilibrium $\mathcal{D}_{\alpha,\lambda}^x(t)$ takes a simple form.

Corollary 10. *For any $\alpha \in (0, 1)$, for all $x \in V$ and all $t \in \mathbb{N}$, if π_0 is a probability vector such that $\pi_0 P = \pi_0$, then taking $\lambda = \pi_0$,*

$$(2.12) \qquad \mathcal{D}_{\alpha,\pi_0}^x(t) = (1 - \alpha)^t \|P^t(x, \cdot) - \pi_0\|_{\mathrm{TV}}.$$

Proof. From Proposition 7 it follows that $\pi_{\alpha,\pi_0} = \pi_0$, and therefore $\pi_{\alpha,\pi_0}P^t = \pi_0$ for all t. $\square$

Finally, another useful consequence of Proposition 8 is that it allows us to control the distance $\mathcal{D}_{\alpha,\lambda}^x(t)$ in terms of the distance $\mathcal{D}_{\alpha,\pi_0}^x(t)$, for some stationary π_0 as in Corollary 10, by means of the distance between $\pi_{\alpha,\lambda}$ and π_0.

Corollary 11. *For any $\alpha \in (0, 1)$, all $t \in \mathbb{N}$, any probability vector λ, if π_0 is such that $\pi_0 P = \pi_0$,*

$$(2.13) \qquad \max_{x \in V} \left|\mathcal{D}_{\alpha,\lambda}^x(t) - \mathcal{D}_{\alpha,\pi_0}^x(t)\right| \leq \|\pi_{\alpha,\lambda} - \pi_0\|_{TV}.$$

Proof. From the triangle inequality and the fact that $\|\mu P^t - \nu P^t\|_{\mathrm{TV}}$ is monotone in t for all distributions μ, ν, one has

$$(2.14) \qquad \left|\|P^t(x, \cdot) - \pi_{\alpha,\lambda}P^t\|_{\mathrm{TV}} - \|P^t(x, \cdot) - \pi_0\|_{\mathrm{TV}}\right| \leq \|\pi_{\alpha,\lambda} - \pi_0\|_{\mathrm{TV}}.$$

The conclusion then follows from Proposition 8 and Corollary 10. $\square$

3. Main technical estimates

The goal of this section is to prove Lemma 6. The proof is divided into three main steps. The first step is a decomposition of $\pi_{\alpha,\lambda}P^t$ as a mixture of π_0 and a distribution μ_λ defined below. The second and most delicate step is the proof that μ_λ and $P^t(x,\cdot)$ are approximately singular for t and α as in Lemma 6. The third step concludes the desired result collecting the technical estimates established in the first two steps.

3.1. Decomposition of $\pi_{\alpha,\lambda}P^t$.

We start with a useful decomposition of $\pi_{\alpha,\lambda}P^t$ as a mixture of π_0 and a distribution μ_λ defined as follows. Fix $\eta \in (0,1/2)$, $t \leq (1-2\eta)T_{\mathrm{ENT}}$, and define $\mu_\lambda = \mu_\lambda^{\eta,t}$ and $A = A^{\eta,t}$ as

$$(3.1) \qquad A = \sum_{k=0}^{(1-\eta)T_{\mathrm{ENT}}-t} \alpha(1-\alpha)^k, \qquad \mu_\lambda = \frac{1}{A}\sum_{k=0}^{(1-\eta)T_{\mathrm{ENT}}-t} \alpha(1-\alpha)^k \lambda P^{k+t}.$$

Note that μ_λ depends on the graph G while A is deterministic. We consider the case $\alpha T_{\mathrm{ENT}} \to \gamma \in (0,\infty]$ and treat the two cases $\gamma = \infty$ and $\gamma \in (0,\infty)$ separately.

Lemma 12. *Fix $s \in (0,\infty)$ and assume $\alpha T_{\mathrm{ENT}} \to +\infty$ and $t = s/\alpha$. For all $\varepsilon > 0$, there exists $\eta > 0$ such that with high probability:*

$$(3.2) \qquad \max_{\lambda \in \mathcal{S}_n} \|\pi_{\alpha,\lambda}P^t - \mu_\lambda\|_{TV} \leq \varepsilon,$$

and the normalization in (3.1) satisfies $A \geq 1 - \varepsilon$.

Proof. Since $\alpha T_{\mathrm{ENT}} \to +\infty$ we have $t \ll T_{\mathrm{ENT}}$. It follows that $A \to 1$. Using Proposition 7,

$$(3.3) \qquad \pi_{\alpha,\lambda}P^t - A\mu_\lambda = \sum_{k=(1-\eta)T_{\mathrm{ENT}}-t}^{\infty} \alpha(1-\alpha)^k \lambda P^{k+t},$$

and therefore $\|\pi_{\alpha,\lambda}P^t - \mu_\lambda\|_{TV} \leq (1-A)$. $\qquad\square$

Lemma 13. *Fix $\gamma \in (0,\infty)$, $s \in (0,\gamma)$ and assume $\alpha T_{\mathrm{ENT}} \to \gamma$ and $t = s/\alpha$. For all $\varepsilon > 0$, there exists $\eta > 0$ such that with high probability:*

$$(3.4) \qquad \max_{\lambda \in \mathcal{S}_n} \|\pi_{\alpha,\lambda}P^t - A\mu_\lambda - (1-A)\pi_0\|_{TV} \leq \varepsilon,$$

where $A = A^{\eta,t}$ and $\mu_\lambda = \mu_\lambda^{\eta,t}$ are given in (3.1).

Proof. For any $a < b$, $z \in [n]$, define the probability vector

$$(3.5) \qquad \nu_{a,b}^z = \frac{1}{Z_{a,b}}\sum_{k=aT_{\mathrm{ENT}}}^{bT_{\mathrm{ENT}}-1} \alpha(1-\alpha)^k P^{k+t}(z,\cdot), \qquad Z_{a,b} = \sum_{k=aT_{\mathrm{ENT}}}^{bT_{\mathrm{ENT}}-1} \alpha(1-\alpha)^k.$$

Since $t = s/\alpha$, $s \in (0,\gamma)$, and $\alpha T_{\mathrm{ENT}} \to \gamma$ we may take $\eta > 0$ small enough and assume that $t = \kappa T_{\mathrm{ENT}}$, $\kappa \in (0, 1-2\eta)$. Using Proposition 7, letting δ_z denote the Dirac mass at z:

$$(3.6) \qquad \pi_{\alpha,\delta_z}P^t = Z_{0,1-\eta-\kappa}\,\nu_{0,1-\eta-\kappa}^z + Z_{1-\eta-\kappa,1+\eta-\kappa}\,\nu_{1-\eta-\kappa,1+\eta-\kappa}^z + Z_{1+\eta-\kappa,\infty}\,\nu_{1+\eta-\kappa,\infty}^z$$

Note that $Z_{0,1-\eta-\kappa} = A$, and $\nu_{0,1-\eta-\kappa}^z = \mu_{\delta_z}$. We show that the middle term above is negligible and that $\nu_{1+\eta-\kappa,\infty}^z$ is well approximated by π_0. If $\alpha T_{\mathrm{ENT}} \to \gamma \in (0,\infty)$, by Riemann integration it follows that for

all n large enough

$$(3.7) \qquad Z_{1-\eta-\kappa,1+\eta-\kappa} \leq \sum_{k=(1-\eta-\kappa)T_{\mathrm{ENT}}}^{(1+\eta-\kappa)T_{\mathrm{ENT}}} \alpha e^{-k\alpha} \leq C\eta,$$

for some constant $C > 0$. Next, using the monotonicity in time of total variation distance and Theorem 1, w.h.p.

$$(3.8) \qquad \max_{z\in[n]} \sup_{k\geq(1+\eta-\kappa)T_{\mathrm{ENT}}} \|P^{k+t}(z,\cdot) - \pi_0\|_{\mathrm{TV}} \leq \max_{z\in[n]} \|P^{(1+\eta)T_{\mathrm{ENT}}}(z,\cdot) - \pi_0\|_{\mathrm{TV}} \leq \eta.$$

It follows that w.h.p.

$$(3.9) \qquad \max_{z} \|\nu^z_{1+\eta-\kappa,\infty} - \pi_0\|_{\mathrm{TV}} \leq \eta.$$

Writing $\pi_{\alpha,\lambda}P^t = \sum_z \lambda(z)\pi_{\alpha,\delta_z}P^t$ and taking A and μ_λ as in (3.1) concludes the proof. $\qquad\square$

3.2. **Singularity of μ_λ and $P^t(x,\cdot)$.** The key to this result is a property of the random walk that was established in [6, 7]. Roughly speaking this says that with high probability, for most vertices x, the trajectory $\{X^x_u,\ u \leq t\}$ of the walk started at x up to time t is supported by a "small" directed tree $\mathcal{T}_x(t)$ rooted at x provided that $t \leq (1 - \eta)T_{\mathrm{ENT}}$ where η is an arbitrary positive constant. As a result the distribution $P^t(x,\cdot)$ is rather strongly localized. We shall see that the distribution μ_λ, depending on the nature of λ, could be either supported on a small subset of $[n]$ (e.g. if $\lambda = \delta_z$ for some $z \in [n]$) or very spread out (e.g. if λ is widespread). The approximate singularity of μ_λ and $P^t(x,\cdot)$ turns out to be the result of a delicate structural property of the digraph G which guarantees that even if μ_λ is localized it must be sufficiently smeared out and cannot concentrate on the support of $P^t(x,\cdot)$. We first recall the construction of the tree $\mathcal{T}_x(t)$ and then address the structural properties ensuring this anti-concentration.

3.2.1. *The tree $\mathcal{T}_z(t)$.* Given the digraph G, the tree $\mathcal{T}_z(t)$, for fixed $t \leq (1 - \eta)T_{\mathrm{ENT}}$, can be discovered algorithmically as described in [6, Section 6.2] and [7, Section 4.1]. We recall the detailed construction for model 1. A very similar construction can be given for model 2; see [7, Section 4.1].

Below we describe a sequence of digraphs $\mathcal{G}^0, \mathcal{G}^1, \ldots, \mathcal{G}^\kappa$ such that at each step $\mathcal{G}^\ell$ is a subset of the out-neighborhood of z of height t in G and such that $\mathcal{G}^\ell$ is obtained from $\mathcal{G}^{\ell-1}$ by adding a single edge of G. Moreover, we obtain a sequence of directed trees $\mathcal{T}^0, \mathcal{T}^1, \ldots, \mathcal{T}^\kappa$ such that for every ℓ, $\mathcal{T}^\ell$ is a spanning tree of $\mathcal{G}^\ell$. The tree $\mathcal{T}_z(t)$ will be defined as $\mathcal{T}_z(t) = \mathcal{T}^\kappa$.

Initially all matchings of tails and heads in G are unrevealed and $\mathcal{G}^0 = \mathcal{T}^0 = \{z\}$; let $\partial_-\mathcal{T}^\ell$ (resp. $\partial_+\mathcal{T}^\ell$) denote the set of unrevealed heads (resp. tails) whose endpoint belongs to $\mathcal{T}^\ell$; the height $h(e)$ of a tail $e \in \partial_+\mathcal{T}^\ell$ is defined as 1 plus the number of edges in the unique path in $\mathcal{T}^\ell$ from z to the endpoint of e; the weight of $e \in \partial_+\mathcal{T}^\ell$ is defined as

$$(3.10) \qquad \mathbf{w}(e) = \prod_{i=0}^{h(e)-1} \frac{1}{d^+_{x_i}},$$

where $(z = x_0, x_1, \ldots, x_{h(e)-1})$ denotes the path in $\mathcal{T}^\ell$ from z to the endpoint of e; we then iterate the following steps:

- a tail $e \in \partial_+\mathcal{T}^\ell$ is selected with maximal weight among all $e \in \partial_+\mathcal{T}^\ell$ with $h(e) \leq t$ and $\mathbf{w}(e) \geq \mathbf{w}_{\min} := n^{-1+\eta^2}$ (using an arbitrary ordering of the tails to break ties);
- the head f matched to e in G is revealed, and $\mathcal{G}^{\ell+1}$ is obtained from $\mathcal{G}^\ell$ by adding the edge ef;
- if f was not in $\partial_-\mathcal{T}^\ell$, then its endpoint and the edge ef are added to $\mathcal{T}^\ell$ to form $\mathcal{T}^{\ell+1}$.

The process stops at $\ell = \kappa$ when there are no tails $e \in \partial_+ \mathcal{T}^\kappa$ with height $h(e) \le t$ and weight $\mathbf{w}(e) \ge \mathbf{w}_{\min}$. Note that $\mathcal{T}^\ell$ is a directed spanning tree of $\mathcal{G}^\ell$ at each step. The tree $\mathcal{T}_z(t)$ is defined as $\mathcal{T}^\kappa$. After the construction of the tree $\mathcal{T}_z(t)$, exactly κ edges of G have been revealed, some of which may not belong to $\mathcal{T}_z(t)$. Note that $\mathcal{G}^\kappa$ has κ edges and coincides with the union of all directed paths from z which have length at most t and at least probability $\mathbf{w}_{\min}$ with respect to the random walk started at z. As in [6, Lemma 11], [7, Lemma 7], it is not difficult to see that when exploring the out-neighborhood of z in this way the number κ, regardless of the realization of G, is bounded as

$$(3.11) \qquad \kappa \le n^{1 - \frac{\eta^2}{2}}.$$

Let us recall the following key facts established in [6, Section 6] for model 1 and in [7, Section 4] for model 2. For every $\eta > 0$, for every $t \le (1 - \eta)T_{\mathrm{ENT}}$, the trajectory $(X_0, \dots, X_t)$ of the random walk started at z in G satisfies with high probability $(X_0, \dots, X_t) \subset \mathcal{T}_z(t)$ for most initial positions z. More precisely, let $Q_z(\cdot)$ denote the quenched law of the random walk $(X_0, X_1 \dots)$ in G started at z. Let V_* denote the set of $z \in [n]$ such that $B_{z,\hbar}^+$ is a directed tree, where $\hbar := \frac{1}{10} \log_\Delta(n)$, and $B_{z,\hbar}^+$ denotes the out neighborhood of z of height $\hbar$ in G (that is the subgraph of G induced by the set of vertices which can be reached from z with a path of length at most $\hbar$). Then, from [6, Proposition 10, part (ii)], and [7, Lemma 11], one has

$$(3.12) \qquad \min_{z \in V_*} Q_z\left((X_0, \dots, X_t) \subset \mathcal{T}_z(t)\right) \xrightarrow{\mathbb{P}} 1,$$

where the notation $(X_0, \dots, X_t) \subset \mathcal{T}_z(t)$ indicates that the walk up to time t traverses only edges of $\mathcal{T}_z(t)$.

3.2.2. *Key technical estimate.* Let $\mathcal{A}_x(t)$ denote the set of vertices in $\mathcal{T}_x(t)$ that have distance from x exactly t in $\mathcal{T}_x(t)$. Recall the definition $\mu_\lambda = \mu_\lambda^{\eta,t}$ in (3.1).

Lemma 14. *Assume $\alpha T_{\mathrm{ENT}} \to \gamma \in (0, \infty]$. Fix $\eta \in (0, 1/2)$ and take $t \le (1 - 2\eta)T_{\mathrm{ENT}}$. Then, for all $\varepsilon > 0$, with high probability*

$$(3.13) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \mu_\lambda(\mathcal{A}_x(t)) \le \varepsilon.$$

The proof of Lemma 14 is based on a structural property of the graph G which says that the intersections of the trees $\mathcal{T}_z(u)$ and $\mathcal{T}_x(t)$, where x, z are two arbitrary vertices and $t \le u \le (1 - \eta)T_{\mathrm{ENT}}$, are such that, with high probability, for all $x, z \in [n]$, no path in $\mathcal{T}_z(u)$ can intersect more than K times the set $\mathcal{A}_x(t)$ where K is a suitably large constant. Let us use the notation $\mathcal{P}_z(u)$ for the set of paths in $\mathcal{T}_z(u)$ having length exactly u and, for all $\mathfrak{p} \in \mathcal{P}_z(u)$, let $V(\mathfrak{p})$ denote the set of vertices along that path. Note that the endpoint of $\mathfrak{p} \in \mathcal{P}_z(u)$ is necessarily a vertex of $\mathcal{A}_z(u)$ and $|V(\mathfrak{p})| = u$, since $\mathcal{T}_z(u)$ is a tree.

Lemma 15. *Fix $\eta \in (0, 1/2)$. For every $x, z \in [n]$ and $t \le u \le (1 - \eta)T_{\mathrm{ENT}}$,*

$$(3.14) \qquad \mathbb{P}\left(\exists\, \mathfrak{p} \in \mathcal{P}_z(u) : \mathcal{A}_x(t) \cap V(\mathfrak{p}) > K\right) \le n^{-3},$$

for all n large enough, where $K = (9 + 3 \log_2 \Delta)/\eta^2$. In particular, the event

$$\mathcal{E}_K = \{\forall x, z \in [n],\ \forall \mathfrak{p} \in \mathcal{P}_z(u),\ \mathcal{A}_x(t) \cap V(\mathfrak{p}) \le K\}$$

holds with high probability.

Proof. We sample the pair $(\mathcal{T}_x(t), \mathcal{T}_z(u))$ in the random digraph G by generating first the subgraph $\mathcal{T}_x(t)$, and then the subgraph $\mathcal{T}_z(u)$ conditionally on $\mathcal{T}_x(t)$. The construction of $\mathcal{T}_x(t)$ follows the steps described by the algorithm in Section 3.2.1 with the understanding that, for model 1 the head f to be matched to
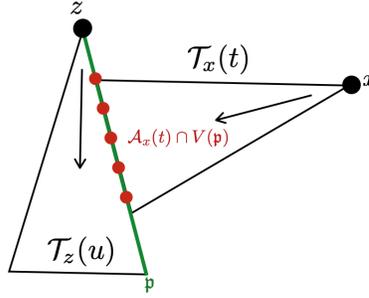
Figure 2. A sketch of the unlikely event in Lemma 15

the tail e to form $\mathcal{G}^{\ell+1}$ is chosen uniformly at random among all $m - \ell$ heads that are unmatched after the ℓ-th step, while for model 2 the tail e has to be connected to a uniformly random vertex in $[n]$. The process terminates when the tree $\mathcal{T}_x(t)$ has been fully generated after κ_x steps. A crucial feature of this construction is that the tails of all vertices $v \in \mathcal{A}_x(t)$ are unmatched once the tree $\mathcal{T}_x(t)$ has been generated. Moreover, the number of vertices of $\mathcal{T}_x(t)$ satisfies, as in (3.11)

$$(3.15) \qquad\qquad |\mathcal{T}_x(t)| \leq \kappa_x \leq n^{1-\frac{\eta^2}{2}}.$$

Next, we generate the tree $\mathcal{T}_z(u)$, conditionally on $\mathcal{T}_x(t)$. This is done by starting at z and by repeating the same steps for the construction described in Section 3.2.1 with the difference that if at step ℓ a tail e is chosen which had already been matched during the generation of $\mathcal{T}_x(t)$ then the corresponding edge is included in the construction (and possibly in the tree being generated). The process terminates when the tree $\mathcal{T}_z(u)$ has been fully generated after κ_z steps. Thus, after $\kappa_x + \kappa_z$ steps we have a sample from the joint distribution of $\mathcal{T}_x(t)$ and $\mathcal{T}_z(u)$ in G. Note that the total number of edges of G discovered after the generation of both trees is $\kappa_x + \kappa_z \leq 2n^{1-\eta^2/2}$. Let $\{\mathcal{F}_\ell\}$ denote the filtration associated to this generation process, so that $\mathcal{F}_{\kappa_x}$ is the σ-field associated to the tree $\mathcal{T}_x(t)$.

During the process generating $\mathcal{T}_z(u)$ conditionally on $\mathcal{T}_x(t)$, we say that a *bad matching* occurs at step ℓ if the tail chosen at that step is currently unmatched (that is it was not revealed during the sampling of $\mathcal{T}_x(t)$) and it gets connected to a vertex y that was already discovered in $\mathcal{T}_x(t)$. The first key observation is that the conditional probability of a bad matching at step ℓ given $\mathcal{F}_\ell$ is uniformly bounded above by

$$(3.16) \qquad\qquad p := 2\Delta n^{-\eta^2/2}.$$

Indeed, in the case of model 1 this probability is at most $\kappa_x \Delta/(m - \kappa_x - \kappa_z)$, while for model 2 this probability is at most κ_x/n. In either case it is less than the number p defined in (3.16) for all n large enough.

The second key observation is that if a path $\mathfrak{p} \in \mathcal{P}_z(u)$ is such that $\mathcal{A}_x(t) \cap V(\mathfrak{p}) > K$ then at least K bad matchings have occurred during the formation of that given path. To see this, observe that after a vertex $y \in \mathcal{A}_x(t)$ is visited for the first time during the construction of $\mathcal{T}_z(u)$, the tails of y will be all matched (at suitable steps $\ell_1, \ldots, \ell_{d_y^+}$) to a uniformly sampled head among the ones that are currently unmatched (for model 1) or to a uniformly random vertex (for model 2). Indeed, the tails of all vertices $v \in \mathcal{A}_x(t)$ all have the same weight $\mathbf{w}(e)$ and all of them are unmatched after the tree $\mathcal{T}_x(t)$ has been generated. Also, by definition, every path in $\mathcal{T}_z(u)$ can visit a given vertex y at most once, and after a visit to $y \in \mathcal{A}_x(t)$ it has to return to $\mathcal{T}_x(t)$ with a bad matching in order to visit some other $y' \in \mathcal{A}_x(t)$.

Hence, the number of visits to $\mathcal{A}_x(t)$ in a given path $\mathfrak{p} \in \mathcal{P}_z(u)$ is at most the number of bad matchings occurred along that path +1. The extra 1 comes from the fact that z could have started already inside $\mathcal{T}_x(t)$, for instance if $z \in \mathcal{A}_x(t)$.
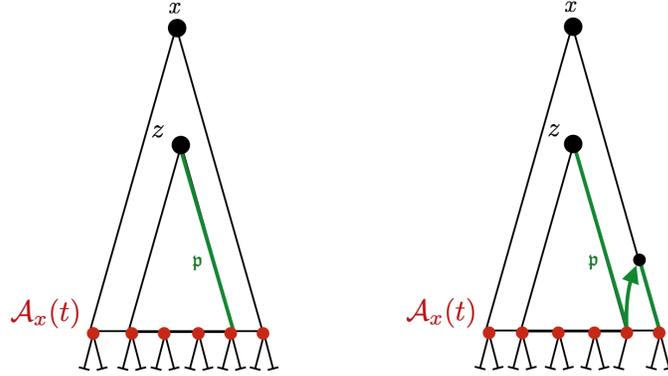
Figure 3. A first visit to $\mathcal{A}_x(t)$ of a path $\mathfrak{p} \in \mathcal{P}_z(u)$ (left) and an illustration of a bad matching (right).

Next, consider an auxiliary directed tree with random marks $\tilde{\mathcal{T}}$ defined as follows: $\tilde{\mathcal{T}}$ is a directed regular tree with deterministic offspring Δ and height u, with independent and identically distributed Bernoulli(p) marks on its edges, where p is as in (3.16). Edges whose Bernoulli mark is 1 are colored red. A path of length u from the root to one of the leaves is called *bad* if it has at least K red edges. The previous construction then shows that the number of $\mathfrak{p} \in \mathcal{P}_z(u)$ such that $\mathcal{A}_x(t) \cap V(\mathfrak{p}) > K$ is stochastically dominated by the number of bad paths in $\tilde{\mathcal{T}}$. The probability that a given path in $\tilde{\mathcal{T}}$ is bad is given by

$$\mathbb{P}\left(\mathrm{Bin}(u, p) \geq K\right) \leq (up)^K.$$

Therefore, the probability that there exists a bad path in $\tilde{\mathcal{T}}$ is at most $\Delta^u (up)^K$. Since $u \leq T_{\mathrm{ENT}} \leq \log n / \log 2$, it follows that

$$(3.17) \qquad \mathbb{P}\left(\exists\, \mathfrak{p} \in \mathcal{P}_z(u),\, \mathcal{A}_x(t) \cap V(\mathfrak{p}) > K\right) \leq \Delta^{T_{\mathrm{ENT}}} (T_{\mathrm{ENT}} p)^K \leq n^{\frac{\log \Delta}{\log 2} - K\eta^2/3},$$

for all n sufficiently large. Taking $K = (9 + 3\log_2 \Delta)/\eta^2$ concludes the proof of (3.14). A union bound then implies that the event $\mathcal{E}_K$ in the statement of the lemma holds with probability at least $1 - 1/n$. $\quad\square$

Once Lemma 15 is available, we can prove Lemma 14.

Proof of Lemma 14. The distribution μ_λ satisfies $\mu_\lambda = \sum_{z \in [n]} \lambda(z) \mu_z$, where $\mu_z := \mu_{\delta_z}$. Hence, it is sufficient to prove that w.h.p.

$$(3.18) \qquad \max_{z,x \in [n]} \mu_z(\mathcal{A}_x(t)) \leq \varepsilon.$$

We write

$$(3.19) \qquad \mu_z(\mathcal{A}_x(t)) = \frac{1}{A} \sum_{k=0}^{(1-\eta)T_{\mathrm{ENT}} - t} \alpha(1-\alpha)^k Q_z(X_{k+t} \in \mathcal{A}_x(t)),$$

where Q_z is defined as in (3.12). As in [6, Propositon 6] one shows that for both models, with high probability:

$$\text{(3.20)} \qquad \max_{z \in [n]} Q_z(X_{k_0} \in V \setminus V_*) \leq 2^{-k_0},$$

for any fixed constant k_0. Hence, for all non negative integers k, t with $k_0 \leq k + t \leq (1 - \eta)T_{\text{ENT}}$:

$$\max_{z \in [n]} Q_z(X_{k+t} \in \mathcal{A}_x(t)) \leq \max_{v \in V_*} Q_v(X_{k+t-k_0} \in \mathcal{A}_x(t)) + 2^{-k_0}.$$

Set $u = (1 - \eta)T_{\text{ENT}}$. For any $v \in V_*$ we write

$$Q_v(X_{k+t-k_0} \in \mathcal{A}_x(t)) \leq Q_v\left((X_0, \ldots, X_u) \not\subset \mathcal{T}_v(u)\right) + \sum_{\mathfrak{p} \in \mathcal{P}_v(u)} Q_v\left((X_0, \ldots, X_u) = \mathfrak{p}, \ X_{k+t-k_0} \in \mathcal{A}_x(t)\right).$$

By (3.12), the first term in the right hand side is w.h.p. less than ε uniformly in $v \in V_*$, for any fixed $\varepsilon > 0$. The second term, taking the summation over $k \in [k_0, (1 - \eta)T_{\text{ENT}} - t]$ satisfies, w.h.p.

$$\sum_{\mathfrak{p} \in \mathcal{P}_v(u)} \sum_{k=k_0}^{(1-\eta)T_{\text{ENT}} - t} \alpha(1 - \alpha)^k Q_v\left((X_0, \ldots, X_u) = \mathfrak{p}, \ X_{k+t-k_0} \in \mathcal{A}_x(t)\right)$$

$$\leq \alpha \sum_{\mathfrak{p} \in \mathcal{P}_v(u)} \sum_{k=0}^{u} Q_v\left((X_0, \ldots, X_u) = \mathfrak{p}, \ X_k \in \mathcal{A}_x(t)\right)$$

$$\text{(3.21)} \qquad \leq \alpha K \sum_{\mathfrak{p} \in \mathcal{P}_v(u)} Q_v\left((X_0, \ldots, X_u) = \mathfrak{p}\right) \leq \alpha K,$$

where K is the constant from Lemma 15 and we have used the fact that the event $\mathcal{E}_K$ from Lemma 15 holds with high probability. From (3.19)-(3.21), noting that the first k_0 terms in the summation over k contribute to (3.19) at most $\alpha k_0 / A$, we then obtain, w.h.p.

$$\text{(3.22)} \qquad \max_{z, x \in [n]} \mu_z(\mathcal{A}_x(t)) \leq \varepsilon + \frac{\alpha k_0}{A} + \frac{\alpha K}{A} + 2^{-k_0}.$$

Since $\alpha T_{\text{ENT}} \to \gamma \in (0, \infty]$ and $\alpha \to 0$, it follows that for all fixed $\eta > 0$ one has $\alpha/A \to 0$ as $n \to \infty$. Since the parameters $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ are arbitrary this implies the desired conclusion. $\qquad\square$

3.3. **Proof of Lemma 6.** Assume $\alpha T_{\text{ENT}} \to \gamma$, and $t = s/\alpha$, with $s \in (0, \gamma)$ as in the statement of Lemma 6. The proof below applies to both cases $\gamma \in (0, \infty)$ and $\gamma = \infty$. We are going to show that for every fixed $\varepsilon > 0$, there exists an event $\mathcal{E}_\varepsilon = \mathcal{E}_\varepsilon(n)$ such that $\mathbb{P}(\mathcal{E}_\varepsilon) \to 1$, $n \to \infty$, and such that on $\mathcal{E}_\varepsilon$ for all $x \in [n]$ there are sets $\mathcal{C}_x \subset V$ satisfying

$$\text{(3.23)} \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \mu_\lambda(\mathcal{C}_x) \leq \varepsilon, \quad \max_{x \in [n]} \pi_0(\mathcal{C}_x) \leq \varepsilon, \quad \min_{x \in [n]} P^t(x, \mathcal{C}_x) \geq 1 - \varepsilon.$$

Indeed, using the decompositions in Lemma 12 (for the case $\gamma = \infty$) and in Lemma 13 (for the case $\gamma \in (0, \infty)$), if (3.23) holds then w.h.p.

$$\left\|P^t(x, \cdot) - \pi_{\alpha, \lambda} P^t\right\|_{\text{TV}} \geq \left\|P^t(x, \cdot) - A\mu_\lambda - (1 - A)\pi_0\right\|_{\text{TV}} - \varepsilon$$

$$\text{(3.24)} \qquad \geq P^t(x, \mathcal{C}_x) - A\mu_\lambda(\mathcal{C}_x) - (1 - A)\pi_0(\mathcal{C}_x) - \varepsilon \geq 1 - 3\varepsilon,$$

where it is understood that $A = 1$ if $\gamma = \infty$. Since (3.24) holds uniformly in λ and x, this completes the proof of Lemma 6. We turn to the proof of (3.23).

It is important that the estimates in (3.23) hold for all $\varepsilon > 0$ and $\eta > 0$ small enough (but fixed), where η is the parameter implicit in the definition of $\mu_\lambda = \mu_\lambda^{\eta, t}$. Since $s \in (0, \gamma)$, we may assume

$t \leq (1 - 2\eta)T_{\mathrm{ENT}}$ by taking η small enough. By Theorem 1 we know that for each $\delta > 0$, with high probability there exists sets $\mathcal{B}_x$, such that for all $x \in [n]$:

$$(3.25) \qquad\qquad \pi_0(\mathcal{B}_x) \leq \delta, \qquad P^t(x, \mathcal{B}_x) \geq 1 - \delta.$$

For $\varepsilon > 0$, take $k_0 = k_0(\varepsilon)$ such that $2^{-k_0} \leq \varepsilon/2$, and call $t' = t - k_0$. Since $t = \Theta(\alpha^{-1})$ and $\alpha \to 0$, we have $t' > 0$ for all n large enough. For all $x \in [n]$, call V_x the subset of vertices in $y \in V_*$ such that $P^{k_0}(x, y) > 0$. Define

$$\mathcal{C}_x = \mathcal{B}_x \cap \left(\cup_{y \in V_x} \mathcal{A}_y(t') \right).$$

From (3.12) we know that, for all $\delta > 0$, with high probability,

$$(3.26) \qquad\qquad \min_{y \in V_*} P^{t'}(y, \mathcal{A}_y(t')) \geq 1 - \delta.$$

By (3.20), (3.25) and (3.26) we obtain

$$P^t(x, \mathcal{C}_x) \geq P^t(x, \cup_{y \in V_x} \mathcal{A}_y(t')) - P^t(x, \mathcal{B}_x^c)$$

$$(3.27) \qquad\qquad \geq \min_{y \in V_x} P^{t'}(y, \mathcal{A}_y(t')) - 2^{-k_0} - \delta \geq 1 - 2^{-k_0} - 2\delta.$$

From (3.25) we also know that $\pi_0(\mathcal{C}_x) \leq \pi_0(\mathcal{B}_x) \leq \delta$. Taking $\delta = \varepsilon/4$, this and (3.27) imply the last two items in (3.23) since $2^{-k_0} \leq \varepsilon/2$. It remains to estimate $\mu_\lambda(\mathcal{C}_x)$. Since $\max_x |V_x| \leq \Delta^{k_0}$ we obtain

$$(3.28) \qquad\qquad \mu_\lambda(\mathcal{C}_x) \leq \mu_\lambda(\cup_{y \in V_x} \mathcal{A}_y(t')) \leq \Delta^{k_0} \max_{y \in [n]} \mu_\lambda(\mathcal{A}_y(t')).$$

From Lemma 14 we see that with high probability, uniformly in λ and x, (3.28) is at most $\Delta^{k_0}\delta$ for any fixed $\delta > 0$. Thus taking $\delta = \Delta^{-k_0}\varepsilon$ concludes the proof of Lemma 6.

4. PROOF OF THE TRICHOTOMY

In this section we show how to prove Theorem 2 from the facts established above. Thus, G is a random graph from either the directed configuration model $\mathrm{DCM}(\mathbf{d}^\pm)$ or the out-configuration model $\mathrm{OCM}(\mathbf{d}^+)$, where the degree sequences satisfy the assumptions (1.6) and (1.7) respectively, and π_0 denotes the (w.h.p.) unique stationary distribution for the simple random walk on G.

4.1. **Scenario 1.** We begin with scenario 1, namely when $\alpha T_{\mathrm{ENT}} \to 0$.

Proposition 16. *For any sequence α such that $\alpha T_{\mathrm{ENT}} \to 0$,*

$$(4.1) \qquad\qquad \max_{\lambda \in \mathcal{S}_n} \|\pi_{\alpha,\lambda} - \pi_0\|_{TV} \xrightarrow{\mathbb{P}} 0.$$

Proof. We need to show that, uniformly in λ, for any $\delta > 0$,

$$(4.2) \qquad\qquad \|\pi_{\alpha,\lambda} - \pi_0\| \leq \delta, \qquad w.h.p.$$

The upper bound (2.2) shows that for all $t \in \mathbb{N}$:

$$(4.3) \qquad\qquad \|\pi_{\alpha,\lambda} - \pi_0\|_{\mathrm{TV}} \leq (1 - (1 - \alpha)^t) + \sum_{k > t} \alpha(1 - \alpha)^k \|\lambda P^k - \pi_0\|_{\mathrm{TV}}.$$

Take $t = sT_{\mathrm{ENT}}$, with some fixed $s > 1$, and observe that by Theorem 1 we know that for all $k > t$, for all λ:

$$\|\lambda P^k - \pi_0\|_{\mathrm{TV}} \leq \|\lambda P^{sT_{\mathrm{ENT}}} - \pi_0\|_{\mathrm{TV}}$$

$$(4.4) \qquad\qquad \leq \max_{x \in V} \|P^{sT_{\mathrm{ENT}}}(x, \cdot) - \pi_0\|_{\mathrm{TV}} \leq \delta/2, \qquad w.h.p.$$

In particular, using $\alpha t \to 0$:

$$(4.5) \qquad \max_{\lambda \in \mathcal{S}_n} \|\pi_{\alpha,\lambda} - \pi_0\|_{\mathrm{TV}} \leq (1 - (1-\alpha)^t) + \delta/2 \leq \delta, \qquad w.h.p.$$

$\square$

The claim (1.17) is thus a consequence of Corollary 10, Corollary 11 and Theorem 1.

4.2. **Scenario 3.** Suppose $\alpha T_{\mathrm{ENT}} \to +\infty$, and $t = s/\alpha$ for some fixed $s \in (0, \infty)$. From Lemma 6, Proposition 8 and the upper bound (2.10) we obtain:

$$(4.6) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \left| \left\| P_{\alpha,\lambda}^t(x, \cdot) - \pi_{\alpha,\lambda} \right\|_{\mathrm{TV}} - (1-\alpha)^t \right| \xrightarrow{\mathbb{P}} 0.$$

Equivalently,

$$(4.7) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \left| \mathcal{D}_{\alpha,\lambda}^x(s/\alpha) - e^{-s} \right| \xrightarrow{\mathbb{P}} 0.$$

This proves (1.19).

4.3. **Scenario 2.** Here $\alpha T_{\mathrm{ENT}} \to \gamma \in (0, \infty)$. We take $t = s/\alpha$, with fixed $s \in (0, \infty)$. We consider separately the case $s \in (\gamma, \infty)$ and the case $s \in (0, \gamma)$.

Suppose first $s \in (\gamma, \infty)$. By Proposition 8 and the triangle inequality

$$\mathcal{D}_{\alpha,\lambda}^x(s/\alpha) \leq \left\| P^t(x, \cdot) - \pi_{\alpha,\lambda} P^t \right\|_{\mathrm{TV}}$$
$$(4.8) \qquad \leq \left\| P^t(x, \cdot) - \pi_0 \right\|_{\mathrm{TV}} + \max_{y \in V} \left\| \pi_0 - P^t(y, \cdot) \right\|_{\mathrm{TV}}.$$

Since $s \in (\gamma, \infty)$, for some $\varepsilon > 0$ we have $t \geq (1 + \varepsilon) T_{\mathrm{ENT}}$. Therefore, by Theorem 1 it follows that

$$(4.9) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \mathcal{D}_{\alpha,\lambda}^x(s/\alpha) \xrightarrow{\mathbb{P}} 0, \qquad s \in (\gamma, \infty).$$

On the other hand, suppose that $s \in (0, \gamma)$. Here we can apply Lemma 6, Proposition 8 and the upper bound (2.10), as in Section 4.2 above, to obtain

$$(4.10) \qquad \max_{\lambda \in \mathcal{S}_n} \max_{x \in [n]} \left| \mathcal{D}_{\alpha,\lambda}^x(s/\alpha) - e^{-s} \right| \xrightarrow{\mathbb{P}} 0, \qquad s \in (0, \gamma).$$

Combining (4.9) and (4.10), we have proved (1.18).

5. WIDESPREAD MEASURES

The goal of this section is to prove Lemma 5. We remark that the statement $\|\pi_{\alpha,\lambda} - \pi_0\|_{\mathrm{TV}} \to 0$ in probability is a consequence of (1.29). Indeed, fix any sequence $\alpha = \alpha(n) \to 0$, and take $t = t(n) \to \infty$ such that $\alpha t \to 0$. From (1.29) we know that

$$(5.1) \qquad \left\| \lambda P^t - \pi_0 \right\|_{\mathrm{TV}} \xrightarrow{\mathbb{P}} 0.$$

As in (4.3), from the upper bound (2.2) and the monotonicity in time of total variation distance to stationarity we obtain:

$$(5.2) \qquad \|\pi_{\alpha,\lambda} - \pi_0\|_{\mathrm{TV}} \leq (1 - (1-\alpha)^t) + \|\lambda P^t - \pi_0\|_{\mathrm{TV}}.$$

Using (5.1) and $\alpha t \to 0$ we conclude the proof. Thus, we are left to prove (1.29).

In the special case where $\lambda = \mu_{\mathrm{in}}$, and for the directed configuration model DCM($\mathbf{d}^{\pm}$), a similar result was already obtained in [6]. Here we are going to prove it for the case of the out-configuration model OCM($\mathbf{d}^+$) as well, and more importantly we are going to extend it to the case of an arbitrary

widespread probability measure λ. Following the approach in [6], the proof of Lemma 5 will be based on the construction of a martingale approximation for the distribution λP^t. The latter, in turn, rests on a branching approximation which allows one to couple the in-neighbourhood of a uniformly distributed random vertex of G with a marked Galton-Watson tree up to depth $t = o(\log n)$.

We start with the definition of the relevant branching processes and the associated martingales. These will later be used in a coupling argument to provide an approximate description of the in-neighbourhood of a vertex in our random graphs, and of the stationary distribution at that vertex. Since the constructions differ slightly for the two models DCM($\mathbf{d}^\pm$) or OCM($\mathbf{d}^+$) we will define two distinct random trees $\mathcal{T}^-(\mathbf{d}^\pm)$ and $\mathcal{T}^-(\mathbf{d}^+)$.

5.1. **The marked Galton-Watson trees $\mathcal{T}^-(\mathbf{d}^\pm)$, $\mathcal{T}^-(\mathbf{d}^+)$.** Given $n \in \mathbb{N}$, and a double sequence $\mathbf{d}^\pm$ of degrees satisfying (1.5) and (1.6), for each $i \in [n]$, we define the rooted random marked tree $\mathcal{T}_i^-(\mathbf{d}^\pm)$ recursively with the following rules:

- the root is given the mark i;
- every vertex with mark j has d_j^- children, each of which is given independently the mark $k \in [n]$ with probability d_k^+/m.

On the other hand, given $n \in \mathbb{N}$, and a sequence $\mathbf{d}^+$ of degrees satisfying (1.7), for each $i \in [n]$, the rooted random marked tree $\mathcal{T}_i^-(\mathbf{d}^+)$ is defined by:

- the root is given the mark i;
- regardless of its own mark every vertex has, for each $j \in [n]$ independently with probability d_j^+/n, a child with mark j.

There are several differences between the two trees $\mathcal{T}_i^-(\mathbf{d}^\pm)$ and $\mathcal{T}_i^-(\mathbf{d}^+)$. In the first case the number of children of a given vertex is a deterministic function of the vertex's mark, whereas in the second case it is a random variable D that can be written as

$$(5.3) \qquad\qquad D = \sum_{j \in [n]} Y_j\,, \qquad Y_j = \mathrm{Ber}(d_j^+/n),$$

where the Y_j are independent Bernoulli random variables with parameters d_j^+/n. In particular, the average number of children of any given vertex in $\mathcal{T}_i^-(\mathbf{d}^+)$ is

$$(5.4) \qquad\qquad \mathbb{E}[D] = \sum_{j \in [n]} \frac{d_j^+}{n} = \frac{m}{n} = \langle d \rangle\,.$$

Since D can be zero, in contrast with the tree $\mathcal{T}_i^-(\mathbf{d}^\pm)$, the tree $\mathcal{T}^-(\mathbf{d}^+)$ is finite with positive probability. However, the two trees share several common features and we shall try to treat the two cases in a unified fashion as much as possible.

We write $\mathbf{o}$ for the root and $\mathbf{x}, \mathbf{y}$ for other vertices of the tree, with the notation $\mathbf{y} \to \mathbf{x}$ if $\mathbf{y}$ is a child of $\mathbf{x}$. Each vertex $\mathbf{x}$ of the tree has a mark, which we denote by $i(\mathbf{x})$. If $\mathcal{I}$ denotes an independent uniformly random $i \in [n]$, and the root is given the mark $i(\mathbf{o}) = \mathcal{I}$, then we write $\mathcal{T}^-(\mathbf{d}^\pm) = \mathcal{T}_\mathcal{I}^-(\mathbf{d}^\pm)$ and $\mathcal{T}^-(\mathbf{d}^+) = \mathcal{T}_\mathcal{I}^-(\mathbf{d}^+)$. Notice that $\mathcal{T}^-(\mathbf{d}^\pm)$ and $\mathcal{T}^-(\mathbf{d}^+)$ have the same average degree at the root, given by (5.4). We often write $\mathcal{T}^-$ for short if this creates no confusion. For each $t \in \mathbb{N}$ we let $\mathcal{T}^{-,t}$ denote the set of vertices in the generation t of the tree. Each vertex $\mathbf{x} \in \mathcal{T}^{-,t}$ has a unique path $(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1, \mathbf{x}_0)$ connecting it to the root with $\mathbf{x}_t = \mathbf{x}$ and $\mathbf{x}_0 = \mathbf{o}$. To any such $\mathbf{x}$ we associate the weight

$$(5.5) \qquad\qquad w(\mathbf{x}) = \prod_{u=1}^{t} \frac{1}{d_{i(\mathbf{x}_u)}^+}.$$

If $\mathcal{T}^{-,t}$ coincides with the in-neighbourhood of $\mathbf{o}$ in a digraph G, then $w(\mathbf{x})$ is the probability that the simple random walk on G goes from $\mathbf{x}$ to $\mathbf{o}$ in t steps.

5.2. **Martingale approximation.** Given a function $\varphi : [n] \mapsto \mathbb{R}$, we define the process

(5.6) $$X_t(\varphi) = \sum_{\mathbf{x} \in \mathcal{T}^{-,t}} \varphi(i(\mathbf{x}))w(\mathbf{x}), \qquad X_0(\varphi) = \varphi(i(\mathbf{o})).$$

We write $\mathcal{F}_t$ for the σ-algebra generated by the random tree $\mathcal{T}^-$ up to and including generation t.

Lemma 17. *Let $\mathcal{T}^-$ be either $\mathcal{T}^-(\mathbf{d}^{\pm})$ or $\mathcal{T}^-(\mathbf{d}^+)$, and write $\bar{\varphi} = \sum_{j=1}^n \varphi(j)$. Then, for all $t \in \mathbb{N}$:*

(5.7) $$\mathbb{E}[X_t(\varphi)|\mathcal{F}_{t-1}] = X_{t-1}(\bar{\varphi}\mu_{\mathrm{in}}).$$

Proof. Let the symbol $\sum_{\mathbf{y}\to\mathbf{x}}$ denote the sum over the set of children of $\mathbf{x}$ and note the symbolic identity

(5.8) $$\sum_{\mathbf{y}\in\mathcal{T}^{-,t}} \equiv \sum_{\mathbf{x}\in\mathcal{T}^{-,t-1}} \sum_{\mathbf{y}\to\mathbf{x}}.$$

Therefore,

$$\mathbb{E}[X_t(\varphi)|\mathcal{F}_{t-1}] = \sum_{\mathbf{x}\in\mathcal{T}^{-,t-1}} \mathbb{E}\left[\sum_{\mathbf{y}\to\mathbf{x}} \varphi(i(\mathbf{y}))w(\mathbf{y})|\mathcal{F}_{t-1}\right]$$

(5.9) $$= \sum_{\mathbf{x}\in\mathcal{T}^{-,t-1}} w(\mathbf{x}) \mathbb{E}\left[\sum_{\mathbf{y}\to\mathbf{x}} \frac{\varphi(i(\mathbf{y}))}{d_{i(\mathbf{y})}^+} \Big| \mathcal{F}_{t-1}\right].$$

For the tree $\mathcal{T}^-(\mathbf{d}^{\pm})$ we have

(5.10) $$\mathbb{E}\left[\sum_{\mathbf{y}\to\mathbf{x}} \frac{\varphi(i(\mathbf{y}))}{d_{i(\mathbf{y})}^+} \Big| \mathcal{F}_{t-1}\right] = d_{i(\mathbf{x})}^- \sum_{j=1}^n \frac{d_j^+}{m} \frac{\varphi(j)}{d_j^+} = \bar{\varphi}\,\mu_{\mathrm{in}}(i(\mathbf{x})).$$

For the tree $\mathcal{T}^-(\mathbf{d}^+)$ we have

(5.11) $$\mathbb{E}\left[\sum_{\mathbf{y}\to\mathbf{x}} \frac{\varphi(i(\mathbf{y}))}{d_{i(\mathbf{y})}^+} \Big| \mathcal{F}_{t-1}\right] = \sum_{j=1}^n \frac{d_j^+}{n} \frac{\varphi(j)}{d_j^+} = \bar{\varphi}\,\mu_{\mathrm{in}}(i(\mathbf{x})).$$

This proves (5.7). $\qquad\qquad\square$

In particular, when $\varphi = \mu_{\mathrm{in}}$, then

$$\mathbb{E}[X_t(\mu_{\mathrm{in}})|\mathcal{F}_{t-1}] = X_{t-1}(\mu_{\mathrm{in}}), \qquad t \in \mathbb{N}.$$

Therefore, $X_t(\mu_{\mathrm{in}})$ is a martingale with respect to the filtration $\mathcal{F}_t$. It is convenient to normalize it and consider instead the martingale defined as

(5.12) $$M_t = nX_t(\mu_{\mathrm{in}}) = \sum_{\mathbf{x}\in\mathcal{T}^{-,t}} n\mu_{\mathrm{in}}(i(\mathbf{x}))w(\mathbf{x}), \qquad M_0 = n\mu_{\mathrm{in}}(i(\mathbf{o})).$$

Notice that $\mathbb{E}[M_t] = \mathbb{E}[M_0] = n\mathbb{E}[\mu_{\mathrm{in}}(\mathcal{I})] = 1$. In the case of model 1, the following convergence result was already discussed in [6, Proposition 15].

Proposition 18. *For every fixed n, as $t \to \infty$ the martingale M_t converges to a limit M_∞, both almost surely and in L^2 (see [22, Ch. 12]) and for all $t \in \mathbb{N}$:*

$$\mathbb{E}[(M_t - M_\infty)^2] = C\rho^t \tag{5.13}$$

where the constants ρ, C are given by

$$\rho = \sum_{j=1}^{n} \mu_{\text{in}}(j) \frac{1}{d_j^+}, \qquad C = \begin{cases} \frac{n}{m(1-\rho)} \sum_{j=1}^{n} \frac{(d_j^- - d_j^+)^2}{md_j^+} & \text{model 1} \\ \frac{\rho - 1/n}{1-\rho} & \text{model 2} \end{cases} \tag{5.14}$$

Proof. Consider the increments

$$\Delta_t = M_{t+1} - M_t. \tag{5.15}$$

Reasoning as in Lemma 17, for both models we write

$$\Delta_t = \sum_{\mathbf{x} \in \mathcal{T}^{-,t}} n\mu_{\text{in}}(i(\mathbf{x}))w(\mathbf{x})\psi(\mathbf{x}), \tag{5.16}$$

where ψ is defined as

$$\psi(\mathbf{x}) = \sum_{\mathbf{y} \to \mathbf{x}} \frac{\mu_{\text{in}}(i(\mathbf{y}))}{\mu_{\text{in}}(i(\mathbf{x}))d_{i(\mathbf{y})}^+} - 1. \tag{5.17}$$

As in Lemma 17 one has $\mathbb{E}[\psi(\mathbf{x}) \,|\, \mathcal{F}_t] = 0$. Let us compute $\mathbb{E}[\psi(\mathbf{x})^2 \,|\, \mathcal{F}_t]$. For the tree $\mathcal{T}^-(\mathbf{d}^\pm)$, we can rewrite

$$\psi(\mathbf{x}) = \sum_{\mathbf{y} \to \mathbf{x}} \left[\frac{\mu_{\text{in}}(i(\mathbf{y}))}{\mu_{\text{in}}(i(\mathbf{x}))d_{i(\mathbf{y})}^+} - \frac{1}{d_{i(\mathbf{x})}^-} \right]. \tag{5.18}$$

Therefore,

$$\mathbb{E}[\psi(\mathbf{x})^2 \,|\, \mathcal{F}_t] = d_{i(\mathbf{x})}^- \sum_{j=1}^{n} \frac{d_j^+}{m} \left(\frac{d_j^-}{d_{i(\mathbf{x})}^- d_j^+} - \frac{1}{d_{i(\mathbf{x})}^-} \right)^2 = \frac{C_1}{d_{i(\mathbf{x})}^-}, \tag{5.19}$$

where we use the notation

$$C_1 = \sum_{j=1}^{n} \frac{(d_j^- - d_j^+)^2}{md_j^+}.$$

For the tree $\mathcal{T}^-(\mathbf{d}^+)$ we have

$$\mathbb{E}[\psi(\mathbf{x})^2 \,|\, \mathcal{F}_t] = \mathbb{E}\left[\left(\sum_{\mathbf{y} \to \mathbf{x}} \frac{1}{d_{i(\mathbf{y})}^+} \right)^2 - 2\sum_{\mathbf{y} \to \mathbf{x}} \frac{1}{d_{i(\mathbf{y})}^+} + 1 \,\Big|\, \mathcal{F}_t \right]$$

$$= \sum_{j \neq j'} \frac{d_j^+ d_{j'}^+}{n^2} \frac{1}{d_j^+ d_{j'}^+} + \sum_{j} \frac{d_j^+}{n} \frac{1}{(d_j^+)^2} - 2\sum_{j} \frac{d_j^+}{n} \frac{1}{d_j^+} + 1 = \rho - \frac{1}{n}, \tag{5.20}$$

where ρ is as in (5.14). Since $\mathbb{E}[\psi(\mathbf{x})\psi(\mathbf{x}') \,|\, \mathcal{F}_t] = 0$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{T}^{-,t}$ with $\mathbf{x} \neq \mathbf{x}'$,

$$\mathbb{E}[\Delta_t^2 \,|\, \mathcal{F}_t] = \sum_{\mathbf{x} \in \mathcal{T}^{-,t}} n^2 \mu_{\text{in}}(i(\mathbf{x}))^2 w(\mathbf{x})^2 \mathbb{E}[\psi(\mathbf{x})^2 \,|\, \mathcal{F}_t]. \tag{5.21}$$

Therefore, combining (5.19) and (5.20) we have

$$(5.22) \qquad \mathbb{E}[\Delta_t^2 \,|\, \mathcal{F}_t] = C(1-\rho) \sum_{\mathbf{x} \in \mathcal{T}^{-,t}} n\mu_{\mathrm{in}}(i(\mathbf{x}))w(\mathbf{x})^2 \,,$$

where ρ, C are given by (5.14). Furthermore, observe that in both models one has

$$\mathbb{E}[\Delta_t^2 \,|\, \mathcal{F}_{t-1}] = \mathbb{E}\left[\mathbb{E}[\Delta_t^2 \,|\, \mathcal{F}_t] \,|\, \mathcal{F}_{t-1}\right]$$

$$= C(1-\rho) \sum_{\mathbf{x} \in \mathcal{T}^{-,t-1}} n\mu_{\mathrm{in}}(i(\mathbf{x}))w(\mathbf{x})^2 \mathbb{E}\left[\sum_{\mathbf{y} \to \mathbf{x}} \frac{\mu_{\mathrm{in}}(i(\mathbf{y}))}{\mu_{\mathrm{in}}(i(\mathbf{x}))(d_{i(\mathbf{y})}^+)^2} \,|\, \mathcal{F}_{t-1}\right]$$

$$(5.23) \qquad = C(1-\rho)\rho \sum_{\mathbf{x} \in \mathcal{T}^{-,t-1}} n\mu_{\mathrm{in}}(i(\mathbf{x}))w(\mathbf{x})^2 = \rho\,\mathbb{E}[\Delta_{t-1}^2 \,|\, \mathcal{F}_{t-1}].$$

Thus, iterating we obtain

$$(5.24) \qquad \mathbb{E}[\Delta_t^2] = \mathbb{E}[\Delta_0^2]\rho^t = C(1-\rho)\mathbb{E}[n\mu_{\mathrm{in}}(\mathcal{I})]\rho^t = C(1-\rho)\rho^t.$$

Since $d_j^+ \geq 2$ one has $\rho \leq 1/2$. Thus M_t is a martingale bounded in L^2, and therefore $M_t \to M_\infty$ almost surely and in L^2, for some $M_\infty \in L^2$. Using the orthogonality $\mathbb{E}[\Delta_t \Delta_{t'}] = 0$ for all $t \neq t'$, (5.13) follows by summing (5.24) from t to $+\infty$. □

Remark 19. *For each fixed $n \in \mathbb{N}$, one can characterise the random variable M_∞ as the solution to a distributional fixed point equation. For the directed configuration model DCM($\mathbf{d}^\pm$) this is discussed in [6, Lemma 16]. With a similar reasoning, for the out-configuration model OCM($\mathbf{d}^+$) one obtains that*

$$(5.25) \qquad M_\infty \overset{d}{=} \sum_{j=1}^{n} \frac{Y_j}{d_j^+} M_{\infty,j},$$

where $\overset{d}{=}$ stands for equality of distributions, $M_{\infty,j}$ are i.i.d. copies of M_∞ and Y_j are independent Bernoulli random variables with parameter d_j^+/n.

The next result will be crucial for the analysis of widespread measures. Notice that the constant $\gamma(\lambda)$ appearing in the estimate below is bounded uniformly in n if and only if λ satisfies (1.28).

Proposition 20. *For any probability vector λ, and any $t \in \mathbb{N}$:*

$$(5.26) \qquad \mathbb{E}[(M_t - nX_t(\lambda))^2] \leq \gamma(\lambda)\rho^t \,,$$

where $\rho \in (0,1)$ is as in Proposition 18 and $\gamma(\lambda)$ is defined as

$$(5.27) \qquad \gamma(\lambda) = \frac{n}{2} \sum_{j=1}^{n} (\lambda(j) - \mu_{\mathrm{in}}(j))^2$$

Proof. Setting $\varphi(j) = n(\mu_{\mathrm{in}}(j) - \lambda(j))$, we write $M_t - nX_t(\lambda) = X_t(\varphi)$. Since $\bar{\varphi} = 0$, Lemma 17 shows that $\mathbb{E}[M_t - nX_t(\lambda)|\mathcal{F}_{t-1}] = 0$. We now compute

$$\Gamma_t := \mathbb{E}[(M_{t+1} - nX_{t+1}(\lambda))^2|\mathcal{F}_t].$$

Using $\bar{\varphi} = 0$ one has

$$\Gamma_t = \mathbb{E}[X_{t+1}(\varphi)^2|\mathcal{F}_t]$$

$$(5.28) \qquad = \sum_{\mathbf{x} \in \mathcal{T}^{-,t}} w(\mathbf{x})^2 \,\mathbb{E}\left[\left(\sum_{\mathbf{y} \to \mathbf{x}} \frac{\varphi(i(\mathbf{y}))}{d_{i(\mathbf{y})}^+}\right)^2 \,\Big|\, \mathcal{F}_t\right].$$

For the tree $\mathcal{T}^-(\mathbf{d}^\pm)$ we have

$$(5.29)\qquad \mathbb{E}\left[\left(\sum_{\mathbf{y}\to\mathbf{x}}\frac{\varphi(i(\mathbf{y}))}{d^+_{i(\mathbf{y})}}\right)^2\Bigg|\,\mathcal{F}_t\right]=d^-_{i(\mathbf{x})}\sum_{j=1}^n\frac{d^+_j}{m}\frac{\varphi(j)^2}{(d^+_j)^2}=\mu_{\mathrm{in}}(i(\mathbf{x}))\sum_{j=1}^n\frac{\varphi(j)^2}{d^+_j}.$$

On the other hand for the tree $\mathcal{T}^-(\mathbf{d}^+)$ we have

$$\mathbb{E}\left[\left(\sum_{\mathbf{y}\to\mathbf{x}}\frac{\varphi(i(\mathbf{y}))}{d^+_{i(\mathbf{y})}}\right)^2\Bigg|\,\mathcal{F}_t\right]=\sum_{j\neq j'}\frac{\varphi(j)\varphi(j')}{n^2}+\sum_j\frac{\varphi(j)^2}{nd^+_j}$$

$$(5.30)\qquad\qquad\qquad\qquad\qquad\quad =\frac{1}{n}\sum_{j=1}^n\frac{\varphi(j)^2}{d^+_j}\left(1-\frac{d^+_j}{n}\right)$$

Summarising, we have shown that

$$(5.31)\qquad \Gamma_t=C(\lambda)\sum_{\mathbf{x}\in\mathcal{T}^{-,t}}n\mu_{\mathrm{in}}(i(\mathbf{x}))w(\mathbf{x})^2\,,\qquad C(\lambda)=\frac{1}{n}\begin{cases}\sum_{j=1}^n\frac{\varphi(j)^2}{d^+_j}&\text{model 1}\\[2mm]\sum_{j=1}^n\frac{\varphi(j)^2}{d^+_j}\left(1-\frac{d^+_j}{n}\right)&\text{model 2}\end{cases}$$

Thus, the same argument used in (5.23) implies that in both models

$$(5.32)\qquad\qquad\qquad\qquad \mathbb{E}[\Gamma_t\,|\,\mathcal{F}_{t-1}]=\rho\Gamma_{t-1},$$

where ρ is defined as in (5.14). Therefore,

$$(5.33)\qquad\qquad\qquad \mathbb{E}[\Gamma_t]=\mathbb{E}[\Gamma_0]\rho^t=C(\lambda)\mathbb{E}[n\mu_{\mathrm{in}}(\mathcal{I})]\rho^t=C(\lambda)\rho^t.$$

The desired bound follows from the fact that in both models $C(\lambda)\leq\gamma(\lambda)$. $\qquad\qquad\square$

5.3. Branching approximation for in-neighbourhoods.
The t-in-neighbourhood of a vertex v, denoted $B^-_{v,t}$, is defined as the subgraph of G induced by the set of directed paths of length t in G which terminate at vertex v. Here we observe that for any fixed $v\in[n]$, if t is a small multiple of $\log n$ then with high probability $B^-_{v,t}$ can be coupled to the first t generations of the random trees defined in Section 5.1. We consider the two models separately.

5.3.1. $B^-_{v,t}$ for DCM($\mathbf{d}^\pm$).
Recall that each vertex x has d^-_x heads and d^+_x tails. Call E^-_x and E^+_x the sets of heads and tails at x respectively. The uniform bijection ω between heads and tails, viewed as a matching, can be sampled by iterating the following steps until there are no unmatched heads left:

1) pick an unmatched head e_- according to some priority rule;
2) pick an unmatched tail e_+ uniformly at random;
3) match e_- with e_+, i.e. set $\omega(e_+)=e_-$.

Note that this gives the desired uniform distribution over matchings regardless of the priority rule chosen at step 1. The graph G is obtained by adding a directed edge (x,y) whenever $e_-\in E^-_y$ and $e_+\in E^+_x$ in step 3 above.

To generate $B^-_{v,t}$ only, one can start at vertex v and run the previous sequence of steps, by giving priority to those unmatched heads which have minimal distance from vertex v, until this minimal distance exceeds t, at which point the process stops. During the process, say that a vertex x is *exposed* if at least one of the tails $e_+\in E^+_x$ or heads $e_-\in E^-_x$ has been already matched. Notice that as long as in step 2 no tail e_+ is picked from exposed vertices, the resulting digraph is a directed tree.

Let us now describe a coupling of the in-neighbourhood $B_{v,t}^-$ and the marked tree $\mathcal{T}_{v,t}^-(\mathbf{d}^\pm)$, where $\mathcal{T}_{v,t}^-(\mathbf{d}^\pm)$ stands for the marked tree $\mathcal{T}_v^-(\mathbf{d}^\pm)$ up to generation t; see Section 5.1 for the definition of $\mathcal{T}_v^-(\mathbf{d}^\pm)$. Clearly, step 2 above can be modified by picking e uniformly at random among all (matched or unmatched) tails and rejecting the proposal if the tail was already matched. The tree can then be generated by iteration of the same sequence of steps with the difference that at step 2 we never reject the proposal and at step 3 we add a new leaf to the current tree, with mark x if $e_+ \in E_x^+$, together with a new set of d_x^- unmatched heads attached to it. Call τ the first time that a uniform random choice among all tails gives $e_+ \in E_x^+$ with x already in the tree. By construction, the in-neighbourhood and the tree coincide up to time τ. At the k-th iteration, the probability of picking a tail with a mark already used is at most $k\Delta/m$, where Δ is the maximum degree. Therefore, by a union bound,

$$(5.34) \qquad \mathbb{P}(\tau \le k) \le \frac{k^2 \Delta}{m}.$$

Taking $k = \Delta^{t+1}$ steps, we have necessarily uncovered the whole in-neighbourhood $B_{v,t}^-$. Thus, we have proved the following statement.

Lemma 21. *The t-in-neighbourhood $B_{v,t}^-$ and the marked tree $\mathcal{T}_{v,t}^-(\mathbf{d}^\pm)$ can be coupled in such a way that*

$$(5.35) \qquad \mathbb{P}\left(B_{v,t}^- \ne \mathcal{T}_{v,t}^-(\mathbf{d}^\pm)\right) \le \frac{\Delta^{2t+3}}{m}.$$

5.3.2. $B_{v,t}^-$ *for* OCM($\mathbf{d}^+$). Recall that each vertex x has d_x^+ tails, and call E_x^+ the sets of tails at x. Consider the following *exploration process* of the in-neighbourhood at a fixed vertex v. The process is defined as a triple $(\mathcal{C}_\ell, \mathcal{A}_\ell, \phi_\ell)$ where $\mathcal{C}_\ell, \mathcal{A}_\ell \subset [n]$ are respectively the *completed* set and the *active* set at time ℓ, and $\phi_\ell : [n] \mapsto \mathbb{Z}_+$ is a map such that $\phi_\ell(y) \in \{0, \dots, d_y^+\}$ for each $y \in [n]$, $\ell \in \mathbb{Z}_+$. At time zero we set $\mathcal{C}_0 = \emptyset, \mathcal{A}_0 = \{v\}$, and $\phi_0(y) = 0$ for all $y \in [n]$. The ℓ-th iteration of the exploration determines the triple $(\mathcal{C}_\ell, \mathcal{A}_\ell, \phi_\ell)$ by executing the following steps:

1) pick a vertex $x \in \mathcal{A}_{\ell-1}$ according to some priority rule;
2) for each $y = 1, \dots, n$ independently, sample $X_{\ell,y}$ defined as the Bernoulli random variable with parameter

$$(5.36) \qquad p_{\ell,y} = \frac{d_y^+ - \phi_{\ell-1}(y)}{n - \ell + 1},$$

call V_ℓ the set of $y \in [n]$ such that $X_{\ell,y} = 1$, and define

$$(5.37) \qquad W_\ell = (\mathcal{C}_{\ell-1} \cup \mathcal{A}_{\ell-1})^c \cap V_\ell.$$

3) define the new triple $(\mathcal{C}_\ell, \mathcal{A}_\ell, \phi_\ell)$ as

$$\mathcal{C}_\ell = \mathcal{C}_{\ell-1} \cup \{x\}, \quad \mathcal{A}_\ell = \mathcal{A}_{\ell-1} \setminus \{x\} \cup W_\ell, \quad \phi_\ell(y) = \phi_{\ell-1}(y) + \mathbf{1}(y \in V_\ell), \ y = 1, \dots, n.$$

Note that this process stops when $\mathcal{A}_\ell$ becomes empty. Let us call $\tau_\emptyset$ this random time:

$$(5.38) \qquad \tau_\emptyset = \min\{\ell \ge 1 : \ \mathcal{A}_\ell = \emptyset\}.$$

For instance, $\tau_\emptyset = 1$ with probability $\prod_{y=1}^n (1 - d_y^+/n)$. We may construct a digraph $G_v(\ell)$ along with the above process by adding the directed edges (y, x) for all $y \in V_\ell$ at step 2. Notice that when the process stops $G_v(\tau_\emptyset)$ is a sample of the subgraph of G induced by all directed paths in G that terminate at v. In particular, if the priority in step 1 is given to x which have minimal distance to v, and if we stop the process as soon as all active vertices have distance to v larger than t in the current graph $G_v(\ell)$, we obtain the in-neighbourhood of v at distance t, namely the digraph $B_{v,t}^-$ for the model OCM($\mathbf{d}^+$). More formally, if τ_t denotes the minimal ℓ such that all $x \in \mathcal{A}_\ell$ have distance to v at least $t + 1$ in $G_v(\ell)$ then,

$B_{v,t}^-$ is given by the subgraph of $G_v(\tau_t \wedge \tau_\emptyset)$ induced by the completed set $\mathcal{C}_{\tau_t \wedge \tau_\emptyset}$, where $a \wedge b$ denotes the minimum of a, b.

Let us remark that the quantity $p_{\ell,y}$ in (5.36) cannot exceed 1. In fact, in case there exists some $\ell \in \mathbb{N}$ such that $p_{\ell,y} = 1$ then it means that at most d_y^+ vertices need to be discovered at step ℓ, and vertex y needs to link to all of them. Hence, $p_{\ell,y}$ stays 1 up to the end of the process.

Let us now describe a coupling of $B_{v,t}^-$ and the marked tree $\mathcal{T}_{v,t}^-(\mathbf{d}^+)$, where we write $\mathcal{T}_{v,t}^-(\mathbf{d}^+)$ for the marked tree $\mathcal{T}_v^-(\mathbf{d}^+)$ up to generation t; see Section 5.1. First, observe that the tree $\mathcal{T}_v^-(\mathbf{d}^+)$ is obtained by iterating the steps above with the difference that at step 2 the probability $p_{\ell,y}$ must be taken always equal to d_ℓ^+/n, and that each $y \in V_\ell$ yields a new child with mark y in the current tree. Let $\mathcal{T}_v^-(\ell)$ denote the tree obtained after ℓ iterations, and let $\Delta = \max_x d_x^+$.

Lemma 22. *The random variables $G_v(\ell), \mathcal{T}_v^-(\ell)$ can be coupled in such a way that for every $\ell \in \mathbb{N}$:*

$$(5.39) \qquad \mathbb{P}(G_v(\ell) \neq \mathcal{T}_v^-(\ell)) \leq \frac{\Delta^2 \ell^2}{n - \ell}.$$

Proof. Let $E_\ell = \{G_v(\ell) \neq \mathcal{T}_v^-(\ell)\}$. Since at time 0 one has $G_v(0) = \mathcal{T}_v^-(0) = \{v\}$, the event E_ℓ satisfies $E_\ell = \cup_{k=1}^\ell E_{k-1}^c \cap E_k$, so that

$$(5.40) \qquad \mathbb{P}\left(G_v(\ell) \neq \mathcal{T}_v^-(\ell)\right) \leq \sum_{k=1}^\ell \mathbb{P}(E_{k-1}^c \cap E_k)$$

Consider now the k-th iteration, and assume that $G_v(k-1) = \mathcal{T}_v^-(k-1)$. Thus, we may pick the same x in step 1 for both samples. At step 2, let $X_{k,y}$ denote the Bernoulli random variables with parameter $p_{k,y}$ used for the sampling of $G_v(k)$ and let $\tilde{X}_{k,y}$ be the Bernoulli random variables with parameter d_y^+/n used for the sampling of $\mathcal{T}_v^-(k)$. The total variation distance between two Bernoulli random variables equals the absolute value of the difference of their parameters. Therefore, for each y independently we may couple $(X_{k,y}, \tilde{X}_{k,y})$ with probability $1 - |p_{k,y} - d_y^+/n|$. Notice that if $G_v(k) \neq \mathcal{T}_v^-(k)$, then either at least one of the pairs $(X_{k,y}, \tilde{X}_{k,y})$ fails to couple, or at least one of the $y \in \mathcal{C}_{k-1} \cup \mathcal{A}_{k-1}$ has $\tilde{X}_{k,y} = 1$. Thus, on the event E_{k-1}^c, the probability of E_k given the history up to the $(k-1)$-th iteration is bounded above by

$$(5.41) \qquad \sum_{y \notin \mathcal{C}_{k-1} \cup \mathcal{A}_{k-1}} |p_{k,y} - d_y^+/n| + \sum_{y \in \mathcal{C}_{k-1} \cup \mathcal{A}_{k-1}} p_{k,y}.$$

If $y \notin \mathcal{C}_{k-1} \cup \mathcal{A}_{k-1}$, then $\phi_{k-1}(y) = 0$ and $p_{k,y} - d_y^+/n = \frac{d_y^+}{n} \cdot \frac{k-1}{n-k+1}$. For the second term we write $|\mathcal{C}_{k-1} \cup \mathcal{A}_{k-1}| \leq Z_{k-1}$, where Z_ℓ denotes the number of edges in the tree $\mathcal{T}_v^-(\ell)$. In conclusion, (5.41) is bounded by

$$\frac{\Delta}{n-k} (k + Z_{k-1}).$$

Thus, letting $\mathcal{F}_\ell$ denote the σ-algebra generated by the two processes up to time ℓ, we have obtained

$$\mathbb{P}(E_{k-1}^c \cap E_k) = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}(E_{k-1}^c \cap E_k) \,|\, \mathcal{F}_{k-1}\right]\right]$$

$$(5.42) \qquad \leq \frac{\Delta}{n-k}(k + \mathbb{E}[Z_{k-1}]).$$

From (5.4) we deduce $\mathbb{E}[Z_{k-1}] = (k-1)\langle d \rangle \leq (k-1)\Delta$. Therefore, the estimate (5.39) follows from (5.40) and (5.42). $\qquad \square$

The next lemma establishes the coupling estimate for the t-in-neighbourhood $B_{v,t}^-$ and the tree $\mathcal{T}_{v,t}^-(\mathbf{d}^+)$. The estimate could be refined but (5.43) below will be more than sufficient for our purposes.

Lemma 23. *The random variables $B_{v,t}^-$ and the tree $\mathcal{T}_{v,t}^-(\mathbf{d}^+)$ can be coupled in such a way that for every $t \leq \frac{\log n}{4 \log \Delta}$, for all n large enough:*

$$(5.43) \qquad \mathbb{P}\left(B_{v,t}^- \neq \mathcal{T}_{v,t}^-(\mathbf{d}^+)\right) \leq \frac{\Delta^{3t}(\log n)^4}{n}.$$

Proof. Let $|\mathcal{T}_{v,t}^-|$ denote the number of edges in the tree $\mathcal{T}_{v,t}^- = \mathcal{T}_{v,t}^-(\mathbf{d}^+)$. Since at each iteration the number of edges added is stochastically dominated by a binomial random variable with parameters n and Δ/n, one has a large deviation bound for $|\mathcal{T}_{v,t}^-|$ of the form: there exist absolute constants $a, A > 0$ such that

$$(5.44) \qquad \mathbb{P}\left(|\mathcal{T}_{v,t}^-| > s\Delta^t\right) \leq A\,e^{-a\,s}, \qquad s \geq 1.$$

The estimate (5.44) can be proved e.g. by repeating the argument in [8, Lemma 23]. Next, observe that if $|\mathcal{T}_{v,t}^-| \leq s\Delta^t$ and $B_{v,t}^- \neq \mathcal{T}_{v,t}^-$, then there must exist $\ell = 1, \dots, s\Delta^t$ such that $G_v(\ell) \neq \mathcal{T}_v^-(\ell)$. The latter probability can be bounded via Lemma 22. Summarizing,

$$\mathbb{P}\left(B_{v,t}^- \neq \mathcal{T}_{v,t}^-(\mathbf{d}^+)\right) \leq \mathbb{P}\left(|\mathcal{T}_{v,t}^-| > s\Delta^t\right) + \mathbb{P}\left(B_{v,t}^- \neq \mathcal{T}_{v,t}^-(\mathbf{d}^+); |\mathcal{T}_{v,t}^-| \leq s\Delta^t\right)$$

$$(5.45) \qquad \leq A\,e^{-a\,s} + \sum_{\ell=1}^{s\Delta^t} \mathbb{P}(G_v(\ell) \neq \mathcal{T}_v^-(\ell)) \leq A\,e^{-a\,s} + \frac{s^3\Delta^{3t+2}}{n - s\Delta^t}.$$

The estimate (5.43) follows by taking $s = K \log n$ for some large enough constant K, and by taking n sufficiently large. $\qquad \square$

5.4. Proof of Lemma 5.

Recall that in both models DCM($\mathbf{d}^\pm$) and OCM($\mathbf{d}^+$) one has w.h.p. a unique stationary distribution for the simple random walk on G, which we denote π_0. The starting point is a result that follows directly from [6, 7], which allows us to replace the unknown distribution π_0 with a local approximation.

Proposition 24. *For any fixed $\varepsilon > 0$, taking $h = \varepsilon T_{\mathrm{ENT}}$, as $n \to \infty$ both models satisfy*

$$(5.46) \qquad \left\|\mu_{\mathrm{in}}P^h - \pi_0\right\|_{TV} \xrightarrow{\mathbb{P}} 0.$$

Proof. For a specific choice of $\varepsilon = \varepsilon_0$, this result appears in [6, Eq. (11)] for model 1 and [7, Eq. (12)] for model 2. In fact, the proofs in [6, 7] apply to any fixed $\varepsilon \in (0, \varepsilon_0)$ without modifications. Since $\left\|\mu_{\mathrm{in}}P^h - \pi_0\right\|_{TV}$ is monotone in h the statement (5.46) holds for all $\varepsilon > 0$. $\qquad \square$

To prove Lemma 5, by monotonicity of $\|\lambda P^t - \pi_0\|_{TV}$ as a function of t, we may restrict to sequences $t = t(n) \to \infty$ with $t = o(\log n)$. Thus, taking advantage of Proposition 24, the conclusion of Lemma 5 is a consequence of the following result.

Proposition 25. *There exists $\varepsilon > 0$ such that if $h = \varepsilon T_{\mathrm{ENT}}$, then for any $t = t(n) \to \infty$ with $t = o(\log n)$, for any widespread measure λ:*

$$(5.47) \qquad \left\|\lambda P^t - \mu_{\mathrm{in}}P^h\right\|_{TV} \xrightarrow{\mathbb{P}} 0.$$

Proof. The proof is based on a first moment argument. Indeed, it suffices to show that

$$(5.48) \qquad \lim_{n \to \infty} \mathbb{E}\left[\left\|\lambda P^t - \mu_{\mathrm{in}}P^h\right\|_{TV}\right] = 0.$$

Observe that

$$\mathbb{E}\left[\left\|\lambda P^t - \mu_{\mathrm{in}} P^h\right\|_{\mathrm{TV}}\right] = \frac{1}{2}\sum_{j\in[n]}\mathbb{E}\left[\left|\lambda P^t(j) - \mu_{\mathrm{in}} P^h(j)\right|\right]$$

$$(5.49) \qquad\qquad \leq \frac{1}{2}\mathbb{E}\left[\left|n\lambda P^t(\mathcal{I}) - n\mu_{\mathrm{in}} P^t(\mathcal{I})\right|\right] + \frac{1}{2}\mathbb{E}\left[\left|n\mu_{\mathrm{in}} P^t(\mathcal{I}) - n\mu_{\mathrm{in}} P^h(\mathcal{I})\right|\right],$$

where $\mathcal{I}$ denotes an independent uniformly random vertex in $[n]$ and the expectation $\mathbb{E}$ is understood to include the expectation over $\mathcal{I}$ as well. Consider the first term above. We are going to use Lemma 21 for model 1 and Lemma 23 for model 2. Notice that since these estimates apply to any fixed vertex v, they apply just as well if the vertex v is taken to be uniformly random in $[n]$, i.e. if $v = \mathcal{I}$ as it is the case here. In particular, since $t = o(\log n)$, as $n \to \infty$,

$$(5.50) \qquad\qquad \mathbb{P}\left(B_{\mathcal{I},t}^- \neq \mathcal{T}_t^-\right) \to 0,$$

where we use the unified notation $\mathcal{T}_t^-$ for the first t generations of the tree $\mathcal{T}_{\mathcal{I}}^-$ in either model 1 or model 2. Next, note that by definition, if $B_{\mathcal{I},t}^- = \mathcal{T}_t^-$, then

$$n\lambda P^t(\mathcal{I}) - n\mu_{\mathrm{in}} P^t(\mathcal{I}) = nX_t(\lambda) - M_t,$$

where we use the notation from (5.6) and (5.12). Therefore,

$$(5.51) \qquad \mathbb{E}\left[\left|n\lambda P^t(\mathcal{I}) - n\mu_{\mathrm{in}} P^t(\mathcal{I})\right|\right] \leq 2\mathbb{P}\left(B_{\mathcal{I},t}^- \neq \mathcal{T}_t^-\right) + \mathbb{E}\left[\left|M_t - nX_t(\lambda)\right|\right],$$

where we used the fact that

$$\mathbb{E}\left[\left|n\lambda P^t(\mathcal{I}) - n\mu_{\mathrm{in}} P^t(\mathcal{I})\right| \, \big| \, B_{\mathcal{I},t}^- \neq \mathcal{T}_t^-\right] \leq 2,$$

which follows from $\|\lambda P^t - \mu_{\mathrm{in}} P^t\|_{\mathrm{TV}} \leq 1$. Using Schwarz' inequality and Proposition 20 it follows that

$$\mathbb{E}\left[\left|M_t - nX_t(\lambda)\right|\right]^2 \leq \gamma(\lambda)\rho^t.$$

Since $t = t(n) \to \infty$ as $n \to \infty$ and $\rho \in (0,1)$, using (5.50) we conclude that

$$\lim_{n\to\infty} \mathbb{E}\left[\left|n\lambda P^t(\mathcal{I}) - n\mu_{\mathrm{in}} P^t(\mathcal{I})\right|\right] = 0,$$

for all widespread measure λ. This settles the convergence of the first term in (5.49). To handle the second term, reasoning as in (5.51) we obtain

$$\mathbb{E}\left[\left|n\mu_{\mathrm{in}} P^t(\mathcal{I}) - n\mu_{\mathrm{in}} P^h(\mathcal{I})\right|\right] \leq 2\mathbb{P}\left(B_{\mathcal{I},h}^- \neq \mathcal{T}_h^-\right) + \mathbb{E}\left[\left|M_t - M_h\right|\right]$$

If $h \leq \frac{\log n}{4\log\Delta}$, Lemma 21 and Lemma 23 imply that both models satisfy

$$(5.52) \qquad\qquad \mathbb{P}\left(B_{\mathcal{I},h}^- \neq \mathcal{T}_h^-\right) \to 0.$$

Moreover, Schwarz' inequality, Proposition 18 and standard facts about square integrable martingales (see, e.g., [22, Ch. 12]) imply

$$\mathbb{E}\left[\left|M_t - M_h\right|\right]^2 \leq \mathbb{E}[(M_t - M_h)^2] \leq \mathbb{E}[(M_t - M_\infty)^2] = C\rho^t.$$

Since the constant C is bounded, letting $n \to \infty$ concludes the proof.

$$\square$$

REFERENCES

[1] Louigi Addario-Berry, Borja Balle, and Guillem Perarnau. Diameter and stationary distribution of random r-out digraphs. *The Electronic Journal of Combinatorics*, pages P3–28, 2020.

[2] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using pagerank vectors. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 475–486. IEEE, 2006.

[3] Luca Avena, Hakan Güldas, Remco van der Hofstad, and Frank den Hollander. Random walks on dynamic configuration models: a trichotomy. *Stochastic Processes and their Applications*, 2018.

[4] Anna Ben-Hamou and Justin Salez. Cutoff for nonbacktracking random walks on sparse random graphs. *The Annals of Probability*, 45(3):1752–1770, 2017.

[5] Nathanael Berestycki, Eyal Lubetzky, Yuval Peres, and Allan Sly. Random walks on the random graph. *The Annals of Probability*, 46(1):456–490, 2018.

[6] Charles Bordenave, Pietro Caputo, and Justin Salez. Random walk on sparse random digraphs. *Probability Theory and Related Fields*, 170(3):933–960, Apr 2018.

[7] Charles Bordenave, Pietro Caputo, and Justin Salez. Cutoff at the "entropic time" for sparse markov chains. *Probability Theory and Related Fields*, 173(1):261–292, Feb 2019.

[8] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular ramanujan graphs. *The Annals of Probability*, 46(1):1–71, 01 2018.

[9] Marco Bressan and Enoch Peserico. Choose the damping, choose the ranking? *Journal of Discrete Algorithms*, 8(2):199 – 213, 2010. Selected papers from the 3rd Algorithms and Complexity in Durham Workshop ACiD 2007.

[10] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.

[11] Ningyuan Chen, Nelly Litvak, and Mariana Olvera-Cravioto. Generalized pagerank on directed configuration networks. *Random Structures & Algorithms*, 51(2):237–274, 2017.

[12] Ningyuan Chen and Mariana Olvera-Cravioto. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013.

[13] Colin Cooper and Alan Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combinatorics, Probability and Computing*, 13, 05 2002.

[14] Persi Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

[15] David Easley and Jon Kleinberg. *Networks, crowds, and markets*, volume 8. Cambridge university press Cambridge, 2010.

[16] Alessandro Garavaglia, Remco van der Hofstad, and Nelly Litvak. Local weak convergence for pagerank. *The Annals of Applied Probability*, 30(1):40–79, 2020.

[17] David F Gleich. Pagerank beyond the web. *SIAM Review*, 57(3):321–363, 2015.

[18] Glen Jeh and Jennifer Widom. Scaling personalized web search. In *Proceedings of the 12th international conference on World Wide Web*, pages 271–279. Acm, 2003.

[19] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2017. Second edition. With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[20] Eyal Lubetzky and Allan Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010.

[21] Daniel Vial and Vijay Subramanian. A structural result for personalized pagerank and its algorithmic consequences. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2):1–88, 2019.

[22] David Williams. *Probability with martingales*. Cambridge university press, 1991.

♯ DIPARTIMENTO DI MATEMATICA E FISICA, UNIVERSITÀ ROMA TRE, LARGO MURIALDO 1, 00146 ROMA, ITALY.
Email address: caputo@mat.uniroma3.it

♭ DIPARTIMENTO DI MATEMATICA E FISICA, UNIVERSITÀ ROMA TRE, LARGO MURIALDO 1, 00146 ROMA, ITALY.
Email address: matteo.quattropani@uniroma3.it