# STRONG COUPLINGS FOR STATIC LOCALLY TREE-LIKE RANDOM GRAPHS

MARIANA OLVERA-CRAVIOTO ⓘ,* *University of North Carolina at Chapel Hill*

### Abstract

We provide a general purpose result for the coupling of exploration processes of random graphs, both undirected and directed, with their local weak limits when this limit is a marked Galton–Watson process. This class includes in particular the configuration model and the family of inhomogeneous random graphs with rank-1 kernel. Vertices in the graph are allowed to have attributes on a general separable metric space and can potentially influence the construction of the graph itself. The coupling holds for any fixed depth of a breadth-first exploration process.

*Keywords:* Random graphs; complex networks; Galton–Watson processes; configuration model; inhomogeneous random graph; local-weak limits

2020 Mathematics Subject Classification: Primary 05C80
Secondary 60J80; 05C82

## 1. Introduction

There is a growing literature of problems in physics, mathematics, computer science, and operations research that are set up as processes, random or not, on large sparse graphs. The range of problems being studied is wide, and includes problems related to the classification, sorting, and ranking of large networks, as well as the analysis of Markov chains and interacting particle systems on graphs. Popular among the types of graphs used for these purposes are the locally tree-like random graph models such as the configuration model and the inhomogeneous random graph family (which includes the classical Erdős–Rényi model). These random graph models are quite versatile in the types of graphs they can mimic, and have important mathematical properties that make their analysis tractable.

In particular, the mathematical tractability of locally tree-like random graphs comes from the fact that their local neighborhoods resemble trees. This property makes it easy to transfer questions about the process of interest on a graph to the often easier analysis of the process on the limiting tree. Mathematically, this transfer is enabled by the notion of local weak convergence [1, 2, 3, 15]. However, as is the case for many problems involving usual weak convergence of random variables, it is often desirable to construct the original set of random variables and their corresponding weak limits on the same probability space; in other words, to have a coupling. In addition, many problems studying processes on graphs require that we keep track of additional vertex attributes not usually included in the local weak limits, attributes that

may not be discrete. The recent work in [14] gives several examples of Markov chains and systems of equations on directed graphs whose analysis relies on the kind of couplings presented here, and include the study of the personalized PageRank distribution [9, 15, 20], among others. Further applications include the analysis of interacting diffusions [16, 17], where we may wish to allow the vertex attributes to influence the dynamics of the processes being studied. The results in this paper were designed to solve these two problems simultaneously by providing a general-purpose coupling between the exploration of the neighborhood of a uniformly chosen vertex in a locally tree-like graph and its local weak limit, including general vertex attributes that may indirectly influence the construction of the graph.

The main results focus only on the two families of random graph models that are known to converge, in the local weak sense, to a marked Galton–Watson process. It is worth mentioning that other locally tree-like graphs like the preferential attachment models do not fall into this category, since their local weak limits are continuous-time branching processes. In particular, we focus on random graphs constructed according to either a configuration model or any of the inhomogeneous random graph models with rank-1 kernels (see Sections 1.1 and 1.2 for the precise definitions). Our results include both undirected and directed graphs, and are given under minimal moment conditions in order to include scale-free graphs. In particular, under our assumptions, it is possible for the offspring distribution in the limiting marked Galton–Watson process to have infinite mean, and in the directed case for the limiting joint distribution of the in-degree and out-degree of a vertex to have infinite covariance.

Before describing the two families of random graph models for which our coupling theorems hold, we introduce some definitions to be used throughout the paper. We use $G(V_n, E_n)$ to denote a graph, or multigraph, having vertices $V_n = \{1, 2, \ldots, n\}$ and edges in the set $E_n$. A directed edge from vertex $i$ to vertex $j$ is denoted by $(i, j)$. For multigraphs, we also need to keep track of the multiplicity of each edge or self-loop, so we use $l(i)$ to denote the number of self-loops of vertex $i$, and $e(i, j)$ to denote the number of edges from vertex $i$ to vertex $j$. If the graph is undirected, we simply ignore the direction. In the undirected case, we use $D_i$ to denote the degree of vertex $i$, which corresponds to the number of adjacent neighbors of vertex $i$. In the directed case, we use $D_i^-$ to denote the in-degree of vertex $i$ and $D_i^+$ to denote its out-degree; the in-degree counts the number of inbound neighbors, and the out-degree the number of outbound ones. All our results are given in terms of the large-graph limit, which corresponds to taking a sequence of graphs $\{G(V_n, E_n) : n \geq 1\}$ and taking the limit as $|V_n| = n \to \infty$, where $|A|$ denotes the cardinality of set $A$. Both the configuration model and the family of inhomogeneous random graphs are meant to model large *static* graphs, since there may be no relation between $G(V_n, E_n)$ and $G(V_m, E_m)$ for $n \geq m$. Strong couplings for *evolving* graphs such as the preferential attachment models are a topic for future work.

## 1.1. Configuration model

The configuration model [4, 21] produces graphs from any prescribed (graphical) degree sequence. In the undirected version of this model, each vertex is assigned a number of stubs or half-edges equal to its target degree. Then, these half-edges are randomly paired to create edges in the graph.

For an undirected configuration model (CM), we assume that each vertex $i \in V_n$ is assigned an *attribute* vector $\mathbf{a}_i = (D_i, \mathbf{b}_i)$, where $D_i \in \mathbb{N}$ is its degree, and $\mathbf{b}_i$ encodes additional information about vertex $i$ that does not directly affect the construction of the graph but may depend on $D_i$. The attributes $\{\mathbf{b}_i\}$ are assumed to take values on a separable metric space $\mathcal{S}'$. For the sequence $\{D_i : 1 \leq i \leq n\}$ to define the degree sequence of an undirected graph, we must

have that $L_n := \sum_{i=1}^{n} D_i$ is even. Note that this may require us to consider a double sequence $\{\mathbf{a}_i^{(n)} : i \geq 1, n \geq 1\}$ rather than a unique sequence, i.e. one where $\mathbf{a}_i^{(n)} \neq \mathbf{a}_i^{(m)}$ for $n \neq m$. In applications it is often convenient to allow the vertex attributes to be random themselves; [14] studies various stochastic recursions on graphs that include vertex attributes such as the ones we envision here.

Assuming that $L_n$ is even, enumerate all the stubs and pick one stub to pair; suppose that stub belongs to vertex $i$. Next, choose one of the remaining $L_n - 1$ stubs uniformly at random, and if the stub belongs to vertex $j$, draw an edge between vertices $i$ and $j$. Then, pick another stub to pair. In general, a stub being paired chooses uniformly at random from the set of unpaired stubs, then identifies the vertex to which the chosen stub belongs, and creates an edge between its vertex and the one to which the chosen stub belongs.

The directed version of the configuration model (DCM) is such that each vertex $i \in V_n$ is assigned an *attribute* of the form $\mathbf{a}_i = (D_i^-, D_i^+, \mathbf{b}_i) \in \mathbb{N}^2 \times \mathcal{S}'$. Similarly to the undirected case, $D_i^-$ and $D_i^+$ denote the in-degree and the out-degree, respectively, of vertex $i$, and $\mathbf{b}_i$ is allowed to depend on $(D_i^-, D_i^+)$. The condition needed to ensure we can draw a graph is now $L_n := \sum_{i=1}^{n} D_i^+ = \sum_{i=1}^{n} D_i^-$, which again may require us to consider a double sequence $\{\mathbf{a}_i^{(n)} : i \geq 1, n \geq 1\}$.

As for the CM, we give to each vertex $i$ a number $D_i^-$ of inbound stubs, and a number $D_i^+$ of outbound stubs. To construct the graph, we start by choosing an inbound (outbound) stub, say belonging to vertex $i$, and choose uniformly at random one of the $L_n$ outbound (inbound) stubs. If the chosen stub belongs to vertex $j$, draw an edge from $j$ to $i$ (from $i$ to $j$); then pick another inbound (outbound) stub to pair. In general, when pairing an inbound (outbound) stub, we pick uniformly at random from all the remaining unpaired outbound (inbound) stubs. If the stub being paired belongs to vertex $i$, and the one to which the chosen stub belongs to is $j$, we draw a directed edge from $j$ to $i$ (from $i$ to $j$).

We emphasize that both the CM and the DCM are in general multi-graphs, that is, they can have self-loops and multiple edges (in the same direction) between a given pair of vertices. However, provided the pairing process does not create self-loops or multiple edges, the resulting graph is uniformly chosen among all graphs having the prescribed degree sequence. It is well known that when the empirical degree distribution converges weakly and its second moment converges to that of the limit, the pairing process results in a simple graph with a probability that remains bounded away from zero even as the graph grows [10, 21].

We use $\mathscr{F}_n = \sigma(\mathbf{a}_i : 1 \leq i \leq n)$ to denote the sigma algebra generated by the attribute sequence, which does not include the edge structure of the graph. To simplify the notation, we use $\mathbb{P}_n(\cdot) = \mathbb{P}(\cdot \mid \mathscr{F}_n)$ and $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid \mathscr{F}_n]$ to denote the conditional probability and conditional expectation, respectively, given $\mathscr{F}_n$.

## 1.2. Inhomogeneous random graphs

The second class of random graph models we consider is the family of inhomogeneous random graphs (digraphs), in which the presence of an edge is determined by the toss of a coin, independently of any other edge. This family includes the classical Erdős–Rényi graph [13], but also several generalizations that allow the edge probabilities to depend on the two vertices being connected, e.g. the Chung–Lu model [11], the Norros–Reittu model (or Poissonian random graph) [19], and the generalized random graph [7], to name a few. Unlike the Erdős–Rényi model, these generalizations are capable of producing graphs with inhomogeneous degree sequences, and can mimic almost any degree distribution whose support is $\mathbb{N}$ (or $\mathbb{N}^2$ in the

directed case). This paper focuses only on inhomogeneous random graphs (digraphs) having rank-1 kernels (see [6]), which excludes models such as the stochastic block model.

Collectively, this family of models has a long history in the random graph literature, and their connectivity properties, phase transitions, and degree distributions are well known. Rather than attempting to name all the existing references where these models have appeared, we refer the interested reader to the books [5, 12, 21, 22], where many of their properties have been compiled.

To define an undirected inhomogeneous random graph (IR), assign to each vertex $i \in V_n$ an *attribute* $\mathbf{a}_i = (W_i, \mathbf{b}_i) \in \mathbb{R}_+ \times \mathcal{S}'$. The $W_i$ will be used to determine how likely vertex $i$ is to have neighbors, while the $\mathbf{b}_i$ can be used to include vertex characteristics that are not needed for the construction of the graph but that are allowed to depend on $W_i$. If convenient, one can consider using a double sequence $\{\mathbf{a}_i^{(n)} : i \geq 1, n \geq 1\}$ as with the configuration model, but this is not as important since the sequence $\mathscr{W}_n := \{W_i : 1 \leq i \leq n\}$ does not need to satisfy any additional conditions in order for us to draw the graph. As with the CM (DCM), the vertex attributes are allowed to be random.

We use the notation $\mathscr{F}_n = \sigma(\mathbf{a}_i : 1 \leq i \leq n)$, as for the configuration model, to denote the sigma algebra generated by the vertex attributes, as well as the notation for the corresponding conditional probability, $\mathbb{P}_n(\cdot) = \mathbb{P}(\cdot \mid \mathscr{F}_n)$, and expectation, $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid \mathscr{F}_n]$.

For the IR, the edge probabilities are given by

$$p_{ij}^{(n)} := \mathbb{P}_n((i, j) \in E_n) = 1 \wedge \frac{W_i W_j}{\theta n}(1 + \varphi_n(W_i, W_j)), \qquad 1 \leq i < j \leq n,$$

where $-1 < \varphi_n(W_i, W_j) = \varphi(n, W_i, W_j, \mathscr{W}_n)$ almost surely (a.s.) is a function that may depend on the entire sequence $\mathscr{W}_n$, on the types of the vertices $\{i, j\}$, or exclusively on $n$, and $0 < \theta < \infty$ satisfies $\frac{1}{n}\sum_{i=1}^{n} W_i \overset{\mathrm{P}}{\longrightarrow} \theta$, $n \to \infty$. Throughout this paper, $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. Since the graph is to be simple by construction, $p_{ii}^{(n)} \equiv 0$ for all $i \in V_n$.

For the directed version, which we refer to as an inhomogeneous random digraph (IRD), the vertex attributes take the form $\mathbf{a}_i = (W_i^-, W_i^+, \mathbf{b}_i) \in \mathbb{R}_+^2 \times \mathcal{S}'$. The parameter $W_i^-$ controls the in-degree of vertex $i$, and $W_i^+$ its out-degree. If we write $\mathbf{W}_i = (W_i^-, W_i^+)$, the edge probabilities in the IRD are given by

$$p_{ij}^{(n)} := \mathbb{P}_n((i, j) \in E_n) = 1 \wedge \frac{W_i^+ W_j^-}{\theta n}(1 + \varphi_n(\mathbf{W}_i, \mathbf{W}_j)), \qquad 1 \leq i \neq j \leq n,$$

where $-1 < \varphi_n(\mathbf{W}_i, \mathbf{W}_j) = \varphi(n, \mathbf{W}_i, \mathbf{W}_j, \mathscr{W}_n)$ a.s. is a function that may depend on the entire sequence $\mathscr{W}_n := \{\mathbf{W}_i : 1 \leq i \leq n\}$, on the types of the vertices $\{i, j\}$, or exclusively on $n$, and $0 < \theta < \infty$ satisfies $\frac{1}{n}\sum_{i=1}^{n}(W_i^- + W_i^+) \overset{\mathrm{P}}{\longrightarrow} \theta$, $n \to \infty$. Since the graphs are again simple by construction, we have $p_{ii}^{(n)} \equiv 0$ for all $i \in V_n$.

## 2. Main result for undirected graphs

For an undirected graph constructed according to one of the two models (CM or IR), our main result shows that there exists a coupling between the breadth-first exploration of the component of a uniformly chosen vertex and that of the root node of a marked Galton–Watson process. Before we can state the theorem, we need to introduce some notation on the graph and describe the Galton–Watson process that describes its local weak limit.

Each vertex $i$ in an undirected graph (multigraph) $G(V_n, E_n)$ is given a vertex *attribute* of the form

$$\mathbf{a}_i = \begin{cases} (D_i, \mathbf{b}_i) & \text{if } G(V_n, E_n) \text{ is a CM,} \\ (W_i, \mathbf{b}_i) & \text{if } G(V_n, E_n) \text{ is an IR.} \end{cases}$$

In addition, define for each vertex $i$ its *full mark* $\mathbf{X}_i = (D_i, \mathbf{a}_i)$, where $D_i$ is the degree of vertex $i$. We point out that the definition of $\mathbf{X}_i$ is redundant when the graph is a CM; however, this is not so if the graph is an IR. In both cases the vertex attributes are measurable with respect to $\mathscr{F}_n$, while the full marks are not if the graph is an IR.

The main assumption needed for the coupling to hold is given in terms of the empirical measure for the vertex attributes, i.e.

$$\nu_n(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\mathbf{a}_i \in \cdot). \tag{2.1}$$

In order to state the assumption, recall that the state space for the vertex attributes, $\mathcal{S}'$, is assumed to be a separable metric space under metric $\rho'$. Now define the metric $\rho(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_2 - y_2| + \rho'(\mathbf{x}_3, \mathbf{y}_3)$, $\mathbf{x} = (x_1, x_2, \mathbf{x}_3)$, $\mathbf{y} = (y_1, y_2, \mathbf{y}_3)$, on the space $\mathcal{S} := \mathbb{N} \times \mathbb{R} \times \mathcal{S}'$, which makes $\mathcal{S}$ a separable metric space as well. Using $\rho$, and for any probability measures $\nu_n$, $\mu_n$ on the conditional probability space $(\mathcal{S}, \mathscr{F}_n, \mathbb{P}_n)$, define the Wasserstein metric of order one: $W_1(\nu_n, \mu_n) = \inf\{\mathbb{E}_n[\rho(\hat{\mathbf{Y}}, Y)] : \text{law}(\hat{\mathbf{Y}} \mid \mathscr{F}_n) = \nu_n, \text{law}(Y \mid \mathscr{F}_n) = \mu_n\}$.

**Assumption 2.1.** (Undirected.) *Let $\nu_n$ be defined according to* (2.1)*, and suppose there exists a non-random probability measure $\nu$ such that $W_1(\nu_n, \nu) \xrightarrow{\mathbb{P}} 0$, $n \to \infty$. In addition, assume that the following conditions hold:*

A. *In the CM, let $(\mathscr{D}, \mathbf{B})$ be distributed according to $\nu$, and suppose there exists a non-random $\mathbf{b}_0 \in \mathcal{S}'$ such that $\mathbb{E}[\mathscr{D} + \rho'(\mathbf{B}, \mathbf{b}_0)] < \infty$.*

B. *In the IR, let $(W, \mathbf{B})$ be distributed according to $\nu$, and suppose the following hold:*

  (i) $\mathcal{E}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{1 \leq i \neq j \leq n} \left| p_{ij}^{(n)} - (r_{ij}^{(n)} \wedge 1) \right| \xrightarrow{\mathrm{P}} 0$ *as* $n \to \infty$, *where* $r_{ij}^{(n)} = W_i W_j / (\theta n)$.

  (ii) *There exists a non-random $\mathbf{b}_0 \in \mathcal{S}'$ such that $\mathbb{E}[W + \rho'(\mathbf{B}, \mathbf{b}_0)] < \infty$.*

Now that we have stated the assumptions for our theorem, we need to describe the local neighborhood of a vertex in the graph $G(V_n, E_n)$. To do this, let $I \in V_n$ denote a uniformly chosen vertex in $G(V_n, E_n)$; vertices are identified with their labels in $\{1, 2, \ldots, n\}$. Define $A_0 = \{I\}$, and let $A_k$ denote the set of vertices at hop distance $k$ from $I$. Now write $\mathcal{G}_I^{(k)}$ to be the subgraph of $G(V_n, E_n)$ consisting of the vertices in $\bigcup_{r=0}^{k} A_r$ along with their (multiple) edges and self-loops. We also use the notation $\mathcal{G}_I^{(k)}(\mathbf{a})$ to refer to the graph $\mathcal{G}_I^{(k)}$ including all the attributes of its vertices.

**Definition 2.1.** We say that two simple graphs $G(V, E)$ and $G'(V', E')$ are *isomorphic* if there exists a bijection $\sigma : V \to V'$ such that edge $(i, j) \in E$ if and only if edge $(\sigma(i), \sigma(j)) \in E'$. We say that two multigraphs $G(V, E)$ and $G'(V', E')$ are *isomorphic* if there exists a bijection $\sigma : V \to V'$ such that $l(i) = l(\sigma(i))$ and $e(i, j) = e(\sigma(i), \sigma(j))$ for all $i \in V$ and all $(i, j) \in E$, where $l(i)$ is the number of self-loops of vertex $i$, and $e(i, j)$ is the number of edges from vertex $i$ to vertex $j$. In both cases, we write $G \simeq G'$.

To describe the limit of $\mathcal{G}_I^{(k)}$ as $n \to \infty$, we construct a delayed marked Galton–Watson process, denoted $\mathcal{T}(A)$, using the measure $\nu$ in Assumption 2.1. Here, 'delayed' refers to the fact that the root will, in general, have a different distribution than all other nodes in the tree.

To start, let $\mathcal{U} := \bigcup_{k=0}^{\infty} \mathbb{N}_+^k$ denote the set of labels for nodes in a tree, with the convention that $\mathbb{N}_+^0 := \{\emptyset\}$ contains the root. For a label $\mathbf{i} = (i_1, \ldots, i_k)$ we write $|\mathbf{i}| = k$ to denote its length, and use $(\mathbf{i}, j) = (i_1, \ldots, i_k, j)$ to denote the index concatenation operation.

The tree $\mathcal{T}$ is constructed as follows. Let $\{(\mathcal{N}_\mathbf{i}, A_\mathbf{i}) : \mathbf{i} \in \mathcal{U}\}$ denote a sequence of independent vectors in $\mathcal{S}$, with $\{(\mathcal{N}_\mathbf{i}, A_\mathbf{i}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ independent and identically distributed (i.i.d.). For any $\mathbf{i} \in \mathcal{U}$, $\mathcal{N}_\mathbf{i}$ denotes the number of offspring of node $\mathbf{i}$, and $A_\mathbf{i}$ denotes its attribute (mark). As with the graph, we use the notation $\mathcal{T}$ to denote the tree without its attributes. Let $\mathcal{A}_0 = \{\emptyset\}$ and recursively define $\mathcal{A}_k = \{(\mathbf{i}, j) : \mathbf{i} \in \mathcal{A}_{k-1}, 1 \le j \le \mathcal{N}_\mathbf{i}\}$, $k \ge 1$, to be the $k$th generation of $\mathcal{T}$. To match the notation on the graph, we write $X_\emptyset = (\mathcal{N}_\emptyset, A_\emptyset)$ and $X_\mathbf{i} = (\mathcal{N}_\mathbf{i} + 1, A_\mathbf{i})$, $\mathbf{i} \neq \emptyset$. The marked tree is then given by $\mathcal{T}(A) = \{X_\mathbf{i} : \mathbf{i} \in \mathcal{T}\}$; note that the marks include the number of offspring of each node, from which the edges in the tree can be deduced. We denote by $\mathcal{T}^{(k)}$ $(\mathcal{T}^{(k)}(A))$ the restriction of $\mathcal{T}$ $(\mathcal{T}(A))$ to its first $k$ generations.

It only remains to identify the distribution of $X_\mathbf{i}$, for both $\mathbf{i} = \emptyset$ and $\mathbf{i} \neq \emptyset$, in terms of the probability measure $\nu$ in Assumption 2.1. For a CM, let $A = (\mathscr{D}, B)$ be distributed according to $\nu$; then, $\mathbb{P}(X_\emptyset \in \cdot) = \mathbb{P}((\mathscr{D}, A) \in \cdot)$, and $\mathbb{P}(X_\mathbf{i} \in \cdot) = (1/\mathbb{E}[\mathscr{D}])\mathbb{E}[\mathscr{D}\mathbf{1}((\mathscr{D}, A) \in \cdot)]$, $\mathbf{i} \neq \emptyset$. For an IR, let $A = (W, B)$ be distributed according to $\nu$; then, $\mathbb{P}(X_\emptyset \in \cdot) = \mathbb{P}((D, A) \in \cdot)$, and $\mathbb{P}(X_\mathbf{i} \in \cdot) = (1/\mathbb{E}[W])\mathbb{E}[W\mathbf{1}((D + 1, A) \in \cdot)]$, $\mathbf{i} \neq \emptyset$, where $D$ is a mixed Poisson random variable with mean $W$. Note that the distribution of $X_\mathbf{i}$ for $\mathbf{i} \neq \emptyset$ corresponds to a *size-biased* version of the distribution of $X_\emptyset$ with respect to its first coordinate.

We are now ready to state the main coupling theorem for undirected graphs.

**Theorem 2.1.** *Suppose $G(V_n, E_n)$ is either a CM or an IR satisfying Assumption 2.1. Then, for any fixed $k$ and $\mathcal{G}_I^{(k)}(\mathbf{a})$ the depth-$k$ neighborhood of a uniformly chosen vertex $I \in V_n$, there exists a marked Galton–Watson tree $\mathcal{T}^{(k)}(A)$ restricted to its first $k$ generations whose root corresponds to vertex $I$ and which is such that $\mathbb{P}_n(\mathcal{G}_I^{(k)} \not\simeq \mathcal{T}^{(k)}) \xrightarrow{\mathrm{P}} 0$, $n \to \infty$; if we let $\sigma(\mathbf{i}) \in V_n$ denote the vertex in the graph corresponding to node $\mathbf{i} \in \mathcal{T}^{(k)}$, and define for any $\varepsilon > 0$ the event $C_I^{(k,\varepsilon)} = \{\bigcap_{\mathbf{i} \in \mathcal{T}^{(k)}} \{\rho(\mathbf{X}_{\sigma(\mathbf{i})}, X_\mathbf{i}) \le \varepsilon\}, \mathcal{G}_I^{(k)} \simeq \mathcal{T}^{(k)}\}$, then $\mathbb{E}_n[\rho(\mathbf{X}_I, X_\emptyset)] \xrightarrow{\mathrm{P}} 0$ and $\mathbb{P}_n(C_I^{(k,\varepsilon)}) \xrightarrow{\mathrm{P}} 1$, $n \to \infty$. Moreover, for any fixed $m, k \ge 1$, $\varepsilon > 0$, and $\{I_j : 1 \le j \le m\}$ i.i.d. random variables uniformly chosen in $V_n$, there exist i.i.d. copies of $\mathcal{T}^{(k)}(A)$, denoted $\{\mathcal{T}_{\emptyset(I_j)}^{(k)}(A) : 1 \le j \le m\}$, whose roots correspond to the vertices $\{I_j : 1 \le j \le m\}$ in $G(V_n, E_n)$, such that $\sum_{j=1}^m \mathbb{E}_n[\rho(\mathbf{X}_{I_j}, X_{\emptyset(I_j)})] \xrightarrow{\mathrm{P}} 0$ and $\mathbb{P}_n(\bigcap_{j=1}^m C_{I_j}^{(k,\varepsilon)}) \xrightarrow{\mathrm{P}} 1$, $n \to \infty$.*

**Remark 2.1.** Theorem 2.1 implies that the graph $G(V_n, E_n)$ converges in the local weak sense in probability, as introduced in [1, 2, 3] for undirected graphs and later extended in [15] to marked undirected graphs. The statement involving more than one exploration is related to the notion of propagation of chaos in the interacting particles literature (see, for example, [8, Definition 4.1]), and can be of independent interest in that context.

## 3. Main result for directed graphs

In the directed case, our main result will allow us to couple the breadth-first exploration of either the in-component or the out-component of a uniformly chosen vertex. Since the two cases are clearly symmetric, we state our results only for the in-component.

As with the undirected graph, each vertex $i$ in the graph $G(V_n, E_n)$ has an *attribute*,

$$\mathbf{a}_i = \begin{cases} (D_i^-, D_i^+, \mathbf{b}_i) & \text{if } G(V_n, E_n) \text{ is a DCM,} \\ (W_i^-, W_i^+, \mathbf{b}_i) & \text{if } G(V_n, E_n) \text{ is an IRD.} \end{cases}$$

The *full mark* of vertex $i$ is now given by $\mathbf{X}_i = (D_i^-, D_i^+, \mathbf{a}_i)$, where $D_i^-$ and $D_i^+$ are the in-degree and out-degree, respectively, of vertex $i$.

With some abuse of notation, we again use $\nu_n$, as defined in (2.1), to denote the empirical measure for the vertex attributes. However, the state space for the full marks is now $\mathcal{S} := \mathbb{N}^2 \times \mathbb{R}^2 \times \mathcal{S}'$, equipped with the metric $\rho(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3| + |x_4 - y_4| + \rho'(\mathbf{x}_5, \mathbf{y}_5)$ for $\mathbf{x} = (x_1, x_2, x_3, x_4, \mathbf{x}_5)$ and $\mathbf{y} = (y_1, y_2, y_3, y_4, \mathbf{y}_5)$. The Wasserstein metric $W_1$ defined on the conditional probability space $(\mathcal{S}, \mathcal{F}_n, \mathbb{P}_n)$ remains the same after the adjustments made to $\mathcal{S}$ and $\rho$.

**Assumption 3.1.** (Directed.) *Let $\nu_n$ be defined according to (2.1), and suppose there exists a non-random probability measure $\nu$ such that $W_1(\nu_n, \nu) \xrightarrow{\mathrm{P}} 0$, $n \to \infty$. In addition, assume that the following conditions hold:*

A. *In the DCM, let $(\mathscr{D}^-, \mathscr{D}^+, \boldsymbol{B})$ be distributed according to $\nu$, and suppose there exists a non-random $\mathbf{b}_0 \in \mathcal{S}'$ such that $\mathbb{E}[\mathscr{D}^- + \mathscr{D}^+ + \rho(\boldsymbol{B}, \mathbf{b}_0)] < \infty$.*

B. *In the IRD, let $(W^-, W^+, \boldsymbol{B})$ be distributed according to $\nu$, and suppose the following hold:*

   (i) *$\mathcal{E}_n = \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq i \neq j \leq n} \left| p_{ij}^{(n)} - (r_{ij}^{(n)} \wedge 1) \right| \xrightarrow{\mathrm{P}} 0$  as  $n \to \infty$,  where  $r_{ij}^{(n)} = W_i^+ W_j^- / (\theta n)$.*

   (ii) *There exists a non-random $\mathbf{b}_0 \in \mathcal{S}'$ such that $\mathbb{E}[W^- + W^+ + \rho(\boldsymbol{B}, \mathbf{b}_0)] < \infty$.*

Since we will state our result for the exploration of the in-component of a uniformly chosen vertex, the structure of the coupled tree will be determined by the vertices that we encounter during a breadth-first exploration. This exploration starts with a uniformly chosen vertex $I \in V_n$, which is used to create the set $A_0 = \{I\}$. We then explore each of the inbound edges of $I$ to discover all the vertices at inbound distance one from $I$, which become the set $A_1$. In general, to identify the vertices in the set $A_k$, we explore all the inbound edges of vertices in $A_{k-1}$. As we perform the exploration, we also discover the out-degrees of the vertices we have encountered; however, we do not follow any outbound edges. We then define $\mathcal{G}_I^{(k)}$ to be the subgraph of $G(V_n, E_n)$ whose vertex set is $\bigcup_{r=0}^k A_r$ and whose edges are those that are encountered during the breadth-first exploration we described. The notation $\mathcal{G}_I^{(k)}(\mathbf{a})$ is used to refer to the graph $\mathcal{G}_I^{(k)}$ including the values of the full marks $\{\mathbf{X}_i\}$ for all of its vertices.

In the directed case, the limit of $\mathcal{G}_I^{(k)}$ is again a delayed marked Galton–Watson process, with the convention that all its edges are pointing towards the root. We denote the tree $\mathcal{T}(\boldsymbol{A})$ as before; however, it will be constructed using a sequence of independent vectors of the form $\{(\mathcal{N}_{\mathbf{i}}, \mathcal{D}_{\mathbf{i}}, \boldsymbol{A}_{\mathbf{i}}) : \mathbf{i} \in \mathcal{U}\}$, with $\{(\mathcal{N}_{\mathbf{i}}, \mathcal{D}_{\mathbf{i}}, \boldsymbol{A}_{\mathbf{i}}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ i.i.d. In other words, the full marks now take the form $\mathbf{X}_{\mathbf{i}} = (\mathcal{N}_{\mathbf{i}}, \mathcal{D}_{\mathbf{i}}, \boldsymbol{A}_{\mathbf{i}})$, $\mathbf{i} \in \mathcal{U}$. The construction of the tree $\mathcal{T}$ is done as in the undirected case using the $\{\mathcal{N}_{\mathbf{i}} : \mathbf{i} \in \mathcal{U}\}$, and the marked tree is given by $\mathcal{T}(\boldsymbol{A}) = \{\mathbf{X}_{\mathbf{i}} : \mathbf{i} \in \mathcal{T}\}$. The notation $\mathcal{T}^{(k)}(\mathcal{T}^{(k)}(\boldsymbol{A}))$ again refers to the restriction of $\mathcal{T}(\mathcal{T}(\boldsymbol{A}))$ to its first $k$ generations.

The distributions of the full marks $\mathbf{X_i}$ for both $\mathbf{i} = \emptyset$ and $\mathbf{i} \neq \emptyset$ are also different than in the undirected case. For a DCM, let $A = (\mathscr{D}^-, \mathscr{D}^+, \mathbf{B})$ be distributed according to $\nu$; then, $\mathbb{P}(X_\emptyset \in \cdot) = \mathbb{P}((\mathscr{D}^-, \mathscr{D}^+, A) \in \cdot)$, and $\mathbb{P}(X_\mathbf{i} \in \cdot) = (1/\mathbb{E}[\mathscr{D}^+])\mathbb{E}[\mathscr{D}^+ \mathbf{1}((\mathscr{D}^-, \mathscr{D}^+, A) \in \cdot)]$, $\mathbf{i} \neq \emptyset$. For an IRD, let $A = (W^-, W^+, \mathbf{B})$ be distributed according to $\nu$; then, $\mathbb{P}(X_\emptyset \in \cdot) = \mathbb{P}((D^-, D^+, A) \in \cdot)$, and $\mathbb{P}(X_\mathbf{i} \in \cdot) = (1/\mathbb{E}[W^+])\mathbb{E}[W^+ \mathbf{1}((D^-, D^+ + 1, A) \in \cdot)]$, $\mathbf{i} \neq \emptyset$, where $D^-$ and $D^+$ are conditionally independent (given $(W^-, W^+)$) Poisson random variables with means $cW^-$ and $(1-c)W^+$, respectively, and $c = \mathbb{E}[W^+]/\mathbb{E}[W^- + W^+]$. Note that in this case, the distribution of $X_\mathbf{i}$ for $\mathbf{i} \neq \emptyset$ corresponds to a *size-biased* version of the distribution of $X_\emptyset$ with respect to its second coordinate.

The following is our main coupling theorem for directed graphs.

**Theorem 3.1.** *Suppose $G(V_n, E_n)$ is either a DCM or an IRD satisfying Assumption* 3.1. *Then, for any fixed $k$ and $\mathcal{G}_I^{(k)}(\mathbf{a})$ the depth-$k$ neighborhood of a uniformly chosen vertex $I \in V_n$, there exists a marked Galton–Watson tree $\mathcal{T}^{(k)}(A)$ restricted to its first $k$ generations, whose root corresponds to vertex $I$, and such that $\mathbb{P}_n(\mathcal{G}_I^{(k)} \not\simeq \mathcal{T}^{(k)}) \xrightarrow{\mathrm{P}} 0$, $n \to \infty$, and if we let $\sigma(\mathbf{i}) \in V_n$ denote the vertex in the graph corresponding to node $\mathbf{i} \in \mathcal{T}^{(k)}$, and define for any $\varepsilon > 0$ the event $C_I^{(k,\varepsilon)} = \big\{ \bigcap_{\mathbf{i} \in \mathcal{T}^{(k)}} \{\rho(\mathbf{X}_{\sigma(\mathbf{i})}, \mathbf{X_i}) \leq \varepsilon\}, \mathcal{G}_I^{(k)} \simeq \mathcal{T}^{(k)} \big\}$, then $\mathbb{E}_n[\rho(\mathbf{X}_I, X_\emptyset)] \xrightarrow{\mathrm{P}} 0$ and $\mathbb{P}_n(C_I^{(k,\varepsilon)}) \xrightarrow{\mathrm{P}} 1$, $n \to \infty$. Moreover, for any fixed $m, k \geq 1$, $\varepsilon > 0$, and $\{I_j : 1 \leq j \leq m\}$ i.i.d. random variables uniformly chosen in $V_n$, there exist i.i.d. copies of $\mathcal{T}^{(k)}(A)$, denoted $\{\mathcal{T}_{\emptyset(I_j)}^{(k)}(A) : 1 \leq j \leq m\}$, whose roots correspond to the vertices $\{I_j : 1 \leq j \leq m\}$ in $G(V_n, E_n)$, such that $\sum_{j=1}^m \mathbb{E}_n[\rho(\mathbf{X}_{I_j}, X_{\emptyset(I_j)})] \xrightarrow{\mathrm{P}} 0$ and $\mathbb{P}_n\big(\bigcap_{j=1}^m C_{I_j}^{(k,\varepsilon)}\big) \xrightarrow{\mathrm{P}} 1$, $n \to \infty$.*

The remainder of the paper contains the proofs of Theorems 2.1 and 3.1.

## 4. Proofs

The proofs of Theorems 2.1 and 3.1 are based on an intermediate coupling between the breadth-first exploration of the graph $G(V_n, E_n)$ and a delayed marked Galton–Watson process whose offspring distribution and marks still depend on the sigma algebra $\mathscr{F}_n$. This intermediate step consists in coupling $\mathcal{G}_I^{(k)}$ with a marked tree denoted $\hat{T}^{(k)}(\hat{\mathbf{A}})$. Interestingly, this coupling will be perfect, in the sense that the vertex/node marks in each of the two graphs will also be identical to each other. The proofs of Theorem 2.1 and 3.1 will be complete once we show that $\hat{T}^{(k)}(\hat{\mathbf{A}})$ can be coupled with the limiting $\mathcal{T}^{(k)}(A)$.

To organize the exposition, we will separate the undirected case from the directed one. Most of the proofs for the directed case have been established in prior work by the author, and therefore will be omitted; the precise references and the missing details are given in Section 4.2. Once the intermediate coupling theorems are proved, the coupling between the two trees can be done similarly for the undirected and directed cases (on the trees, the direction of the edges is irrelevant).

### 4.1. Discrete coupling for undirected graphs

As mentioned above, the main difference between the intermediate tree and the limiting one lies in the distribution of the marks. As before, we start with the construction of the possibly infinite tree $\hat{T}$, which is done with the conditionally independent (given $\mathscr{F}_n$) sequence of random vectors in $\mathcal{S}$, $\{(\hat{N}_\mathbf{i}, \hat{\mathbf{A}}_\mathbf{i}) : \mathbf{i} \in \mathcal{U}\}$, with $\{(\hat{N}_\mathbf{i}, \hat{\mathbf{A}}_\mathbf{i}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ conditionally i.i.d. Let

$\hat{A}_0 = \{\emptyset\}$, and recursively define $\hat{A}_k = \{(\mathbf{i}, j) : \mathbf{i} \in \hat{A}_{k-1}, 1 \le j \le \hat{N}_\mathbf{i}\}$, $k \ge 1$. Next, define the full marks according to $\hat{\mathbf{X}}_\emptyset = (\hat{N}_\emptyset, \hat{\mathbf{A}}_\emptyset)$ and $\hat{\mathbf{X}}_\mathbf{i} = (\hat{N}_\mathbf{i} + 1, \hat{\mathbf{A}}_\mathbf{i})$, $\mathbf{i} \neq \emptyset$, and let $\hat{T}(\hat{\mathbf{A}}) = \{\hat{\mathbf{X}}_\mathbf{i} : \mathbf{i} \in \hat{T}\}$.

For a CM, the distribution of the full marks is given by

$$\mathbb{P}_n(\hat{\mathbf{X}}_\emptyset \in \cdot) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}((D_i, \mathbf{a}_i) \in \cdot),$$

$$\mathbb{P}_n(\hat{\mathbf{X}}_\mathbf{i} \in \cdot) = \sum_{i=1}^n \frac{D_i}{L_n} \mathbf{1}((D_i, \mathbf{a}_i) \in \cdot), \qquad \mathbf{i} \neq \emptyset.$$

For the IR model, first let $\{b_n\}$ be a sequence such that $b_n \overset{\mathrm{P}}{\to} \infty$ and $b_n/\sqrt{n} \overset{\mathrm{P}}{\to} 0$ as $n \to \infty$, and use it to define $\bar{W}_i = W_i \wedge b_n$ and $\Lambda_n = \sum_{i=1}^n \bar{W}_i$. The marks on the coupled marked Galton–Watson process are given by

$$\mathbb{P}_n(\hat{\mathbf{X}}_\emptyset \in \cdot) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}((D_i, \mathbf{a}_i) \in \cdot \mid \mathbf{a}_i),$$

$$\mathbb{P}_n(\hat{\mathbf{X}}_\mathbf{i} \in \cdot) = \sum_{i=1}^n \frac{\bar{W}_i}{\Lambda_n} \mathbb{P}((D_i + 1, \mathbf{a}_i) \in \cdot \mid \mathbf{a}_i), \qquad \mathbf{i} \neq \emptyset,$$

where conditionally on $\mathbf{a}_i$, $D_i$ is a Poisson random variable with mean $\Lambda_n \bar{W}_i/(\theta n)$.

We will also need to extend our definition of an isomorphism for marked graphs.

**Definition 4.1.** A graph $G(V, E)$ is called a *vertex-weighted* graph if each of its vertices has a mark (weight) assigned to it. We say that the two vertex-weighted simple graphs $G(V, E)$ and $G(V', E')$ are *isomorphic* if there exists a bijection $\sigma : V \to V'$ such that edge $(i, j) \in E$ if and only if edge $(\sigma(i), \sigma(j)) \in E'$, and in addition, the marks of $i$ and $\sigma(i)$ are the same. For vertex-weighted multigraphs, we say that $G(V, E)$ and $G'(V', E')$ are *isomorphic* if there exists a bijection $\sigma : V \to V'$ such that $l(i) = l(\sigma(i))$ and $e(i, j) = e(\sigma(i), \sigma(j))$ for all $i \in V$ and all $(i, j) \in E$, where $l(i)$ is the number of self-loops of vertex $i$ and $e(i, j)$ is the number of edges from vertex $i$ to vertex $j$, and in addition, the marks of $i$ and $\sigma(i)$ are the same. In both cases, we write $G \simeq G'$.

The intermediate coupling theorem is given below.

**Theorem 4.1.** *Suppose* $G(V_n, E_n)$ *is either a CM or an IR satisfying Assumption* 2.1. *Then, for* $\mathcal{G}_I^{(k)}(\mathbf{a})$ *the depth-k neighborhood of a uniformly chosen vertex* $I \in V_n$, *there exists a marked Galton–Watson tree* $\hat{T}^{(k)}(\hat{\mathbf{A}})$ *restricted to its first k generations whose root corresponds to vertex* $I$ *and such that, for any fixed* $k \ge 1$, $\mathbb{P}_n(\mathcal{G}_I^{(k)}(\mathbf{a}) \not\simeq \hat{T}^{(k)}(\hat{\mathbf{A}})) \overset{\mathrm{P}}{\to} 0$, $n \to \infty$.

The proof of Theorem 4.1 is given separately for the two models being considered, the CM and the IR.

4.1.1. *Coupling for the configuration model.* To explore the neighborhood of depth $k$ of vertex $I \in G(V_n, E_n)$ we start by labeling the set of $L_n$ stubs in such a way that stubs $\{1, \ldots, D_1\}$ belong to vertex 1, stubs $\{D_1 + 1, \ldots, D_1 + D_2\}$ belong to vertex 2, and, in general, stubs $\{D_1 + \cdots + D_{m-1} + 1, \ldots, D_1 + D_m\}$ belong to vertex $m$.

For any $k \geq 0$, define the sets:

$$A_k = \text{set of vertices in } G(V_n, E_n) \text{ at distance } k \text{ from vertex } I,$$
$$J_k = \text{set of stubs belonging to vertices in } A_k,$$
$$V_k = \bigcup_{r=0}^{k} A_r,$$
$$\hat{A}_k = \text{set of nodes in } \hat{T} \text{ at distance } k \text{ from the root } \emptyset,$$
$$\hat{V}_k = \bigcup_{r=0}^{k} \hat{A}_r.$$

These sets will be constructed as we explore the graph.

To do a breadth-first exploration of $G(V_n, E_n)$, we start by selecting vertex $I$ uniformly at random. Next, let $J_0$ denote the set of stubs belonging to vertex $I$ and set $A_0 = \{I\}$. For $k \geq 1$, step $k$ in the exploration will identify all the stubs belonging to nodes in $A_k$:

*Step $k$, $k \geq 1$:*

(a) Initialize the sets $A_k = J_k = \varnothing$.

(b) For each vertex $i \in A_{k-1}$:

    (i) For each of the unpaired stubs of vertex $i$:

        (1) Pick an unpaired stub of vertex $i$ and sample uniformly at random a stub from the $L_n$ available. If the chosen stub is the stub currently being paired, or if it had already been paired, sample again until an unpaired stub is sampled.

        (2) If the chosen stub belongs to vertex $j$, draw an edge between vertices $i$ and $j$ using the chosen stub. If vertex $j$ had not yet been discovered, add it to $A_k$ and add all of its unpaired stubs to $J_k$.

The exploration terminates at step $k$ if $J_k = \varnothing$, at which point the component of $I$ will have been fully explored.

To couple the construction of $\hat{T}$, initialize $\hat{A}_0 = \{\emptyset\}$, identify $\emptyset$ with vertex $I$ in $G(V_n, E_n)$, and set $\hat{N}_\emptyset = D_I$, $\mathbf{a}_\emptyset = \mathbf{a}_I$. For $k \geq 1$, step $k$ in the construction will identify all the nodes in $\hat{A}_k$ by adding nodes in agreement with the exploration of the graph. Each node that is added to the tree will have a number of stubs equal to the total number of stubs of the corresponding vertex minus one (the one being used to create the edge), regardless of whether some of those stubs may already have been paired.

*Step $k$, $k \geq 1$:*

(a) Initialize the set $\hat{A}_k = \emptyset$.

(b) For each node $\mathbf{i} = (i_1, \dots, i_{k-1}) \in \hat{A}_{k-1}$:

    (i) For each $1 \leq r \leq \hat{N}_\mathbf{i}$:

        (1) Pick a stub uniformly at random from the $L_n$ available.

        (2) If the chosen stub belongs to vertex $j$, then add node $(\mathbf{i}, r)$ to $\hat{A}_k$ and set $\hat{N}_{(\mathbf{i},r)} = D_j - 1$, $\hat{\mathbf{A}}_{(\mathbf{i},r)} = \mathbf{a}_j$.

This process will end in step $k$ if $\hat{N}_\mathbf{i} = 0$ for all $\mathbf{i} \in \hat{A}_k$, or it may continue indefinitely.

Note that the coupling relies on using the same uniform random numbers in step (b)(i)(1) for the two constructions. Specifically, there is one uniform for each of the $L_n$ stubs which is used in step (b)(i)(1) of the tree construction, and in the first pick in the acceptance–rejection

step (b)(i)(1) of the graph exploration; in case of a rejection, independent uniforms are used until a stub is accepted.

**Definition 4.2.** We say that the coupling breaks in generation $\tau = k$ if:

- the first time we have to resample a stub in step (b)(i)(1) occurs while exploring a stub belonging to a vertex in $A_{k-1}$; or

- given that the above has not happened, a stub belonging to a vertex in $A_{k-1}$ is paired with a stub belonging to a previously encountered vertex (this vertex could be in either $A_{k-1}$ or the current set $A_k$).

Note that the exploration of the component of depth $k$ of vertex $I$ in $G(V_n, E_n)$ and the construction of the first $k$ generations of the tree $\hat{T}$ will be identical provided $\tau > k$, meaning $\mathcal{G}_I^{(k)}(\mathbf{a}) \simeq \hat{T}^{(k)}(\hat{\mathbf{A}})$ in the sense of Definition 4.1. This particular construction is standard in the analysis of the configuration model, with small variations depending on whether we need to keep track of completed generations or simply the number of stubs that have been paired. We refer the reader to [22, Chapter 4] for the stub-by-stub version and a full history of the model. Earlier versions of the coupling imposed finite second moment conditions that more recent proofs can omit (see, e.g., [23]).

*Proof of Theorem 4.1.* ($m = 1$) *for the CM.* From the observation made above, it suffices to show that the exploration of the $k$-neighborhood of vertex $I$ does not require us to resample any stub in step (b)(i)(1), nor does it sample a stub belonging to a vertex that had already been discovered. To compute the probability of successfully completing $k$ generations in $\hat{T}$ before the coupling breaks, write $\mathbb{P}_n\big(\mathcal{G}_I^{(k)}(\mathbf{a}) \neq \hat{T}^{(k)}(\hat{\mathbf{A}})\big) \leq \mathbb{P}_n(\tau \leq k)$.

The coupling breaks the first time we draw a stub belonging to a vertex that has already been explored: either a stub already paired, or one that is unpaired but already attached to the graph. The number of paired stubs when exploring a vertex in $A_{r-1}$ is less than or equal to $2 \sum_{j=1}^{r} |A_j| + |J_r|$, which corresponds to two stubs for each of the vertices at distance at most $r$ of $I$ and the unpaired stubs belonging to nodes in $J_r$. Note that up to the moment that the coupling breaks, we have $|A_j| = |\hat{A}_j|$ for all $0 \leq j \leq r$, and $|J_r| = |\hat{A}_{r+1}|$, so the probability that we break the coupling while exploring a vertex in $A_{r-1}$ is less than or equal to $P_r := (2/L_n) \sum_{j=1}^{r+1} |\hat{A}_j| \leq 2|\hat{V}_{r+1}|/L_n$, $r \geq 1$.

It follows that, for any $a_n > 0$,

$$
\begin{aligned}
\mathbb{P}_n(\tau \leq k) &= \mathbb{P}_n(\tau \leq k, |\hat{V}_{k+1}| \leq a_n) + \mathbb{P}_n(|\hat{V}_{k+1}| > a_n) \\
&\leq \sum_{r=1}^{k} \mathbb{P}_n(\tau = r, |\hat{V}_{r+1}| \leq a_n) + \mathbb{P}_n(|\hat{V}_{k+1}| > a_n) \\
&\leq \sum_{r=1}^{k} \mathbb{P}_n(\text{Bin}(\hat{A}_{r-1}, P_r) \geq 1, |\hat{V}_{r+1}| \leq a_n) + \mathbb{P}_n(|\hat{V}_{k+1}| > a_n) \\
&\leq \sum_{r=1}^{k} \mathbb{P}_n(\text{Bin}(a_n, 2a_n/L_n) \geq 1) + \mathbb{P}_n(|\hat{V}_{k+1}| > a_n) \\
&\leq \sum_{r=1}^{k} \frac{2a_n^2}{L_n} + \mathbb{P}_n(|\hat{V}_{k+1}| > a_n),
\end{aligned}
$$

where Bin$(n, p)$ represents a binomial random variable with parameters $(n, p)$. Hence, we have

$$\mathbb{P}_n\big(\mathcal{G}_I^{(k)}(\mathbf{a}) \not\simeq \hat{T}^{(k)}(\hat{\mathbf{A}})\big) \le \mathbb{P}_n(\tau \le k) \le (2ka_n^2/L_n) + \mathbb{P}_n(|\hat{V}_{k+1}| > a_n).$$

To analyze the last probability we use the first part of Theorem 4.3 to obtain that, for any fixed $k \ge 1$, there exists a tree $\mathcal{T}^{(k)}$ of depth $k$, whose distribution does not depend on $\mathscr{F}_n$, such that $\mathbb{P}_n\big(\hat{T}^{(k)} \not\simeq \mathcal{T}^{(k)}\big) \xrightarrow{P} 0$ as $n \to \infty$. Let $|\mathcal{A}_k|$ denote the size of the $k$th generation of that tree, define $|\mathcal{V}_{k+1}| = \sum_{j=0}^{k+1} |\mathcal{A}_j|$, and note that $\mathbb{P}_n\big(\mathcal{G}_I^{(k)}(\mathbf{a}) \not\simeq \hat{T}^{(k)}(\hat{\mathbf{A}})\big) \le (2ka_n^2/L_n) + \mathbb{P}(|\mathcal{V}_{k+1}| > a_n) + \mathbb{P}_n\big(\hat{T}^{(k)} \not\simeq \mathcal{T}^{(k)}\big)$. Choosing $a_n \xrightarrow{P} \infty$ so that $a_n^2/n \xrightarrow{P} 0$ as $n \to \infty$, and observing that $|\mathcal{V}_{k+1}| < \infty$ a.s., completes the proof. $\qquad\square$

4.1.2. *Coupling for the inhomogeneous random graph.* We couple the exploration of the component of vertex $I \in G(V_n, E_n)$ with a marked multi-type Galton–Watson process with $n$ types, one for each vertex in $G(V_n, E_n)$. A node of type $i \in \{1, 2, \ldots, n\}$ in the tree will have a Poisson number of offspring of type $j$ with mean $q_{ij}^{(n)} = \bar{W}_i \bar{W}_j/(\theta n)$, $1 \le j \le n$.

Similarly to the CM, define:

$$A_k = \text{set of vertices in } G(V_n, E_n) \text{ at distance } k \text{ from vertex } I,$$
$$V_k = \bigcup_{r=0}^k A_r,$$
$$\hat{A}_k = \text{set of nodes in } \hat{T} \text{ at distance } k \text{ from the root } \emptyset,$$
$$\hat{B}_k = \text{set of types of nodes in } \hat{A}_k,$$
$$\hat{V}_k = \bigcup_{r=0}^k \hat{A}_r.$$

We again do a breadth-first exploration of $G(V_n, E_n)$ starting from a uniformly chosen vertex $I$. To start, let $\{U_{ij} : i, j \ge 1\}$ be a sequence of i.i.d. Uniform[0, 1] random variables, independent of $\mathscr{F}_n$. We will use this sequence of i.i.d. uniforms to realize the Bernoulli random variables that determine the presence/absence of edges in $G(V_n, E_n)$. Set $A_0 = \{I\}$ and initialize the set $J = \emptyset$; the set $J$ keeps track of the vertices that have been fully explored (all its potential edges realized), and coincides with $V_{k-1}$ at the end of step $k$.

*Step $k$, $k \ge 1$:*

(a) Initialize the set $A_k = \emptyset$.

(b) For each vertex $i \in A_{k-1}$:

    (i) Let $X_{ij} = 1(U_{ij} > 1 - p_{ij}^{(n)})$ for each $j \in \{1, 2, \ldots, n\} \setminus J$.

    (ii) If $X_{ij} = 1$ draw an edge between vertices $i$ and $j$ and add vertex $j$ to $A_k$.

    (iii) Add vertex $i$ to set $J$.

The exploration terminates at the end of step $k$ if $A_k = \emptyset$, at which point the component of $I$ will have been fully explored.

To couple the construction of $\hat{T}$, initialize $\hat{A}_0 = \{\emptyset\}$ and identify $\emptyset$ with vertex $I$ in $G(V_n, E_n)$ as before; let $\hat{B}_0 = \{I\}$. To construct the tree, we sample for a node of type $i$ a Poisson number of offspring of type $j$ for each $j \in \{1, \ldots, n\}$. To do this, let $G(\cdot; \lambda)$ be the cumulative distribution function of a Poisson random variable with mean $\lambda$, and let $G^{-1}(u; \lambda) = \inf\{x \in \mathbb{R} : G(x; \lambda) \ge u\}$ denote its pseudoinverse. In order to keep the tree coupled with the exploration of the graph we use the same sequence of i.i.d. uniform random variables used to sample the edges in the graph. Initialize the set $\hat{J} = \emptyset$, which will keep track of the types that have appeared and whose offspring have been sampled. The precise construction is given below:

*Step k, k ≥ 1:*

(a) Initialize the sets $\hat{A}_k = \hat{B}_k = \varnothing$.

(b) For each node $\mathbf{i} = (i_1, \ldots, i_{k-1}) \in \hat{A}_{k-1}$:

   (i) If $\mathbf{i}$ has type $t \notin \hat{J}$:

      (1) For each type $j \in \{1, \ldots, n\} \setminus \hat{J}$, let $Z_{tj} = G^{-1}(U_{tj}; q_{tj}^{(n)})$, and create $Z_{tj}$ children of type $j$ for node $\mathbf{i}$. If $Z_{tj} \geq 1$, create $Z_{tj}$ children of type $j$ for node $\mathbf{i}$, each with node attribute equal to $\mathbf{a}_j$, and add $j$ to set $\hat{B}_k$.

      (2) For each type $j \in \hat{J}$, sample $Z_{tj}^* \sim \text{Poisson}(q_{tj}^{(n)})$, independently of the sequence $\{U_{ij} : i, j \geq 1\}$ and any other random variables. If $Z_{tj}^* \geq 1$ create $Z_{tj}^*$ children of type $j$ for node $\mathbf{i}$, each with attribute equal to $\mathbf{a}_j$.

      (3) Randomly shuffle all the children created in steps (b)(i)(1) and (b)(i)(2) and give them labels of the form $(\mathbf{i}, j)$, then add the labeled nodes to set $\hat{A}_k$. The node attributes will be denoted $\hat{\mathbf{A}}_{(\mathbf{i},j)} = \mathbf{a}_j$. (The shuffling avoids the label from providing information about its type).

      (4) Add type $t$ to set $\hat{J}$.

   (ii) If $\mathbf{i}$ has type $t \in \hat{J}$:

      (1) For each type $j \in \{1, \ldots, n\}$, sample $Z_{tj}^* \sim \text{Poisson}(q_{tj}^{(n)})$, independently of the sequence $\{U_{ij} : i, j \geq 1\}$ and any other random variables; create $Z_{tj}^*$ children of type $j$ for node $\mathbf{i}$, each with attribute equal to $\mathbf{a}_j$.

      (2) Randomly shuffle all the children created in step (b)(ii)(1) and give them labels of the form $(\mathbf{i}, j)$, attributes $\hat{\mathbf{A}}_{(\mathbf{i},j)} = \mathbf{a}_j$, and add the labeled nodes to set $\hat{A}_k$.

This construction may continue indefinitely, or may terminate at the end of step $k$ if $\hat{A}_k = \varnothing$.

We point out that this coupling is not standard, since in most of the literature on IR models the coupling is done between the binomial distribution (the degree of a vertex) and its coupled Poisson limit (see [6] or [22, Chapter 3]), where the proof uses the moments method). Moreover, most of the existing couplings consider first a multi-type Galton–Watson process with finitely many types, and then use a monotonicity argument to obtain the general case. The coupling described above avoids the need for this second step since the number of types grows as $n \to \infty$, and since it is based on coupling the individual Bernoulli random variables (edges) with their Poisson counterparts, it allows us to keep track of the vertex marks with no additional effort.

**Definition 4.3.** We say that the coupling breaks in generation $\tau = k$ if, for any node in $\hat{A}_{k-1}$, either:

- in step (b)(i)(1) we have $Z_{tj} \neq X_{tj}$ for some $j \in \{1, \ldots, n\} \setminus \hat{J}$;

- in step (b)(i)(1) we have $Z_{tj} \geq 1$ for some $j \in (\hat{B}_{k-1} \cup \hat{B}_k) \setminus \hat{J}$, in which case a cycle or self-loop is created; or,

- in step (b)(i)(2) we have $Z_{tj}^* \geq 1$ for some $j \in \hat{J}$.

We start by proving the following preliminary result. Throughout this section, let $\Delta_n := \int_0^1 \left| F_n^{-1}(u) - F^{-1}(u) \right| \, du \le W_1(\nu_n, \nu)$, where $F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(W_i \le x)$ and $F(x) = \mathbb{P}(W \le x)$. We also use the notation $X_n = O_{\mathrm{P}}(x_n)$ as $n \to \infty$ to mean that there exists a random variable $Y_n$ such that $|X_n| \le_{\mathrm{st}} Y_n$ and $Y_n/x_n \xrightarrow{\mathrm{P}} K$ for some finite constant $K$.

**Lemma 4.1.** *For any* $1 \le i \le n$ *we have* $\mathbb{P}_n\left( \max_{1 \le j \le n, j \ne i} |X_{ji} - Z_{ji}| \ge 1 \right) \le \min\{1, \mathbf{1}(W_i > b_n) + \mathcal{P}_n(i) + \bar{W}_i \eta_n\}$, *where* $\mathcal{P}_n(i) = \sum_{1 \le j \le n, j \ne i} \left| p_{ji}^{(n)} - (r_{ji}^{(n)} \wedge 1) \right|$, $\eta_n = (\Delta_n + g(b_n) + b_n^2/n + b_n^2 \Delta_n/(\theta n))/\theta$, *and* $g(x) = \mathbb{E}[(W - x)^+]$.

*Proof.* Let $R_{ij} = 1(U_{ij} > 1 - r_{ij}^{(n)})$, with $r_{ij}^{(n)} = W_i W_j/(\theta n)$. The union bound gives

$$\mathbb{P}_n\left( \max_{1 \le j \le n, j \ne i} |X_{ij} - Z_{ij}| \ge 1 \right) \le \mathbf{1}(W_i > b_n) + \mathbf{1}(W_i \le b_n) \sum_{1 \le j \le n, j \ne i} \mathbb{P}_n(|X_{ij} - Z_{ij}| \ge 1).$$

Now note that

$$
\begin{aligned}
\mathbb{P}_n(|X_{ij} - Z_{ij}| \ge 1) &= \mathbb{P}_n(|X_{ij} - Z_{ij}| \ge 1, |X_{ij} - R_{ij}| \ge 1) \\
&\quad + \mathbb{P}_n(|X_{ij} - Z_{ij}| \ge 1, |X_{ij} - R_{ij}| = 0) \\
&\le \mathbb{P}_n(|X_{ij} - R_{ij}| \ge 1) + \mathbb{P}_n(|R_{ij} - Z_{ij}| \ge 1).
\end{aligned}
$$

The first probability can be computed to be $\mathbb{P}_n(|X_{ij} - Z_{ij}| \ge 1) = \left| p_{ij}^{(n)} - (r_{ij}^{(n)} \wedge 1) \right|$.

To analyze each of the probabilities involving $R_{ij}$ and $Z_{ij}$, note that

$$
\begin{aligned}
\mathbb{P}_n(|R_{ij} - Z_{ij}| \ge 1) &= \mathbb{P}_n(R_{ij} = 0, Z_{ij} \ge 1) + \mathbb{P}_n(R_{ij} = 1, Z_{ij} = 0) + \mathbb{P}_n(R_{ij} = 1, Z_{ij} \ge 2) \\
&= \left( 1 - \left( 1 \wedge r_{ij}^{(n)} \right) - e^{-q_{ij}^{(n)}} \right)^+ + \left( e^{-q_{ij}^{(n)}} - 1 + \left( 1 \wedge r_{ij}^{(n)} \right) \right)^+ \\
&\quad + \min\left\{ 1 - e^{-q_{ij}^{(n)}} \left( 1 + q_{ij}^{(n)} \right), \, \left( 1 \wedge r_{ij}^{(n)} \right) \right\} \\
&= \left| 1 - \left( 1 \wedge r_{ij}^{(n)} \right) - e^{-q_{ij}^{(n)}} \right| + \min\left\{ \left( 1 \wedge r_{ij}^{(n)} \right), \, e^{-q_{ij}^{(n)}} \left( e^{q_{ij}^{(n)}} - 1 - q_{ij}^{(n)} \right) \right\}.
\end{aligned}
$$

Now use the inequalities $e^{-x} \ge 1 - x$, $e^{-x} - 1 + x \le x^2/2$, and $e^x - 1 - x \le x^2 e^x/2$ for $x \ge 0$, to obtain that

$$
\begin{aligned}
\mathbb{P}_n\left( |X_{ij} - Z_{ij}| \ge 1 \right) &\le r_{ij}^{(n)} - q_{ij}^{(n)} + \left| 1 - q_{ij}^{(n)} - e^{-q_{ij}^{(n)}} \right| + e^{-q_{ij}^{(n)}} \left( e^{q_{ij}^{(n)}} - 1 - q_{ij}^{(n)} \right) \\
&= r_{ij}^{(n)} - q_{ij}^{(n)} + e^{-q_{ij}^{(n)}} - 1 + q_{ij}^{(n)} + e^{-q_{ij}^{(n)}} \left( e^{q_{ij}^{(n)}} - 1 - q_{ij}^{(n)} \right) \\
&\le r_{ij}^{(n)} - q_{ij}^{(n)} + \left( q_{ij}^{(n)} \right)^2.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\mathbf{1}(W_i \le b_n) \sum_{1 \le j \le n, j \ne i} \mathbb{P}_n(|X_{ij} - Z_{ij}| \ge 1) \\
&\le \mathbf{1}(W_i \le b_n) \sum_{1 \le j \le n, j \ne i} \left( \left| p_{ij}^{(n)} - (r_{ij}^{(n)} \wedge 1) \right| + r_{ij}^{(n)} - q_{ij}^{(n)} + \left( q_{ij}^{(n)} \right)^2 \right) \\
&\le \mathcal{P}_n(i) + \sum_{1 \le j \le n, j \ne i} \frac{\bar{W}_i (W_j - \bar{W}_j)}{\theta n} + \frac{(\bar{W}_i)^2}{(\theta n)^2} \sum_{1 \le j \le n, j \ne i} (\bar{W}_j)^2 \\
&\le \mathcal{P}_n(i) + \frac{\bar{W}_i}{\theta n} \sum_{j=1}^n (W_j - b_n)^+ + \frac{(\bar{W}_i)^2 b_n \Lambda_n}{(\theta n)^2}.
\end{aligned}
$$

To further bound the second term, note that if we let $W^{(n)}$ denote a random variable distributed according to $F_n$ and $W$ a random variable distributed according to $F$, then $\frac{1}{n} \sum_{j=1}^{n} (W_j - b_n)^+ = \mathbb{E}_n[(W^{(n)} - b_n)^+] \leq \mathbb{E}_n[|W^{(n)} - W| + (W - b_n)^+] = \Delta_n + g(b_n)$. And for the last term,

$$\frac{(\bar{W}_i)^2 b_n \Lambda_n}{(\theta n)^2} \leq \frac{\bar{W}_i b_n^2}{\theta^2 n} \cdot \mathbb{E}_n[W^{(n)}] \leq \frac{\bar{W}_i b_n^2}{\theta^2 n}(\Delta_n + \mathbb{E}[W]).$$

We conclude that for $\mathcal{E}_n$ as defined in the statement of the lemma,

$$\mathbf{1}(W_i \leq b_n) \sum_{1 \leq j \leq n, j \neq i} \mathbb{P}_n(|X_{ij} - Z_{ij}| \geq 1)$$
$$\leq \mathcal{P}_n(i) + \frac{\bar{W}_i}{\theta}(\Delta_n + g(b_n)) + \frac{\bar{W}_i b_n^2}{\theta^2 n}(\Delta_n + \mathbb{E}[W^-])$$
$$\leq \mathcal{P}_n(i) + \eta_n \bar{W}_i,$$

which in turn yields $\mathbb{P}_n(\max_{1 \leq j \leq n, j \neq i} |X_{ij} - Z_{ij}| \geq 1) \leq \min\{1, \mathbf{1}(W_i > b_n) + \mathcal{P}_n(i) + \eta_n \bar{W}_i\}$. $\qquad\square$

*Proof of Theorem* 4.1. $(m = 1)$ *for the IR.* We start by defining the following events:

$$F_i(I, J, L) = \left\{ \max_{j \in I} |X_{ji} - Z_{ji}| = 0, \sum_{j \in J} Z_{ji}^* + \sum_{j \in L} Z_{ji} = 0 \right\},$$

$\mathbb{B}_i = \{\text{current set } \hat{B}_{k-1} \cup \hat{B}_k \text{ when the neighbors of } i \in A_{k-1} \text{ are explored}\},$

$\mathbb{J}_i = \{\text{current set } J \text{ when the neighbors of } i \text{ are explored}\},$

$H_k = \bigcap_{i \in A_{k-1}} F_i(\{1, \ldots, n\} \setminus \mathbb{J}_i, \mathbb{J}_i, \mathbb{B}_i \setminus \mathbb{J}_i),$

$M_k = \{|\hat{V}_k| \leq s_n\}.$

Next, note that $\mathbb{P}_n(\tau \leq k) \leq \mathbb{P}_n(\tau \leq k, M_k) + \mathbb{P}_n(M_k^c) \leq \sum_{r=1}^{k} \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_{n,i}(\tau = r, M_r) + \mathbb{P}_n(M_k^c)$, where the last probability can be bounded using the first part of Theorem 4.3 as at the end of the proof of Theorem 4.1 for the CM. Specifically, $\mathbb{P}_n(M_k^c) \leq \mathbb{P}(|\mathcal{V}_{k+1}| > s_n) + \mathbb{P}_n(\hat{T}^{(k)} \not\simeq \mathcal{T}^{(k)})$, where $|\mathcal{V}_{k+1}| = \sum_{j=0}^{k+1} |\mathcal{A}_j| < \infty$ a.s. and the distribution of $\mathcal{T}^{(k)}$ does not depend on $\mathscr{F}_n$.

Now note that, for any $r \geq 1$, $\mathbb{P}_{n,i}(\tau = r, M_r) = \mathbb{P}_{n,i}(M_r \cap \bigcap_{m=1}^{r-1} H_m \cap H_r^c)$, with the convention that $\bigcap_{m=1}^{0} H_m = \Omega$. Let $\mathcal{F}_t$ denote the sigma-algebra that contains the history of the exploration process in the graph as well as that of its coupled tree, up to the end of step $t$ of the graph exploration process. It follows that we can write $\mathbb{P}_{n,i}(\tau = r, M_r) = \mathbb{E}_{n,i}\big[\mathbf{1}\big(M_{r-1} \cap \bigcap_{m=1}^{r-1} H_m\big)\mathbb{P}_n(M_r \cap H_r^c \mid \mathcal{F}_{r-1})\big]$. To analyze the conditional probability inside the expectation above, note that, conditionally on $\mathcal{F}_{r-1}$, the set $A_{r-1}$ is known, and recall that the set $J = V_{r-2}$ at the beginning of step $r$ (assuming $r \geq 2$, otherwise $J = \varnothing$). Therefore, by the union bound and the independence among the edges, we have:

$$\mathbb{P}_n(M_r \cap H_r^c \mid \mathcal{F}_{r-1}) = \mathbb{P}_n\left(M_r \cap \bigcup_{i \in A_{r-1}} F_i(\{1, \ldots, n\} \setminus \mathbb{J}_i, \mathbb{J}_i, \mathbb{B}_i \setminus \mathbb{J}_i)^c \mid \mathcal{F}_{r-1}\right)$$
$$\leq \sum_{i \in A_{r-1}} \mathbb{P}_n(M_r \cap F_i(\{1, \ldots, n\} \setminus \mathbb{J}_i, \mathbb{J}_i, \mathbb{B}_i \setminus \mathbb{J}_i)^c \mid \mathcal{F}_{r-1})$$

$$\leq \sum_{i \in A_{r-1}} \min \left\{ 1, \ \mathbb{P}_n \Big( \max_{j \in \{1,\dots,n\} \setminus \mathbb{J}_i} |X_{ji} - Z_{ji}| \geq 1 \mid \mathcal{F}_{r-1} \Big) \right.$$
$$\left. + \mathbb{P}_n \Big( M_r \cap \Big\{ \sum_{j \in \mathbb{J}_i} Z_{ji}^* + \sum_{j \in \mathbb{B}_i \setminus \mathbb{J}_i} Z_{ji} \geq 1 \Big\} \mid \mathcal{F}_{r-1} \Big) \right\}.$$

Now use the independence of the edges from the rest of the exploration process and Lemma 4.1 to obtain that

$$\mathbb{P}_n \Big( \max_{\{1,\dots,n\} \setminus \mathbb{J}_i} |X_{ji} - Z_{ji}| \geq 1 \mid \mathcal{F}_{r-1} \Big) \leq \mathbb{P}_n \Big( \max_{1 \leq j \leq n, j \neq i} |X_{ji} - Z_{ji}| \geq 1 \Big)$$
$$\leq \mathbf{1}(W_i > b_n) + \mathcal{P}_n(i) + \bar{W}_i \eta_n.$$

Next, condition further on the exploration up to the moment we are about to explore the neighbors of $i$, and use the independence of the edges from the rest of the exploration process to obtain that

$$\mathbb{P}_n \Big( M_r \cap \Big\{ \sum_{j \in \mathbb{J}_i} Z_{ji}^* + \sum_{j \in \mathbb{B}_i \setminus \mathbb{J}_i} Z_{ji} \geq 1 \Big\} \mid \mathcal{F}_{r-1} \Big)$$
$$\leq \mathbb{E}_n \Big[ \mathbf{1}(|\hat{V}_r| \leq s_n) \Big( 1 - \exp \Big\{ -\sum_{j \in \mathbb{B}_i} q_{ji}^{(n)} \Big\} \Big) \mid \mathcal{F}_{r-1} \Big]$$
$$= \mathbb{E}_n \Big[ \mathbf{1}(|\hat{V}_r| \leq s_n) \Big( 1 - \exp \Big\{ -\frac{\bar{W}_i}{\theta n} \sum_{j \in \cup \mathbb{B}_i} \bar{W}_j \Big\} \Big) \mid \mathcal{F}_{r-1} \Big]$$
$$\leq \mathbb{E}_n \Big[ \mathbf{1}(|\hat{V}_r| \leq s_n) \Big( 1 - \exp \Big\{ -\frac{b_n \bar{W}_i}{\theta n} |\mathbb{B}_i| \Big\} \Big) \mid \mathcal{F}_{r-1} \Big]$$
$$\leq \frac{b_n \bar{W}_i}{\theta n} s_n,$$

where in the last inequality we used $1 - e^{-x} \leq x$ for $x \geq 0$ and $|\mathbb{B}_i| \leq |\hat{V}_r| \leq s_n$.

It follows that

$$\mathbb{P}_{n,i}(\tau = r, M_r) \leq \mathbb{E}_{n,i} \Big[ \mathbf{1} \Big( M_{r-1} \cap \bigcap_{m=1}^{r-1} H_m \Big) \sum_{j \in A_{r-1}} \min \Big\{ 1, \ \mathbf{1}(W_i > b_n) + \mathcal{P}_n(j) \right.$$
$$\left. + \bar{W}_i \eta_n + \frac{b_n \bar{W}_i}{\theta n} s_n \Big\} \Big].$$

To analyze this remaining expectation we note that on the event $\bigcap_{m=1}^{r-1} H_m$ the coupling has not been broken yet, and therefore we can can replace $A_{r-1}$ with its tree counterpart $\hat{A}_{r-1}$. Also, note that by [18, Lemma 3.4], the types of the nodes in each of the sets $\hat{A}_k$ are independent of the type of their parents. We will then identify the nodes in $\hat{A}_{r-1}$ as $\{Y_1, \dots, Y_{|\hat{A}_{r-1}|}\}$, where, for any $t \geq 1$, $\mathbb{P}_n(Y_t = j) = \bar{W}_j / \Lambda_n$, $j = 1, 2, \dots, n$.

It follows that

$$\mathbb{P}_{n,i}(\tau = r, M_r)$$

$$\leq \mathbb{E}_{n,i}\left[\mathbf{1}(M_{r-1}) \sum_{t=1}^{|\hat{A}_{r-1}|} \min\left\{1, \mathbf{1}(W_{Y_t} > b_n) + \mathcal{P}_n(Y_t) + \bar{W}_{Y_t}\eta_n + \frac{b_n\bar{W}_{Y_t}}{\theta n}s_n\right\}\right]$$

$$\leq \mathbb{E}_{n,i}\left[\sum_{t=1}^{\lfloor s_n \rfloor} \min\left\{1, \mathbf{1}(W_{Y_t} > b_n) + \mathcal{P}_n(Y_t) + \bar{W}_{Y_t}\eta_n + \frac{b_n\bar{W}_{Y_t}}{\theta n}s_n\right\}\right]$$

$$\leq \sum_{t=1}^{\lfloor s_n \rfloor} \mathbb{E}_{n,i}\left[\mathbf{1}(W_{Y_t} > b_n) + \mathcal{P}_n(Y_t) + \bar{W}_{Y_t}\eta_n + \frac{b_n\bar{W}_{Y_t}}{\theta n}s_n\right]$$

$$= \lfloor s_n \rfloor \mathbb{E}_n\left[\mathbf{1}(W_{Y_1} > b_n) + \mathcal{P}_n(Y_1) + \bar{W}_{Y_1}\eta_n + \frac{b_n\bar{W}_{Y_1}}{\theta n}s_n\right].$$

To compute the last expectation, let $(W^{(n)}, W)$ be constructed according to an optimal coupling of $F_n$ and $F$. Let $\bar{W}^{(n)} = W^{(n)} \wedge b_n$. Then, for any $c_n \geq 1$,

$$\lfloor s_n \rfloor \mathbb{E}_n\left[\mathbf{1}(W_{Y_1} > b_n) + \mathcal{P}_n(Y_1) + \bar{W}_{Y_1}\eta_n + \frac{b_n\bar{W}_{Y_1}}{\theta n}s_n\right]$$

$$\leq s_n \sum_{j=1}^{n} \frac{\bar{W}_j}{\Lambda_n}\left(\mathbf{1}(W_j > b_n) + \mathcal{P}_n(j) + \bar{W}_j\eta_n + \frac{b_n\bar{W}_j}{\theta n}s_n\right)$$

$$\leq \frac{s_n n}{\Lambda_n} \cdot \frac{1}{n}\sum_{j=1}^{n}(\bar{W}_j - c_n)^+ + \frac{s_n n}{\Lambda_n} \cdot \frac{1}{n}\sum_{j=1}^{n}c_n\left(\mathbf{1}(W_j > b_n) + \mathcal{P}_n(j) + \bar{W}_j\eta_n + \frac{b_n\bar{W}_j}{\theta n}s_n\right)$$

$$= \frac{s_n}{\mathbb{E}_n[\bar{W}^{(n)}]}\left(\mathbb{E}_n\left[(\bar{W}^{(n)} - c_n)^+\right] + c_n\mathbb{P}_n(W^{(n)} > b_n) + c_n\mathcal{E}_n\right.$$

$$\left. + \mathbb{E}_n[\bar{W}^{(n)}]\left(c_n\eta_n + \frac{c_n b_n s_n}{\theta n}\right)\right)$$

$$= O_P\left(s_n\left(g(c_n) + \Delta_n + c_n\mathbb{P}(W > b_n) + c_n\mathcal{E}_n + c_n\eta_n + \frac{c_n b_n s_n}{n}\right)\right)$$

as $n \to \infty$, and since $\eta_n = O_P(\Delta_n + g(b_n) + b_n^2/n)$, we conclude that $\mathbb{P}_{n,i}(\tau = r, M_r) = O_P(s_n(g(c_n) + c_n\mathcal{E}_n + c_n\Delta_n + c_ng(b_n) + c_nb_n^2/n))$ as $n \to \infty$. It now follows from the beginning of the proof that $\mathbb{P}_n(\tau \leq k) \leq O_P(ks_n(g(c_n) + c_n\mathcal{E}_n + c_n\Delta_n + c_ng(b_n) + c_nb_n^2/n)) + \mathbb{P}(|\mathcal{V}_{k+1}| > s_n) + \mathbb{P}_n(\hat{\mathcal{T}}^{(k)} \not\simeq \mathcal{T}^{(k)})$ as $n \to \infty$. Since $\lim_{x\to\infty} g(x) = 0$, choosing $c_n = (\mathcal{E}_n + \Delta_n + g(b_n) + b_n^2/n)^{-1/2}$ and $s_n = (g(c_n) + c_n^{-1/2})^{-1/2}$ proves the theorem. $\square$

## 4.2. Discrete coupling for directed graphs

The equivalent of Theorem 4.1 ($m = 1$) for directed graphs has already been proved, under conditions equivalent to those in Assumption 3.1, in [20, Theorem 6.3] for the DCM, and in [18, Theorem 3.7] for the IRD. Hence, we only need to describe the distribution of the intermediate tree and state the coupling theorem. The descriptions of the couplings follow, with some adjustments, those from Sections 4.1.1 and 4.1.2. However, the precise descriptions in the directed case can be found in [9, Section 5.2] for the DCM and [18, Section 3.2.2] for the IRD.

In the directed case, the intermediate tree $\hat{T}$ is constructed using a sequence of condition-ally independent (given $\mathscr{F}_n$) random vectors $\{(\hat{N}_\mathbf{i}, \hat{D}_\mathbf{i}, \hat{\mathbf{A}}_\mathbf{i}) : \mathbf{i} \in \mathcal{U}\}$ in $\mathcal{S}$, with $\{(\hat{N}_\mathbf{i}, \hat{D}_\mathbf{i}, \hat{\mathbf{A}}_\mathbf{i}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ conditionally i.i.d. The tree $\hat{T}$ is constructed as in the undirected case using the $\{\hat{N}_\mathbf{i}\}$, with all edges pointing towards the root, and the full marks take the form $\hat{\mathbf{X}}_\mathbf{i} = (\hat{N}_\mathbf{i}, \hat{D}_\mathbf{i}, \hat{\mathbf{A}}_\mathbf{i})$, $\mathbf{i} \in \mathcal{U}$. The marked tree is given by $\hat{T}(\hat{\mathbf{A}}) = \{\hat{\mathbf{X}}_\mathbf{i} : \mathbf{i} \in \hat{T}\}$.

We now specify the distribution of the full marks, which in the case of a DCM is given by

$$\mathbb{P}_n\big(\hat{\mathbf{X}}_\emptyset \in \cdot\big) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}((D_i^-, D_i^+, \mathbf{a}_i) \in \cdot),$$

$$\mathbb{P}_n\big(\hat{\mathbf{X}}_\mathbf{i} \in \cdot\big) = \sum_{i=1}^n \frac{D_i^+}{L_n} \mathbf{1}((D_i^-, D_i^+, \mathbf{a}_i) \in \cdot), \quad \mathbf{i} \neq \emptyset.$$

For the IRD model, first let $\{a_n\}$ and $\{b_n\}$ be sequences such that $a_n \wedge b_n \xrightarrow{\text{P}} \infty$ and $a_n b_n / n \xrightarrow{\text{P}} 0$ as $n \to \infty$, and use them to define $\bar{W}_i^- = W_i^- \wedge a_n$ and $\bar{W}_i^+ = W_i^+ \wedge b_n$, with $\Lambda_n^- = \sum_{i=1}^n \bar{W}_i^-$ and $\Lambda_n^+ = \sum_{i=1}^n \bar{W}_i^+$. The marks on the coupled marked Galton–Watson process are given by

$$\mathbb{P}_n\big(\hat{\mathbf{X}}_\emptyset \in \cdot\big) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}((D_i^-, D_i^+, \mathbf{a}_i) \in \cdot \mid \mathbf{a}_i),$$

$$\mathbb{P}_n\big(\hat{\mathbf{X}}_\mathbf{i} \in \cdot\big) = \sum_{i=1}^n \frac{\bar{W}_i^+}{\Lambda_n^+} \mathbb{P}((D_i^-, D_i^+ + 1, \mathbf{a}_i) \in \cdot \mid \mathbf{a}_i), \quad \mathbf{i} \neq \emptyset,$$

where, conditionally on $\mathbf{a}_i$, $D_i^-$ and $D_i^+$ are independent Poisson random variables with means $\Lambda_n^+ \bar{W}_i^+ / (\theta n)$ and $\Lambda_n^- \bar{W}_i^- / (\theta n)$, respectively.

The intermediate coupling theorem for directed graphs is given below, and it is a direct consequence of [20, Theorem 6.3] and [18, Theorem 3.7].

**Theorem 4.2.** *Suppose $G(V_n, E_n)$ is either a DCM or an IRD satisfying Assumption 3.1. Then, for $\mathcal{G}_I^{(k)}(\mathbf{a})$ the depth-k neighborhood of a uniformly chosen vertex $I \in V_n$, there exists a marked Galton–Watson tree $\hat{T}^{(k)}(\hat{\mathbf{A}})$, restricted to its first k generations, whose root corresponds to vertex I and is such that, for any fixed $k \geq 1$, $\mathbb{P}_n\big(\mathcal{G}_I^{(k)}(\mathbf{a}) \not\simeq \hat{T}^{(k)}(\hat{\mathbf{A}})\big) \xrightarrow{\text{P}} 0$, $n \to \infty$.*

### 4.3. Coupling between two trees

In view of Theorems 4.1 and 4.2, the proofs of the main theorems, Theorems 2.1 and 3.1 ($m = 1$), will be complete once we establish that with high probability the intermediate tree $\hat{T}^{(k)}$ is isomorphic to the limiting tree $\mathcal{T}^{(k)}$, and that the node marks in the two trees are within $\varepsilon$ distance of each other.

Note that there is no need to consider the undirected and directed cases separately, since they only differ on the sample space for the full marks, $\hat{\mathbf{X}}_\mathbf{i} / X_\mathbf{i}$, which take values in $\mathcal{S} = \mathbb{N} \times \mathbb{R} \times \mathcal{S}'$ in the undirected case and $\mathcal{S} = \mathbb{N} \times \mathbb{N} \times \mathbb{R} \times \mathbb{R} \times \mathcal{S}'$ in the directed one. For the directed case, all edges in the trees point towards the root.

The coupling theorem between the two trees is the following. The proof of the main theorems, Theorems 2.1 and 3.1 ($m = 1$), follow directly from combining Theorems 4.1 and 4.3 in the undirected case, and Theorems 4.2 and 4.3 in the directed one.

**Theorem 4.3.** *Under Assumption* 2.1 *or* 3.1, *as appropriate, there exists a coupling of* $\hat{T}^{(k)}(\hat{\mathbf{A}})$ *and* $\mathcal{T}^{(k)}(A)$ *such that* $\mathbb{P}_n(\hat{T}^{(k)} \not\simeq \mathcal{T}^{(k)}) \overset{P}{\to} 0$, $n \to \infty$, *and such that, for any* $\varepsilon > 0$, $\mathbb{E}_n[\rho(\hat{\mathbf{X}}_\emptyset, X_\emptyset)] \overset{P}{\to} 0$ *and* $\mathbb{P}_n(\bigcap_{\mathbf{i} \in \mathcal{T}^{(k)}} \{\rho(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) \le \varepsilon\}, \hat{T}^{(k)} \simeq \mathcal{T}^{(k)}) \overset{P}{\to} 1$ *as* $n \to \infty$.

Before proving Theorem 4.3, we need to prove a couple of technical lemmas. The first establishes the existence of couplings for the node attributes, whose distributions are given by $\nu_n(\cdot) = \mathbb{P}_n(\hat{A} \in \cdot) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathbf{a}_i \in \cdot)$ and $\nu(\cdot) = \mathbb{P}(A \in \cdot)$, and their size-biased versions. Recall that in the undirected case the node attributes are of the form $\mathbf{a}_i = (D_i, \mathbf{b}_i)$ in the CM and $\mathbf{a}_i = (W_i, \mathbf{b}_i)$ in the IR, while in the directed case they take the form $\mathbf{a}_i = (D_i^-, D_i^+, \mathbf{b}_i)$ in the DCM and $\mathbf{a}_i = (W_i^-, W_i^+, \mathbf{b}_i)$ in the IRD. In the undirected case, the size-bias is done with respect to the first coordinate, while in the directed case with respect to the second one. Specifically, the size-biased attributes in the undirected case take the form

$$\mathbb{P}_n(\hat{\mathbf{A}}_b \in \cdot) = \begin{cases} L_n^{-1} \sum_{i=1}^n D_i \mathbf{1}((D_i, \mathbf{b}_i) \in \cdot) & \text{in the CM,} \\ \Lambda_n^{-1} \sum_{i=1}^n \bar{W}_i \mathbf{1}((W_i, \mathbf{b}_i) \in \cdot) & \text{in the IR} \end{cases}$$

and

$$\mathbb{P}(A_b \in \cdot) = \begin{cases} \mathbb{E}[\mathscr{D}\mathbf{1}((D, \mathbf{B}) \in \cdot)]/\mathbb{E}[\mathscr{D}] & \text{in the CM,} \\ \mathbb{E}[W\mathbf{1}((W, \mathbf{B}) \in \cdot)]/\mathbb{E}[W] & \text{in the IR,} \end{cases}$$

while in the directed case they take the form

$$\mathbb{P}_n(\hat{\mathbf{A}}_b \in \cdot) = \begin{cases} L_n^{-1} \sum_{i=1}^n D_i^+ \mathbf{1}((D_i^-, D_i^+, \mathbf{b}_i) \in \cdot) & \text{in the DCM,} \\ (\Lambda_n^+)^{-1} \sum_{i=1}^n \bar{W}_i^+ \mathbf{1}((W_i^-, W_i^+, \mathbf{b}_i) \in \cdot) & \text{in the IRD} \end{cases}$$

and

$$\mathbb{P}(A_b \in \cdot) = \begin{cases} \mathbb{E}[\mathscr{D}^+ \mathbf{1}((\mathscr{D}^-, \mathscr{D}^+, \mathbf{B}) \in \cdot)]/\mathbb{E}[\mathscr{D}^+] & \text{in the DCM,} \\ \mathbb{E}[W^+ \mathbf{1}((W^-, W^+, \mathbf{B}) \in \cdot)]/\mathbb{E}[W^+] & \text{in the IRD.} \end{cases}$$

For the undirected case, let $\rho''$ be the metric on $\mathcal{S}'' = [0, \infty) \times \mathcal{S}'$ given by $\rho''(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + \rho'(\mathbf{x}_2, \mathbf{y}_2)$, $\mathbf{x} = (x_1, \mathbf{x}_2)$, $\mathbf{y} = (y_1, \mathbf{y}_2)$, and for the directed case let $\rho''$ be the metric on $\mathcal{S}'' = [0, \infty) \times [0, \infty) \times \mathcal{S}'$ given by $\rho''(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_1 - y_2| + \rho'(\mathbf{x}_3, \mathbf{y}_3)$, $\mathbf{x} = (x_1, x_2, \mathbf{x}_3)$, $\mathbf{y} = (y_1, y_2, \mathbf{y}_3)$.

**Lemma 4.2.** *Under Assumption* 2.1 *or* 3.1, *as appropriate, there exist couplings* $(\hat{\mathbf{A}}, A)$ *and* $(\hat{\mathbf{A}}_b, A_b)$ *constructed on the same probability space* $(\mathcal{S}'', \mathscr{F}_n, \mathbb{P}_n)$ *such that* $\mathbb{E}_n[\rho''(\hat{\mathbf{A}}, A)] \overset{P}{\to} 0$, $\rho''(\hat{\mathbf{A}}, A) \overset{P}{\to} 0$, *and* $\rho''(\hat{\mathbf{A}}_b, A_b) \overset{P}{\to} 0$ *as* $n \to \infty$.

*Proof.* Assumptions 2.1 and 3.1 state that $W_1(\nu_n, \nu) \overset{P}{\to} 0$ as $n \to \infty$, and by the properties of the Wasserstein metric (see [24, Theorem 4.1]), there exists an optimal coupling $(\hat{\mathbf{A}}, A)$ such that $\mathbb{E}_n[\rho''(\hat{\mathbf{A}}, A)] = W_1(\nu_n, \nu) \overset{P}{\to} 0$ and $\rho''(\hat{\mathbf{A}}, A) \overset{P}{\to} 0$ as $n \to \infty$.

For the biased versions, note that it suffices to prove the lemma for the undirected case, since a simple rearrangement of terms such as $\mathbf{a}_i' = (D_i', \mathbf{b}_i') := (D_i^+, D_i^-, \mathbf{b}_i)$ or $\mathbf{a}_i' = (W_i', \mathbf{b}_i') :=$

$(W_i^+, W_i^-, \mathbf{b}_i)$ reduces the directed case to the undirected one. Through the remainder of the proof, we write $\hat{\mathbf{A}} = (\hat{Y}, \hat{\mathbf{B}})$ and $\mathbf{A} = (Y, \mathbf{B})$ to avoid having to separate the CM and IR cases.

Next, note that we only need to show that $\hat{\mathbf{A}}_b \Rightarrow \mathbf{A}_b$ as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution, since then we can take the almost sure representation to obtain that $\rho''(\hat{\mathbf{A}}_b, \mathbf{A}_b) \xrightarrow{P} 0$. To this end, let $f : \mathcal{S}'' \to \mathbb{R}$ be a bounded and continuous function, and let $(\hat{\mathbf{A}}, \mathbf{A})$ be the one from the beginning of the proof. Let $\tilde{Y} = \hat{Y}$ if the graph is a CM or $\tilde{Y} = \hat{Y} \wedge b_n$ if it is an IR. Then,

$$
\begin{aligned}
\left| \mathbb{E}_n\big[f(\hat{\mathbf{A}}_b)\big] - \mathbb{E}[f(\mathbf{A}_b)] \right| &= \left| \frac{1}{\mathbb{E}_n[\tilde{Y}]} \mathbb{E}_n\big[\tilde{Y} f(\hat{\mathbf{A}})\big] - \frac{1}{\mathbb{E}[Y]} \mathbb{E}[Y f(\mathbf{A})] \right| \\
&\leq \frac{1}{\mathbb{E}_n[\tilde{Y}]} \left( \left| \mathbb{E}_n\big[(\tilde{Y} - Y)f(\hat{\mathbf{A}})\big] \right| + \left| \mathbb{E}_n\big[Y(f(\hat{\mathbf{A}}) - f(\mathbf{A}))\big] \right| \right) \\
&\quad + \left| \frac{1}{\mathbb{E}_n[\tilde{Y}]} - \frac{1}{\mathbb{E}[Y]} \right| \left| \mathbb{E}[Y f(\mathbf{A})] \right| \\
&\leq \frac{1}{\mathbb{E}_n[\tilde{Y}]} \left( \mathbb{E}_n[|\tilde{Y} - Y|] \sup_{\mathbf{a} \in \mathcal{S}''} |f(\mathbf{a})| + \left| \mathbb{E}_n\big[Y(f(\hat{\mathbf{A}}) - f(\mathbf{A}))\big] \right| \right) \\
&\quad + \frac{\mathbb{E}_n[|\tilde{Y} - Y|]}{\mathbb{E}_n[\tilde{Y}]\mathbb{E}[Y]} |\mathbb{E}[Y f(\mathbf{A})]| \\
&\leq \frac{1}{\mathbb{E}_n[\tilde{Y}]} \left( \mathbb{E}_n\big[|\tilde{Y} - Y|\big] 2 \sup_{\mathbf{a} \in \mathcal{S}''} |f(\mathbf{a})| + \mathbb{E}_n\big[Y|f(\hat{\mathbf{A}}) - f(\mathbf{A})|\big] \right).
\end{aligned}
$$

Since $W_1(v_n, v) \xrightarrow{P} 0$ implies that $\mathbb{E}_n[|\hat{Y} - Y|] \xrightarrow{P} 0$ as $n \to \infty$, we have $\mathbb{E}_n\big[|\tilde{Y} - Y|\big] \leq \mathbb{E}_n\big[|\hat{Y} - Y|\big] + \mathbb{E}[Y \mathbf{1}(Y > b_n)] \xrightarrow{P} 0$ and $\mathbb{E}_n[\tilde{Y}] \xrightarrow{P} \mathbb{E}[Y]$ as $n \to \infty$. And, by the dominated convergence theorem, $\lim_{n\to\infty} \mathbb{E}\big[\mathbb{E}_n\big[Y|f(\hat{\mathbf{A}}) - f(\mathbf{A})|\big]\big] = \mathbb{E}\big[\lim_{n\to\infty} Y|f(\hat{\mathbf{A}}) - f(\mathbf{A})|\big] = 0$. Hence, $\mathbb{E}_n\big[Y|f(\hat{\mathbf{A}}) - f(\mathbf{A})|\big] \xrightarrow{P} 0$ as $n \to \infty$, and $\mathbf{A}_b \Rightarrow \mathbf{A}_b$ as required. $\qquad\square$

The second technical lemma relates the convergence of the attributes to that of the full marks.

**Lemma 4.3.** *Suppose Assumption* 2.1 *or* 3.1 *holds, as appropriate, and let* $(\hat{\mathbf{A}}, \mathbf{A})$ *and* $(\hat{\mathbf{A}}_b, \mathbf{A}_b)$ *be the couplings in Lemma* 4.2. *Then, there exist couplings for* $(\hat{\mathbf{X}}_\emptyset, X_0)$ *and* $(\hat{\mathbf{X}}, X)$ *constructed on the same probability space as* $(\hat{\mathbf{A}}, \mathbf{A})$ *and* $(\hat{\mathbf{A}}_b, \mathbf{A}_b)$, *such that* $\mathbb{E}_n\big[\rho(\hat{\mathbf{X}}_\emptyset, X_0)\big] \xrightarrow{P} 0$, $\rho(\hat{\mathbf{X}}_\emptyset, X_0) \xrightarrow{P} 0$, *and* $\rho(\hat{\mathbf{X}}, X) \xrightarrow{P} 0$ *as* $n \to \infty$.

*Proof.* For the two undirected models, CM and IR, write:

$$
\begin{aligned}
\hat{\mathbf{A}} &= (\hat{Y}, \hat{\mathbf{B}}), & \mathbf{A} &= (Y, \mathbf{B}), \\
\hat{\mathbf{A}}_b &= (\hat{Y}_b, \hat{\mathbf{B}}_b), & \mathbf{A}_b &= (Y_b, \mathbf{B}_b).
\end{aligned}
$$

For the two directed models, DCM and IRD, write:

$$
\begin{aligned}
\hat{\mathbf{A}} &= (\hat{Y}^-, \hat{Y}^+, \hat{\mathbf{B}}), & \mathbf{A} &= (Y^-, Y^+, \mathbf{B}), \\
\hat{\mathbf{A}}_b &= (\hat{Y}_b^-, \hat{Y}_b^+, \hat{\mathbf{B}}_b), & \mathbf{A}_b &= (Y_b^-, Y_b^+, \mathbf{B}_b).
\end{aligned}
$$

To obtain the statement of the lemma for the CM, simply set $(\hat{\mathbf{X}}_\emptyset, X_\emptyset) = (\hat{Y}, \hat{\mathbf{A}}, Y, A)$ and $(\hat{\mathbf{X}}_1, X_1) = (\hat{Y}_b, \hat{\mathbf{A}}_b, Y_b, A_b)$. Similarly, for the DCM set $(\hat{\mathbf{X}}_\emptyset, X_\emptyset) = (\hat{Y}^-, \hat{Y}^+, \hat{\mathbf{A}}, Y^-, Y^+, A)$ and $(\hat{\mathbf{X}}_1, X_1) = (\hat{Y}_b^-, \hat{Y}_b^+, \hat{\mathbf{A}}_b, Y_b^-, Y_b^+, A_b)$.

For the IR, construct $(\hat{S}, S) = (\Lambda_n(\hat{Y} \wedge b_n)/(\theta n), Y)$. Note that our assumptions imply that $\mathbb{E}_n[|\hat{S} - S|] \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. Now let $U \sim \mathrm{Uniform}[0, 1]$ be i.i.d. and independent of $(\hat{S}, S)$, and take $(\hat{\mathbf{X}}_\emptyset, X_0) = (G^{-1}(U; \hat{S}), \hat{\mathbf{A}}, G^{-1}(U; Y), A)$, where $G^{-1}(u; \lambda) = \sum_{m=0}^\infty m \mathbf{1}(G(m; \lambda) \leq u < G(m+1; \lambda))$ is the generalized inverse of the Poisson distribution function with mean $\lambda$. Note that since $G(m; \lambda)$ is decreasing in $\lambda$ for all $m \geq 0$, we have $\mathrm{Poi}(\lambda) \geq_{\mathrm{st}} \mathrm{Poi}(\mu)$ whenever $\lambda \geq \mu$, where $\geq_{\mathrm{st}}$ denotes the usual stochastic order and $\mathrm{Poi}(\alpha)$ denotes a Poisson random variable with mean $\alpha$. It follows that $\mathbb{E}[|G^{-1}(U; \lambda) - G^{-1}(U; \mu)|] = |\lambda - \mu|$, which in turn implies that $\mathbb{E}_n[\rho(\hat{\mathbf{X}}_\emptyset, X_0)] = \mathbb{E}_n[|\hat{S} - S| + \rho''(\hat{\mathbf{A}}, A)] \xrightarrow{\mathrm{P}} 0, n \to \infty$.

For the size-biased versions, set $(\hat{S}_b, S_b) = (\Lambda_n(\hat{Y}_b \wedge b_n)/(\theta n), Y_b)$, note that Lemma 4.2 gives $|\hat{S}_b - S_b| \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, and let $(\hat{\mathbf{X}}_1, X_1) = (G^{-1}(U; \hat{S}_b) + 1, \hat{\mathbf{A}}_b, G^{-1}(U; Y_b) + 1, A_b)$. Now use the continuity in $\lambda$ of $G^{-1}(u; \lambda)$ to obtain that $\rho(\hat{\mathbf{X}}_1, X_1) = |G^{-1}(U; \hat{S}_b) - G^{-1}(U; Y_b)| + \rho''(\hat{\mathbf{A}}_b, A_b) \xrightarrow{\mathrm{P}} 0, n \to \infty$.

The same steps also give the result for the IRD by setting

$$(\hat{S}^-, \hat{S}^+, S^-, S^+) = (\Lambda_n^+(\hat{Y}^- \wedge a_n)/(\theta n), \Lambda_n^-(\hat{Y}^+ \wedge b_n)/(\theta n), cY^-, (1-c)Y^+),$$
$$(\hat{S}_b^-, \hat{S}_b^+, S_b^-, S_b^+) = (\Lambda_n^+(\hat{Y}_b^- \wedge a_n)/(\theta n), \Lambda_n^-(\hat{Y}_b^+ \wedge b_n)/(\theta n), cY_b^-, (1-c)Y_b^+),$$

where $c = \mathbb{E}[W^+]/\mathbb{E}[W^- + W^+]$, and setting

$$(\hat{\mathbf{X}}_\emptyset, X_\emptyset) = (G^{-1}(U; \hat{S}^-), G^{-1}(U'; \hat{S}^+) + 1, \hat{\mathbf{A}}, G^{-1}(U; S^-), G^{-1}(U'; S^+) + 1, A),$$
$$(\hat{\mathbf{X}}_1, X_1) = (G^{-1}(U; \hat{S}_b^-), G^{-1}(U'; \hat{S}_b^+) + 1, \hat{\mathbf{A}}_b, G^{-1}(U'; S_b^-), G^{-1}(U; S_b^+) + 1, A_b)$$

for some $U, U'$ i.i.d. Uniform$[0, 1]$ and independent of $\mathscr{F}_n$; this completes the proof. $\square$

Finally, we can give the proof of Theorem 4.3.

*Proof of Theorem* 4.3. By Lemma 4.3 there exist couplings $(\hat{\mathbf{X}}_\emptyset, X_\emptyset)$ and $(\hat{\mathbf{X}}_1, X_1)$ such that $\mathbb{E}_n[\rho(\hat{\mathbf{X}}_\emptyset, X_\emptyset)] \xrightarrow{\mathrm{P}} 0$ and $\rho(\hat{\mathbf{X}}_1, X_1) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. Now let $\{(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) : \mathbf{i} \in \mathcal{U}, \mathbf{i} \neq \emptyset\}$ be i.i.d. copies of $(\hat{\mathbf{X}}_1, X_1)$, independent of $(\hat{\mathbf{X}}_\emptyset, X_\emptyset)$. Recall that $\hat{N}_\mathbf{i}$ ($\mathcal{N}_\mathbf{i}$) can be determined from the first coordinate of $\hat{\mathbf{X}}_\mathbf{i}$ ($X_\mathbf{i}$).

We will now use the sequence $\{(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) : \mathbf{i} \in \mathcal{U}\}$ to construct both $\hat{T}(\hat{\mathbf{A}})$ and $\mathcal{T}(A)$ by determining their nodes according to the recursions $\hat{A}_k = \{(\mathbf{i}, j) : \mathbf{i} \in \hat{A}_{k-1}, 1 \leq j \leq \hat{N}_\mathbf{i}\}$ and $\mathcal{A}_k = \{(\mathbf{i}, j) : \mathbf{i} \in \mathcal{A}_{k-1}, 1 \leq j \leq \mathcal{N}_\mathbf{i}\}$ for $k \geq 1$. Without loss of generality, assume that $0 < \varepsilon < 1$.

Now define the stopping time $\kappa(\varepsilon) = \inf\{k \geq 0 : \rho(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) > \varepsilon$ for some $\mathbf{i} \in \hat{A}_k\}$. Note that since $\hat{N}_\mathbf{i}$ and $\mathcal{N}_\mathbf{i}$ are integer-valued, $\rho(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) < 1$ implies that $\hat{N}_\mathbf{i} = \mathcal{N}_\mathbf{i}$. It follows that, for any $x_n \geq 1$,

$$\mathbb{P}_n(\hat{T}^{(k)} \simeq \mathcal{T}^{(k)}) \geq \mathbb{P}_n\left(\bigcap_{r=0}^k \bigcap_{\mathbf{i} \in \mathcal{A}_r} \{\rho(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) \leq \varepsilon\}, \hat{T}^{(k)} \simeq \mathcal{T}^{(k)}\right)$$
$$= \mathbb{P}_n(\kappa(\varepsilon) > k)$$
$$\geq 1 - \mathbb{P}_n(\kappa(\varepsilon) \leq k, |\mathcal{V}_k| \leq x_n) - \mathbb{P}_n(|\mathcal{V}_k| > x_n)$$
$$= 1 - \sum_{r=0}^k \mathbb{P}_n(\kappa(\varepsilon) = r, |\mathcal{V}_k| \leq x_n) - \mathbb{P}_n(|\mathcal{V}_k| > x_n),$$

where $\mathcal{V}_k = \bigcup_{r=0}^{k} \mathcal{A}_r$. To compute the last probabilities, note that $\mathbb{P}_n(\kappa(\varepsilon) = 0) \leq \varepsilon^{-1}\mathbb{E}_n[\rho(\hat{\mathbf{X}}_\emptyset, X_\emptyset)]$, and, for $r \geq 1$,

$$\mathbb{P}_n(\kappa(\varepsilon) = r, |\mathcal{V}_k| \leq x_n) \leq \mathbb{P}_n\left(\bigcup_{\mathbf{i}\in\mathcal{A}_r}\{\rho(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) > \varepsilon\}, |\mathcal{A}_r| \leq x_n\right)$$
$$\leq \mathbb{E}_n\left[\mathbf{1}(|\mathcal{A}_r| \leq x_n)\sum_{\mathbf{i}\in\mathcal{A}_r}\mathbf{1}(\rho(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) > \varepsilon)\right]$$
$$= \mathbb{E}_n[1(|\mathcal{A}_r| \leq x_n)|\mathcal{A}_r|]\mathbb{P}_n(\rho(\hat{\mathbf{X}}_1, X_1) > \varepsilon)$$
$$\leq x_n\mathbb{P}_n(\rho(\hat{\mathbf{X}}_1, X_1) > \varepsilon),$$

where in the third step we used the independence of $(\hat{\mathbf{X}}_1, X_1)$ from $\mathcal{A}_r$. It follows that, if we choose $x_n = \mathbb{P}_n(\rho(\hat{\mathbf{X}}_1, X_1) > \varepsilon)^{-1/2} \xrightarrow{P} \infty$, then

$$\mathbb{P}_n\left(\bigcap_{r=0}^{k}\bigcap_{\mathbf{i}\in\mathcal{A}_r}\{\rho(\hat{\mathbf{X}}_\mathbf{i}, X_\mathbf{i}) \leq \varepsilon\}, \hat{T}^{(k)} \simeq \mathcal{T}^{(k)}\right)$$
$$\geq 1 - \varepsilon^{-1}\mathbb{E}_n[\rho(\hat{\mathbf{X}}_\emptyset, X_\emptyset)] - kx_n\mathbb{P}_n(\rho(\hat{\mathbf{X}}_1, X_1) > \varepsilon) - \mathbb{P}_n(|\mathcal{V}_k| > x_n)$$
$$\geq 1 - \varepsilon^{-1}\mathbb{E}_n[\rho(\hat{\mathbf{X}}_\emptyset, X_\emptyset)] - kx_n^{-1/2} - \mathbb{P}_n(|\mathcal{V}_k| > x_n) \xrightarrow{P} 0$$

as $n \to \infty$. This completes the proof. $\qquad\square$

The last proof in the paper relates to the case $m \geq 2$ for both Theorems 2.1 and 3.1. Since the proof for the directed case follows exactly the same steps as for the undirected one, we include here only the undirected case.

*Proof of Theorem 2.1.* ($m \geq 2$). Start by sampling $\{I_j : 1 \leq j \leq m\}$ independently and uniformly in $V_n$. Without loss of generality we can assume that $I_1 \neq I_2 \neq \cdots \neq I_m$. Next, note that the couplings for $\mathcal{G}_i^{(k)}(\mathbf{a})$ with their corresponding intermediate trees can be done simultaneously for all $i \in V_n$, so let $\hat{T}_{\emptyset(i)}^{(k)}(\hat{\mathbf{A}})$ be the coupled tree for $\mathcal{G}_i^{(k)}(\mathbf{a})$. Note that the $\{\hat{T}_{\emptyset(I_j)}^{(k)}(\hat{\mathbf{A}}) : 1 \leq j \leq m\}$ are not independent of each other, but by Theorem 2.1 for the $m = 1$ case, they satisfy $\sum_{j=1}^{m}\mathbb{E}_n[\rho(\mathbf{X}_{I_j}, \hat{\mathbf{X}}_{\emptyset(I_j)})] \xrightarrow{P} 0$ and $\mathbb{P}_n(\bigcup_{j=1}^{m}\{\mathcal{G}_{I_j}^{(k)}(\mathbf{a}) \not\simeq \hat{T}_{\emptyset(I_j)}^{(k)}(\hat{\mathbf{A}})\}) \xrightarrow{P} 0$ as $n \to \infty$. We now explain how to construct a set of i.i.d. copies of $\hat{T}_{\emptyset(I_1)}^{(k)}(\hat{\mathbf{A}})$, denoted $\{\tilde{T}_{\emptyset(I_j)}^{(k)}(\tilde{\mathbf{A}}) : 1 \leq j \leq m\}$, satisfying

$$\mathbb{P}_n\left(\bigcup_{j=1}^{m}\{\hat{T}_{\emptyset(I_j)}^{(k)}(\hat{\mathbf{A}}) \not\simeq \tilde{T}_{\emptyset(I_j)}^{(k)}(\tilde{\mathbf{A}})\}\right) \xrightarrow{P} 0, \qquad n \to \infty. \tag{4.1}$$

To start, let $\tilde{T}_{\emptyset(I_1)}^{(k)}(\tilde{\mathbf{A}}) = \hat{T}_{\emptyset(I_1)}^{(k)}(\hat{\mathbf{A}})$. Next, note that the trees $\{\hat{T}_{\emptyset(I_j)}^{(k)}(\hat{\mathbf{A}}) : 1 \leq j \leq m\}$ are each (delayed) marked Galton–Watson processes whose roots have distribution $\mu_n^*(\cdot) = \mathbb{P}_n(\hat{\mathbf{X}}_\emptyset \in \cdot)$ and all other nodes have distribution $\mu_n(\cdot) = \mathbb{P}_n(\hat{\mathbf{X}}_1 \in \cdot)$. Since each full mark $\hat{\mathbf{X}}_\mathbf{i}$ contains the vector $\hat{\mathbf{A}}_\mathbf{i}$, we can see which vertex attributes have been sampled. Note that the possible vertex

attributes are $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$; they are such that $p_i^* = \mathbb{P}_n(\hat{\mathbf{A}}_\emptyset = \mathbf{a}_i) = 1/n$, and $p_i = \mathbb{P}_n(\hat{\mathbf{A}}_1 = \mathbf{a}_i)$ is either equal to $D_i/L_n$ in the CM or to $\bar{W}_i/\Lambda_n$ in the IR. We need to keep track of the labels $\{1, 2, \ldots, n\}$ of the vertex attributes $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$. To do this, let $S$ be the set of labels sampled in the construction of $\tilde{T}^{(k)}_{\emptyset(I_1)}(\tilde{\mathbf{A}})$, and then construct each of $\tilde{T}^{(k)}_{\emptyset(I_j)}(\tilde{\mathbf{A}})$, $2 \leq j \leq m$, in a breadth-first fashion according to the following rule:

- If a node in $\hat{T}^{(k)}_{\emptyset(I_j)}(\hat{\mathbf{A}})$ has a vertex attribute whose label is not in the set $S$, copy the node onto $\tilde{T}^{(k)}_{\emptyset(I_j)}(\tilde{\mathbf{A}})$ and add the new observed label to the set $S$.

- Otherwise, attach an independent copy of a Galton–Watson marked tree having full mark distribution $\mu_n$ where the repeated node would be.

Since, as long as we do not sample any vertex attributes from the set $S$, we will have $\tilde{T}^{(k)}_{\emptyset(I_j)}(\tilde{\mathbf{A}}) \simeq \hat{T}^{(k)}_{\emptyset(I_j)}(\hat{\mathbf{A}})$, for any $x_n > 0$ and $M_n$ equal to either $L_n$ for a CM or $\Lambda_n$ for an IR, we have

$$\mathbb{P}_n\left( \bigcup_{j=1}^{m} \{\hat{T}^{(k)}_{\emptyset(I_j)}(\hat{\mathbf{A}}) \not\simeq \tilde{T}^{(k)}_{\emptyset(I_j)}(\tilde{\mathbf{A}})\} \right)$$

$$\leq \mathbb{P}_n\left( \bigcup_{j=1}^{m} \{\hat{T}^{(k)}_{\emptyset(I_j)}(\hat{\mathbf{A}}) \not\simeq \tilde{T}^{(k)}_{\emptyset(I_j)}(\tilde{\mathbf{A}})\}, \sum_{j=1}^{m} |\hat{T}^{(k)}_{\emptyset(I_j)}| \leq m x_n, \sum_{i \in S} p_i \leq m x_n/M_n \right) \qquad (4.2)$$

$$+ \mathbb{P}_n\left( \left\{ \sum_{i \in S} p_i > m x_n/M_n \right\} \cup \left\{ \sum_{j=1}^{m} |\hat{T}^{(k)}_{\emptyset(I_j)}| > m x_n \right\} \right). \qquad (4.3)$$

Now note that since, in the first probability, $\sum_{j=1}^{m} |\hat{T}^{(k)}_{\emptyset(I_j)}| \leq m x_n$ and the chances of sampling a label from $S$ is at most $m x_n/M_n$, we have that (4.2) is bounded by $\mathbb{P}_n(\text{Bin}(m x_n, m x_n/M_n) \geq 1) \leq m^2 x_n^2/M_n$, where $\text{Bin}(n, p)$ is a binomial random variable with parameters $(n, p)$. To analyze (4.3), let $Y_{\mathbf{i}} = p_i M_n$ if $\hat{\mathbf{A}}_{\mathbf{i}} = \mathbf{a}_i$, and note that $\sum_{i \in S} p_i = \sum_{j=1}^{m} \sum_{\mathbf{i} \in \hat{T}^{(k)}_{\emptyset(I_j)}} Y_{\mathbf{i}}/M_n$, so by the union bound we obtain that (4.3) is bounded from above by

$$\sum_{j=1}^{m} \mathbb{P}_n\left( |\hat{T}^{(k)}_{\emptyset(I_j)}| \vee \sum_{\mathbf{i} \in \hat{T}^{(k)}_{\emptyset(I_j)}} Y_{\mathbf{i}} > x_n \right) = m \mathbb{P}_n\left( |\hat{T}^{(k)}| \vee \sum_{\mathbf{i} \in \hat{T}^{(k)}} Y_{\mathbf{i}} > x_n \right).$$

Now use Theorem 4.3 to obtain that, for any $\varepsilon > 0$,

$$\mathbb{P}_n\left( |\hat{T}^{(k)}| \vee \sum_{\mathbf{i} \in \hat{T}^{(k)}} Y_{\mathbf{i}} > x_n \right) \leq \mathbb{P}\left( |\mathcal{T}^{(k)}| \vee \sum_{\mathbf{i} \in \mathcal{T}^{(k)}} (\mathcal{Y}_{\mathbf{i}} + \varepsilon) > x_n \right) + o(1)$$

as $n \to \infty$, where $\mathcal{Y}_{\mathbf{i}}$ is equal to the second component of $X_{\mathbf{i}}$. Since $|\mathcal{T}^{(k)}| < \infty$ almost surely and does not depend on $\mathscr{F}_n$, and $1/M_n = O_{\mathrm{P}}(1/n)$ as $n \to \infty$, choosing $x_n = n^{1/2}/\log n$ proves (4.1).

Finally, use Theorem 4.3 applied to each of the $\{\tilde{T}^{(k)}_{\emptyset(I_j)}(\tilde{\mathbf{A}}) : 1 \leq j \leq m\}$ to obtain that there exists an i.i.d. set $\{\mathcal{T}^{(k)}_{\emptyset(I_j)}(A) : 1 \leq j \leq m\}$ having the same distribution as $\mathcal{T}^{(k)}(A)$ such that, for any $\varepsilon \in (0, 1)$, $\sum_{j=1}^{m} \mathbb{E}_n\big[\rho(\tilde{\mathbf{X}}_{\emptyset(I_j)}, X_{\emptyset(I_j)})\big] \xrightarrow{\text{P}} 0$ and

$$\mathbb{P}_n\left(\bigcap_{j=1}^{m}\left\{\bigcap_{\mathbf{i} \in \mathcal{T}^{(k)}_{\emptyset(I_j)}}\{\rho(\tilde{\mathbf{X}}_{\mathbf{i}}, X_{\mathbf{i}}) \leq \varepsilon\}, \ \tilde{T}^{(k)}_{\emptyset(I_j)} \simeq \mathcal{T}^{(k)}_{\emptyset(I_j)}\right\}\right) \xrightarrow{\text{P}} 1$$

as $n \to \infty$. This completes the proof. $\qquad\square$

## Acknowledgements

## Funding information

## Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

## References

[1] ALDOUS, D. AND LYONS, R. (2007). Processes on unimodular random networks. *Electron. J. Prob.* **12**, 1454–1508.

[2] ALDOUS, D. AND STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*, ed. H. Kesten. Springer, New York, pp. 1–72.

[3] BENJAMINI, I. AND SCHRAMM, O. (2011). Recurrence of distributional limits of finite planar graphs. In *Selected Works of Oded Schramm*, eds I. Benjamini and O. Häggström. Springer, New York, pp. 533–545.

[4] BOLLOBÁS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Europ. J. Combinatorics* **1**, 311–316.

[5] BOLLOBÁS, B. (2001). *Random Graphs*. Cambridge University Press.

[6] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31**, 3–122.

[7] BRITTON, T., DEIJFEN, M. AND MARTIN-LÖF, A. (2006). Generating simple random graphs with prescribed degree distribution. *J. Statist. Phys.* **124**, 1377–1397.

[8] CHAINTRON, L.-P. AND DIEZ, A. (2021). Propagation of chaos: A review of models, methods and applications. Preprint, arXiv:2106.14812.

[9] CHEN, N., LIVTAK, N. AND OLVERA-CRAVIOTO, M. (2017). Generalized PageRank on directed configuration networks. *Random Structures Algorithms* **51**, 237–274.

[10] CHEN, N. AND OLVERA-CRAVIOTO, M. (2013). Directed random graphs with given degree distributions. *Stoch. Syst.* **3**, 147–186.

[11] CHUNG, F. AND LU, L. (2002). Connected components in random graphs with given expected degree sequences. *Ann. Combinatorics* **6**, 125–145.

[12] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge University Press.

[13] ERDŐS, P. E. AND RÉNYI, A. (1959). On random graphs. *Publicationes Mathematicae (Debrecen)* **6**, 290–297.

[14] FRAIMAN, N., LIN, T. AND OLVERA-CRAVIOTO, M. (2021). Stochastic recursions on directed random graphs. Preprint, arXiv:2010.09596.

[15] GARAVAGLIA, A., VAN DER HOFSTAD, R. AND LITVAK, N. (2020). Local weak convergence for PageRank. *Ann. Appl. Prob.* **30**, 40–79.

[16] LACKER, D., RAMANAN, K. AND WU, R. (2020). Local weak convergence and propagation of ergodicity for sparse networks of interacting processes. Preprint, arXiv:1904.02585.

[17] LACKER, D., RAMANAN, K. AND WU, R. (2020). Marginal dynamics of interacting diffusions on unimodular galton-watson trees. Preprint, arXiv:2009.11667.

[18] LEE, J. AND OLVERA-CRAVIOTO, M. (2020). PageRank on inhomogeneous random digraphs. *Stoch. Process. Appl.* **130**, 1–57.

[19] NORROS, I. AND REITTU, H. (2006). On a conditionally Poissonian graph process. *Adv. Appl. Prob.* **38**, 59–75.

[20] OLVERA-CRAVIOTO, M. (2021). PageRank's behavior under degree correlations. *Ann. Appl. Prob.* **1**, 1403–1442.

[21] VAN DER HOFSTAD, R. (2017). *Random Graphs and Complex Networks*, Vol. 1. Cambridge University Press.

[22] VAN DER HOFSTAD, R. (2021). *Random Graphs and Complex Networks*, Vol. 2. Cambridge University Press.

[23] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. AND VAN MIEGHEM, P. (2005). Distances in random graphs with finite variance degrees. *Random Structures Algorithms* **27**, 76–123.

[24] VILLANI, C. (2009). *Optimal Transport, Old and New*. Springer, New York.