

LOCAL WEAK CONVERGENCE AND ITS APPLICATIONS

SAYAN BANERJEE, SHANKAR BHAMIDI, JIANAN SHEN, AND SETH PARKER YOUNG

ABSTRACT. Motivated in part by understanding average case analysis of fundamental algorithms in computer science, and in part by the wide array of network data available over the last decade, a variety of random graph models, with corresponding processes on these objects, have been proposed over the last few years. The main goal of this paper is to give an overview of local weak convergence, which has emerged as a major technique for understanding large network asymptotics for a wide array of functionals and models. As opposed to a survey, the main goal is to try to explain some of the major concepts and their use to junior researchers in the field and indicate potential resources for further reading.

1. INTRODUCTION

Starting graduate courses in probability cover myriad applications including weak convergence of **real-valued** random variables, which one often sees initially via convergence properties of cumulative distribution functions. However this notion has an equivalent definition via convergence of expectations of functions of the random variables. In many cases, only in a more advanced course does one see that this final definition easily extends to “completely general” metric spaces as beautifully articulated in [54] which then went on to play an important role in a host of applications of modern probability ranging from Billingsley’s classic [20] to optimal transport [9]. The goal of this specific article is to convey developments of this general theory to the study of asymptotics of large discrete random structures, leading to one specific approach now described as *local* weak convergence. The goal of this paper is not a survey length description of this approach (pointers to beautiful survey length treatments for further reading can be found in Section 6). Rather, taking a “leaf” from KRP’s interview [1]:

“Audience: *So according to you, the best way to learn a new subject is by giving seminars to each other?”*

KRP: *Yes, you can catch hold of somebody who can follow you. Then both of you should discuss and explain to each other, and the process itself will lead you to something new. Gelfand used to say, “Read the theorems but not their proofs.” You try to construct the proof but that may be too difficult for us. So, everyday learn at least one new lemma and then try to tell somebody about this lemma. While discussing, I think something new will always crop up but that is how new ideas are born. Studying by oneself is boring but if you have four people, and you discuss with each other, go together for tea and discuss again, then life is more interesting. Ranga Rao, Varadarajan, Varadhan and me, always explained things to each other.”*

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, 304 HANES HALL, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599

E-mail addresses: sayan, bhamidi, jshen, sethpar@email.unc.edu .

2010 Mathematics Subject Classification. Primary: 60C05, 05C80.

Key words and phrases. weak convergence on metric spaces, local weak convergence, continuous time branching processes, random trees, random graphs.

Our goal is to describe the power of “abstraction” related to weak convergence on metric spaces, pioneered in [54] in the settings of network models and their applications. The main goal is to convey the joy of mathematical discovery via using some canonical results. Envisioning an interested student with basic technical knowledge at the level of say a measure theoretic probability course, and interest in discrete random structures, our goal is to explain the motivations driving the mathematical development and (our) intuition behind this fundamental technique. Following Paul Halmos’s adage that “*The only way to learn mathematics is to do mathematics*”, we start with motivating questions, including those that have recently arisen in our own research, in Section 2. These questions then motivate the development of the theoretical foundations in Section 3. Next we convey the “awe” behind *some* of the canonical results in this area in Section 4. Since this paper is in no way a survey, as a stand alone paper can do little justice to the far reaching applications of this technique, we point the interested reader to further material in the Section 6.

2. MOTIVATING EXAMPLES AND QUESTIONS

We start by describing motivating questions in a number of different areas.

2.1. Dynamic social network models. Social networks now play an important role in society, for example in the diffusion of information across populations of individuals. This has motivated the study of network valued data where nodes and/or edges have attributes, which modulate the dynamics of both network evolution, and information flow on the network itself. This has spurred the development of probabilistic network models [32, 33, 42, 44] that incorporate three major ingredients: (a) heterogeneity in edge creation across groups; (b) dynamic (i.e. time dependent) network evolution and (c) popularity bias. Formulated models are then used to understand domain specific questions, including bias in network sampling, PageRank and degree centrality scores and their impact in network ranking and recommendation algorithms. One basic model is as follows. Fix $\gamma \in [0, 1]$. Vertices enter the system sequentially at discrete times $n \geq 1$ starting with a base connected graph $\tilde{\mathcal{G}}_0$ at time $n = 0$. Write v_n for the vertex that enters at time n and $a(v_n)$ for the corresponding attribute; every vertex v_n has attribute distribution

$$a(v_n) \sim \boldsymbol{\pi}, \quad \text{independent of } \{\tilde{\mathcal{G}}_s : 0 \leq s \leq n-1\}. \quad (2.1)$$

The dynamics of construction are recursively defined as: for any n and $v \in \tilde{\mathcal{G}}_n$, let $\deg(v, n)$ denote the degree of v at time n (if $\tilde{\mathcal{G}}_0 = \{v_0\}$, initialize $\deg(v_0, 0) = 1$). For $n \geq 0$, v_{n+1} attaches to the network via a single outgoing edge. This edge is created via sampling an existing vertex in $\tilde{\mathcal{G}}_n$ to attach to, with probabilities (conditionally on $\tilde{\mathcal{G}}_n$ and $a(v_{n+1})$) given by:

$$\mathbb{P}(v_{n+1} \rightsquigarrow v | \tilde{\mathcal{G}}_n, a(v_{n+1}) = a^*) = \frac{\kappa(a(v), a^*)[\deg(v, n)]^\gamma}{\sum_{v' \in \tilde{\mathcal{G}}_n} \kappa(a(v'), a^*)[\deg(v', n)]^\gamma}, \quad v \in \tilde{\mathcal{G}}_n. \quad (2.2)$$

Denote this model of evolving random networks by $\mathcal{P}(\gamma, \boldsymbol{\pi}, \kappa)$. The single attribute setting consists of a dynamic network model where new vertices enter the system and connect to existing vertices with probability proportional to their current degree, thus promulgating the “rich get richer phenomenon” (i.e. taking degree as a proxy for popularity, high degree vertices accumulate their advantage faster over time). This was first formulated in the networks community in [16], where they found that this simple mechanistic model gives rise to heavy tailed degree distribution (observed in real world systems); in this specific case, if one writes $N_k(n)$ for the number of vertices with degree k in the network of size n , then [16] found numerically, which was later proved rigorously in [22], that $N_k(n)/n \xrightarrow{\text{a.e.}} p_k$ as $n \rightarrow \infty$, where $p_k \sim C/k^\tau$ as $k \rightarrow \infty$ for some $\tau > 0$.

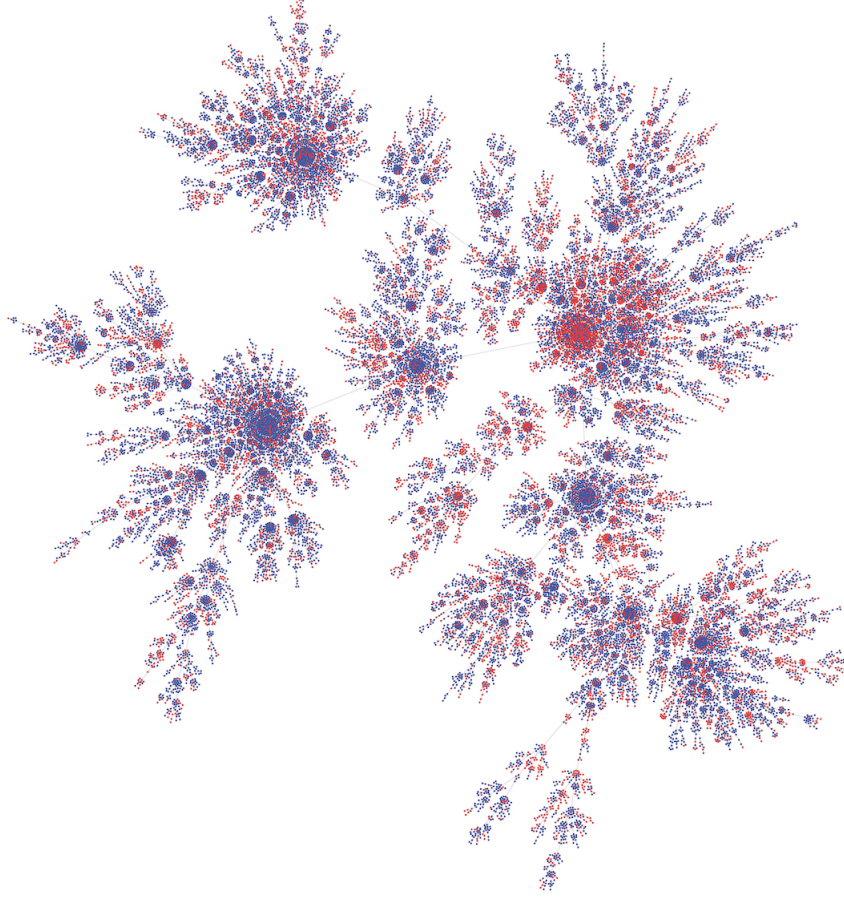


FIGURE 2.1. A simulation with two types ($\mathcal{S} = \{\text{red, blue}\}$) with $n = 30,000$ and $\kappa(a, a) = .75, \kappa(a, a') = .25$ for $a \neq a'$ and $\pi(\text{red}) = .35$ in the linear $\gamma \equiv 1$ case.

2.2. Google's PageRank. PageRank [52] is one of the most important measures of centrality of vertices in the system. Unlike degree centrality, where the popularity of a vertex is solely quantified by the number of its neighbors, PageRank centrality can also be enhanced by attaching to vertices having high PageRank (popularity enhanced by association with other popular members of a social network). Mathematically, this gives rise to the following definition.

View the networks under consideration as directed trees with edges pointing from offspring to parents.

Definition 2.1 (PageRank scores with damping factor c). *For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the PageRank scores of vertices $v \in \mathcal{V}$ with damping factor c is the stationary distribution $(\mathfrak{R}_{v,c} : v \in \mathcal{G})$ of the following random walk: at each step, with probability c , follow an outgoing edge (uniform amongst available choices) from the current location in the graph while, with probability $1 - c$, restart at a uniformly selected vertex in the entire graph. These scores are given by the linear system of equations:*

$$\mathfrak{R}_{v,c} = \frac{1-c}{n} + c \sum_{u \in \mathcal{N}^-(v)} \frac{\mathfrak{R}_{u,c}}{d^+(u)} \quad (2.3)$$

where $\mathcal{N}^-(v)$ is the set of vertices with edges pointed at v and $d^+(u)$ is the out-degree of vertex u .

Motivating question: Understand asymptotics for the empirical distribution of PageRank scores as $n \rightarrow \infty$; in particular does this distribution have a limit? What is the relationship, if any, between this limit distribution and the degree distribution? In the attributed network setting, does extremal behavior of the PageRank distribution vary across different attributes?

2.3. Gibbs distribution on graphs. Suppose $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ is the vertex and edge representation of a finite connected graph describing a social network. Further suppose, for a specific topic of interest, each individual $v \in \mathcal{G}$ has an opinion $x_v \in \{-1, 1\}$ (e.g. two types of political parties); more complex opinion spaces can be considered but we want to describe an important special case to fix ideas. Now in principle opinions do not necessarily vary independently, rather one would expect opinions of an individual to be influenced at least by its neighbors in the network. To model such situations, fix two constants $\beta \geq 0, B \in \mathbb{R}$ and consider the probability measure on opinion configurations $\Omega_{\mathcal{G}} := \{\mathbf{x} = (x_v : v \in \mathcal{G}), x_v \in \{-1, 1\} \ \forall v\}$ given by

$$\mu(\mathbf{x}) = \frac{1}{Z(\beta, B)} \exp \left(\beta \sum_{(u,v) \in E(\mathcal{G})} x_u x_v + B \sum_{u \in V} x_u \right) \quad (2.4)$$

Here:

- (a) $\beta \geq 0$ is sometimes referred to as the inverse temperature and describes the inherent influence of neighbors on one's own opinion; $\beta = 0$ implies no influence so everyone's opinions vary independently, while increasing β denotes increasing (positive or ferromagnetic, namely opinions of neighbors want to align with each other) influence of the neighbors.
- (b) B is sometimes referred to as the external magnetic field, representing baseline propensities of individuals wanting to adopt either of the two opinions; $B > 0$ represents inherent tendency towards $+1$, while $B < 0$ represents inherent tendency towards -1 .
- (c) $Z(\beta, B)$ is the normalizing factor that makes $\mu(\cdot)$ a probability measure (i.e. $\sum_{\mathbf{x} \in \Omega_{\mathcal{G}}} \mu(\mathbf{x}) = 1$).

Motivating question: Understand properties of the above probability measure. This is hard to do without any other assumptions on the graph, so consider a sequence of graphs $\{\mathcal{G}_n : n \geq 1\}$ with $\mathcal{G}_n$ a (random) network on n vertices generated according to some model that incorporates aspects of social networks. For each n consider the Gibbs measure in (2.4) on $\mathcal{G}_n$ with corresponding normalizing factor $Z_n(\beta, B)$. Understand asymptotics as $n \rightarrow \infty$.

2.4. Spectral distribution of adjacency matrices. Random matrix theory now has myriad applications in areas such as statistical physics [47], statistics and machine learning [63, 64]. In the context of networks, the adjacency matrix of a graph and its variants play an important role in downstream applications such as classifying nodes of different types (sometimes referred to as community detection) [2, 25, 45].

Motivating question: Consider a sequence of random network models $\{\mathcal{G}_n : n \geq 1\}$ and for $n \geq 1$, let $\mathbf{A}_n$ denote the adjacency matrix of $\mathcal{G}_n$ with corresponding eigen-values $\{\lambda_i^{(n)} : 1 \leq i \leq n\}$. Let $\mu_{\text{EMP}}^{(n)} = n^{-1} \sum_{i=1}^n \delta_{\lambda_i^{(n)}}$ denote the corresponding empirical distribution of eigen-values. Understand asymptotics of $\mu_{\text{EMP}}^{(n)}$ as $n \rightarrow \infty$.

2.5. Probabilistic combinatorial optimization. To motivate one stream of local weak convergence (geometric random graphs [5]) and also illustrate the breadth of this theory let us now discuss a completely different application. Imagine n faculty who have to be assigned n teaching time slots. Suppose the cost for assigning slot i to faculty j is given by $c_{i,j}$. For a fixed faculty i , the best assignment would be $\arg \min_{j \in [n]} c_{i,j}$, however rarely can we guarantee that everyone will get their best choice. Any specific choice of assignment of these slots to faculty can be represented by a permutation $\pi : [n] \rightarrow [n]$. Define the total cost of a permutation $C_n(\pi) = \sum_{i \in [n]} c_{i,\pi(i)}$. and let Π_n denote the space of all $n!$ permutations on $[n]$. The goal is to compute $\arg \min_{\pi \in \Pi_n} C_n(\pi)$. Solving

this optimization problem is a canonical example of “hardness” from a computational complexity perspective in the sense of worst case upper bounds for algorithms.

Motivating question: Coupled with computational complexity and algorithmic considerations that deal with worst case situations, there has been an enormous literature deal with *average case* properties of such questions (see [59] for a beautiful overview). In the context of the assignment problem consider the following: suppose $\{c_{i,j} : i, j \in [n]\}$ are *i.i.d.* positive random variables (below we will assume exponential rate 1). Can one understand asymptotics for the cost of the optimal assignment $A_n = \min_{\pi \in \Pi_n} C_n(\pi)$ as $n \rightarrow \infty$? After decades of work, in 1987 two statistical physicists Mezard and Parisi (the second of whom won the Nobel prize in Physics in 2022 in part for work related to such questions) via non-rigorous arguments in [49] conjectured that

$$\mathbb{E}(A_n) \rightarrow \zeta(2) = \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}, \quad \text{as } n \rightarrow \infty,$$

where $\zeta(\cdot)$ is the Riemann-Zeta function. Is this not amazing? Our goal is to give an idea of proof of how Aldous in [7] completed a rigorous proof of this result using local weak convergence techniques.

3. DEFINITIONS AND NOTATION

3.1. Weak convergence: General theory. In an initial treatment of weak convergence, typically when one specializes to sequences of *real valued* random variables $\{X_n : n \geq 1\} \subseteq \mathbb{R}$ (say with the goal of describing the Central Limit Theorem), when defining the convergence in distribution of $X_n \xrightarrow{w} X_\infty$ to a limit random variable, one often phrases weak convergence in terms of the (appropriate) convergence of the cumulative distribution functions (cdfs) $F_{X_n}(x) \rightarrow F_{X_\infty}(x)$ for all continuity points x of the limit cdf.

When dealing with sequences $\{X_n : n \geq 1\} \subseteq \mathcal{S}$ in a general metric space, the above definition does not extend. However the beautiful theory laid by giants in the field of probability in the early part of the last century (such as Prohorov, Skorohod, the “Indian school” including Varadarajan, Ranga Rao, Varadhan and KRP) and described so beautifully in the classics including KRP’s [54] and Billingsley’s [20] describe how one can systematically study weak convergence on general (typically Polish) metric space $\mathcal{S}$. Let $\mathcal{B}(\mathcal{S})$ denote the associated Borel sigma-field (i.e. generated by open sets in $\mathcal{S}$) and let $\{\mathbb{P}_n : n \geq 1\}$ be a sequence of probability measures on the measure space $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ and let $\mathbb{P}_\infty$ be another probability measure on the same measure space.

Definition 3.1. Let $C_b(\mathcal{S})$ denote the space of all continuous and bounded functions on $\mathcal{S}$. Say that $\mathbb{P}_n \xrightarrow{w} \mathbb{P}_\infty$ if for all $f \in C_b(\mathcal{S})$,

$$\int_{\mathcal{S}} f(s) d\mathbb{P}_n(s) \rightarrow \int_{\mathcal{S}} f(s) d\mathbb{P}_\infty(s), \quad \text{as } n \rightarrow \infty.$$

3.2. Local weak convergence (LWC): Intuition and general framework. As described in the introduction and motivation, the last decade has witnessed an explosion in the formulation and use of discrete random structures to understand various real world phenomenon. For probabilists, the natural question is to understand asymptotics for such models as the system size $\uparrow \infty$. To fix ideas first consider the following classical network model.

Definition 3.2 (Erdős-Rényi random graph model). Fix $\lambda > 0$, $n \geq 1$ and vertex set $[n] := \{1, 2, \dots, n\}$. Say that a random graph with vertex set $[n]$ has distribution $\mathcal{G}_n \sim \text{ERRG}(n, \lambda/n)$ if it is generated as follows: place each of the possible $\binom{n}{2}$ edges with probability λ/n , independently across edges.

Now suppose we want to understand asymptotics of this random graph model as $n \rightarrow \infty$. Before turning to a rigorous setup let us intuitively try to understand this model. A moment's thought reveals that it is probably hopeless to expect convergence of the entire graph uniformly to a limiting object because there are too many vertices and edges to keep track of. So let us simplify the question: suppose we sample a vertex $V_n \in [n]$ uniformly at random. What does the geometry of this random graph around this vertex look like? Let us start exploring the graph in a depth first manner around this vertex. The number of friends this vertex has $\mathcal{N}_n(V_n) \stackrel{d}{=} \text{Bin}(n-1, \lambda/n)$ and by Poisson approximation to the Binomial we get $\mathcal{N}(V_n) \approx \text{Pois}(\lambda)$. The reader should convince themselves that, as we continue the depth first exploration from this vertex, at least “initially” in the exploration, the number of new friends of each explored vertex has approximately $\text{Pois}(\lambda)$, more or less independent of the previous steps of the explorations; none of these phrases are rigorous but the above mental picture seems to suggest the following (with the key phrases for the reader in bold):

Guess 1. **Asymptotically** $n \rightarrow \infty$, the **local** geometry around a **randomly sampled** vertex should look like a Galton-Watson branching process with offspring distribution $\text{Pois}(\lambda)$. Since such a process survives iff $\lambda > 1$, this also suggests that there is a **large** connected component iff $\lambda > 1$.

Thus the natural question: how does one setup the mathematical architecture to make the above guess both rigorous and amenable to answering questions such as the second line? This is the goal of this Section (largely following [62, Chapter 2]) where we will focus on unweighted graphs. To help the reader, we will discuss special cases of trees and Geometric graphs (i.e. the general edge weighted setting) later. We will first need some notation. Let $\mathbb{G}$ denote the space of all locally finite graphs (i.e either finite or infinite graphs such that every vertex has finite, not necessarily uniformly bounded, degree); any such graph can be specified by its vertex set and edge set $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$. Further, with each edge assumed to have length one, $\mathcal{G}$ can be viewed as a metric space on $V(\mathcal{G})$ with the usual graph distance $d_{\mathcal{G}}(\cdot, \cdot)$. For $u \in \mathcal{G}$ and $r \geq 0$ we let $B_{\mathcal{G}}(u, r)$ denote the ball of radius r about this vertex. Say that two elements $\mathcal{G}_1, \mathcal{G}_2 \in \mathbb{G}$ are *isomorphic* if there exists (at least one) a bijection $\phi: V(\mathcal{G}_1) \rightarrow V(\mathcal{G}_2)$ which preserves edges, i.e. $\{u, v\} \in E(\mathcal{G}_1)$ iff $\{\phi(u), \phi(v)\} \in E(\mathcal{G}_2)$. Write this as $\mathcal{G}_1 \cong \mathcal{G}_2$.

A **rooted** graph is a graph with a special vertex o , namely a pair $(\mathcal{G}, o)$ where $\mathcal{G} \in \mathbb{G}$ and $o \in V(\mathcal{G})$; we will sometimes refer to this as the graph $\mathcal{G}$ *rooted at* o . Let $\mathbb{G}_{\star}$ denote the space of all locally finite rooted graphs. Note that for any $r \geq 0$, we can view the ball of radius r about the root o , $B_{\mathcal{G}}(o, r)$ as an element of $\mathbb{G}_{\star}$, rooted at o . For two elements $\mathcal{G}_{i,\star} = (\mathcal{G}_i, o_i) \in \mathbb{G}_{\star}$ for $i = 1, 2$ say that $\mathcal{G}_{1,\star} \cong \mathcal{G}_{2,\star}$ iff there exists an isomorphism ϕ that preserves the root (i.e. $\phi(o_1) = \phi(o_2)$).

Definition 3.3 ($\mathbb{G}_{\star}$ as a metric space). *For two rooted graphs $\{\mathcal{G}_{i,\star} = (\mathcal{G}_i, o_i) : i = 1, 2\}$ define the distance*

$$d_{\mathbb{G}_{\star}}(\mathcal{G}_{1,\star}, \mathcal{G}_{2,\star}) := \frac{1}{1 + R^*}, \quad R^* = \sup \{r : B_{\mathcal{G}_{1,\star}}(o_1, r) \cong B_{\mathcal{G}_{2,\star}}(o_2, r)\}, \quad (3.1)$$

where the relation $\cong$ denotes root preserving isomorphisms. For the rest of the paper, view $\mathbb{G}_{\star}$ as a metric space with distance $d_{\mathbb{G}_{\star}}(\cdot, \cdot)$.

Two minor comments are in order: **(a)** If one was being pedantic, we would now start thinking about $\mathbb{G}_{\star}$ as equivalence classes of graphs but this leads to far too much overhead. **(b)** It turns out that the above metric makes $\mathbb{G}_{\star}$ a Polish space. Let $\mathcal{B}(\mathbb{G}_{\star})$ denote the Borel measure space on $\mathbb{G}_{\star}$ generated by this metric. We are now in a position to talk about weak convergence on this space!

Definition 3.4 (Local weak convergence and the standard construction). *(a) Let $\{\mathbb{P}_n : n \geq 1\}$ be a sequence of probability measures on $(\mathbb{G}_{\star}, \mathcal{B}(\mathbb{G}_{\star}))$ and $\mathbb{P}_{\infty}$ be another probability measure on the same measure space. If $\mathbb{P}_n \xrightarrow{w} \mathbb{P}_{\infty}$ as $n \rightarrow \infty$ then say that this sequence of measures converges in the local weak convergence sense to $\mathbb{P}_{\infty}$ and denote this by $\mathbb{P}_n \xrightarrow{\text{LWC}} \mathbb{P}_{\infty}$.*

- (b) Suppose $\mathcal{G}$ is a (potentially random element) in $\mathbb{G}_{<\infty}$ (namely the space of unrooted locally finite graphs) each of finite size. Let V be a vertex selected uniformly at random (and independent of any randomization in the construction of $\mathcal{G}$) from $\mathcal{G}$. Let $\mathcal{G}[V]$ denote the connected component of V rooted at V . This operation of converting an unrooted graph in $\mathbb{G}_{<\infty}$ to a random element of the space of rooted graphs $\mathbb{G}_\star$ is sometimes referred to as the **standard construction** [5].
- (c) Consider a sequence of (potentially random) graphs $\{\mathcal{G}_n : n \geq 1\} \subseteq \mathbb{G}_{<\infty}$ and for each n let $\mathcal{G}_n(V_n) \in \mathbb{G}_\star$ be the standard construction as in (b) with V_n uniformly selected at random from $\mathcal{G}_n$ and let $\mathbb{P}_n$ be the corresponding probability distribution of $\mathcal{G}_n[V_n]$. Let $\mathbb{P}_\infty$ be another probability measure on $\mathbb{G}_\star$ and let $\mathcal{G}_{\infty,\star} \sim \mathbb{P}_\infty$. Say that $\mathcal{G}_n \xrightarrow{\text{LWC}} \mathcal{G}_{\infty,\star}$ if $\mathbb{P}_n \xrightarrow{w} \mathbb{P}_\infty$ as $n \rightarrow \infty$.

A natural question at this stage is: if this is just weak convergence in a specific setting, what is the rationale for associating a new name to this convergence? Quoting [5]:

“The topology on the metric space $\mathbb{G}_\star$ turns out to give weak convergence of probability measures on $\mathbb{G}_\star$ a local character that sharply differs from the traditional weak convergence such as one finds in the weak convergence of scaled random walk to Brownian motion. Weak convergence of measures on $\mathbb{G}_\star$ never involves any rescaling, and the special role of the neighborhoods $B_{\mathcal{G}}(o, \cdot)$ means that convergence in $\mathbb{G}_\star$ only informs us about behavior in the neighborhood of the root.”

The power of this methodology and reason for its widespread use is that, despite this local nature of convergence, in a host of examples, a careful analysis can leverage this to the convergence of global functionals. However before proceeding, let us give a more concrete description of what this convergence entails which is the goal of the next Section.



FIGURE 3.1. Generated via imgflip.com

3.3. Putting the “local” in LWC: lifting convergence to the space of probability measures. This Section perhaps best illustrates the power of weak convergence and the general framework developed in KRP’s work on weak convergence and entails unraveling the implications hidden in Definition 3.7. Let,

$$\mathcal{M}(\mathbb{G}_\star) = \{\mu : \mu \text{ finite measure on } (\mathbb{G}_\star, \mathcal{B}(\mathbb{G}_\star))\}. \quad (3.2)$$

It turns out, since $\mathbb{G}_\star$ is a Polish space, one can metrize the above space to also make it a Polish space (foundational treatments of such issues can be found in [43] or [29, Appendix 2, Vol 1]). Thus now one can talk about convergence of random objects $\{\mathfrak{P}_n : n \geq 1\} \subseteq \mathcal{M}(\mathbb{G}_\star)$!

Now for a finite unrooted graph $\mathcal{G} \in \mathbb{G}_{<\infty}$ integer $k \geq 1$ and vertex $v \in \mathcal{G}$, let $B_{\mathcal{G}}(v, k) \in \mathbb{G}_\star$ denote the ball of radius k around v in $\mathcal{G}$, viewed as an element in $\mathbb{G}_\star$ and rooted at v . Consider the map:

$$\mathfrak{P}_k : \mathbb{G}_{<\infty} \rightarrow \mathcal{M}(\mathbb{G}_\star) \quad \text{given by} \quad \mathcal{G} \rightsquigarrow \frac{1}{|\mathcal{G}|} \sum_{v \in \mathcal{G}} \delta\{B_{\mathcal{G}}(v, k)\}. \quad (3.3)$$

If $\mathbb{P}_\infty$ is a probability measure on $(\mathbb{G}_\star, \mathcal{B}(\mathbb{G}_\star))$, then for $\mathcal{G}_{\infty,\star} \sim \mathbb{P}_\infty$, write $\mathfrak{P}_{\mathcal{G}_{\infty,\star},k} = \mathcal{L}(B_{\mathcal{G}_{\infty,\star}}(o, k))$ for the law of the neighborhood upto distance k from the root o of $\mathcal{G}_{\infty,\star}$.

Theorem 3.5 ([62, Theorem 2.7]). *Let $\{\mathcal{G}_n : n \geq 1\}$ be a sequence of (potentially random) elements in $\mathbb{G}_{<\infty}$ and let $\mathbb{P}_\infty$ be a probability measure on $(\mathbb{G}_\star, \mathcal{B}(\mathbb{G}_\star))$ and let $\mathcal{G}_{\infty,\star} \sim \mathbb{P}_\infty$. Then $\mathcal{G}_n \xrightarrow{\text{LWC}} \mathcal{G}_{\infty,\star}$ iff for every $k \geq 1$ the sequence of (random) empirical neighborhood measures satisfy $\mathfrak{P}_k(\mathcal{G}_n) \xrightarrow{w} \mathfrak{P}_{\mathcal{G}_{\infty,\star},k}$ as $n \rightarrow \infty$.*

Thus local weak convergence is equivalent to convergence in distribution of local asymptotics of neighborhoods of randomly sampled vertices. In particular, at first sight, local weak convergence should guarantee convergence of local functionals (e.g. empirical distribution of degrees, empirical distribution of number of vertices at distance two etc) but **not** necessarily global functionals (size of the maximal connected component in the graph etc).

If we wanted to establish convergence of local statistics in a stronger sense than in distribution (say in probability), the local weak convergence criterion needs to be appropriately strengthened as follows.

Definition 3.6 (Local weak convergence in the probability sense). *Let $\{\mathcal{G}_n : n \geq 1\}$ be a sequence of (potentially random) elements in $\mathbb{G}_{<\infty}$ and let $\mathbb{P}_\infty$ be a probability measure on $(\mathbb{G}_\star, \mathcal{B}(\mathbb{G}_\star))$ and let $\mathcal{G}_{\infty,\star} \sim \mathbb{P}_\infty$. Then say that $\mathcal{G}_n \xrightarrow{\text{P-LWC}} \mathcal{G}_{\infty,\star}$ written as $\mathcal{G}_n$ converge in the probability fringe sense to $\mathcal{G}_{\infty,\star}$ iff for every $k \geq 1$ the sequence of empirical neighborhood measures satisfy the convergence in probability relationship $\mathfrak{P}_k(\mathcal{G}_n) \xrightarrow{\text{P}} \mathfrak{P}_{\mathcal{G}_{\infty,\star},k}$ as $n \rightarrow \infty$.*

Let us give an example where one has local weak convergence but not convergence in probability. Consider the following sequence of network models $\{\mathcal{G}_n : n \geq 1\}$ generated as follows:

- (a) For fixed n flip a biased coin which with probability $1/3$ comes out Heads and $2/3$ Tails.
- (b) If the coin is Head, generate an Erdős-Rényi random graph $\mathcal{G}_n \sim \text{ERRG}(n, 4/n)$; if it comes out tails let $\mathcal{G}_n \sim \text{ERRG}(n, 10/n)$.

Then a reader can convince themselves that, writing $\text{BP}(\text{Pois}(\lambda))$ for the genealogy tree generated by Poisson Galton-Watson branching process with mean λ started with a single root, that

$$\mathcal{G}_n \xrightarrow{\text{LWC}} \frac{1}{3} \text{BP}(\text{Pois}(4)) + \frac{2}{3} \text{BP}(\text{Pois}(10)),$$

however this convergence **cannot** be strengthened to convergence in probability. A similar notion is analogously defined for local weak convergence in the almost sure sense. We skip the details.

3.4. Special case: Local weak convergence of trees. Here we will specialize the phenomenon of local weak convergence to the setting where the sequence of models are rooted trees. There are two main reasons for this:

- (a) This will allow the readers to gain more traction on this notion of convergence in a concrete setting which is easier to visualize (see the figures below) various concepts, than in the context of general graphs.
- (b) More importantly: rooted trees have a notion of direction via orienting the geometry using the root and another specified vertex. When this specified vertex V_n is chosen uniformly at random, it is possible, with some amount of work, to show that for many models of random trees, the subtree **below** this chosen vertex (namely vertices for which the unique path from the root has to pass through this vertex), sometimes called the *fringe distribution*, converges weakly to a distribution on the space of finite trees. What is truly amazing is that, under very general conditions, Aldous in [3] showed that this implies local weak convergence of the tree itself (convergence of entire neighborhoods of the uniformly chosen vertex, not just the subtree below it) to a limiting infinite object! Thus this phenomenon gives a tractable tool to prove local weak convergence as all it requires is to understand what happens “below” a randomly selected vertex. This is the main tool we will use in Section 5.

We will first need some notation. For $n \geq 1$, let $\mathbb{T}_n$ be the space of all rooted trees on n vertices. Let $\mathbb{T} = \bigcup_{n=0}^{\infty} \mathbb{T}_n$ be the space of all finite rooted marked trees. Here $\mathbb{T}_0 = \emptyset$ will be used to represent the empty tree (tree on zero vertices). Let $\rho_{\mathbf{t}}$ denote the root of $\mathbf{t}$. For any $r \geq 0$ and $\mathbf{t} \in \mathbb{T}$, let

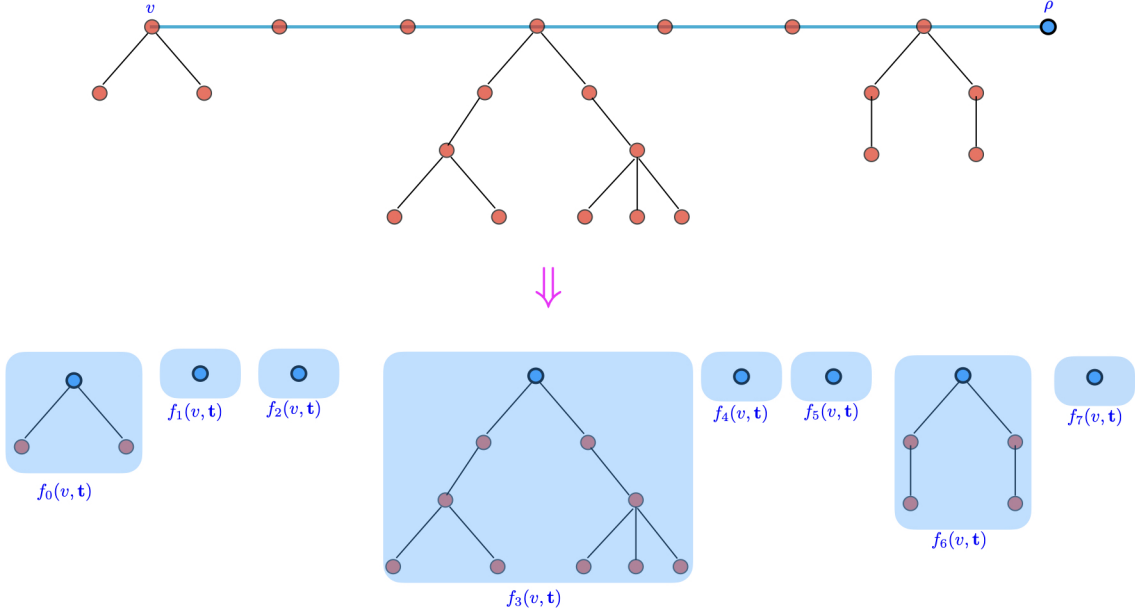


FIGURE 3.2. Fringe decomposition around vertex v of a finite tree rooted at ρ . Here the blue colors represent roots of the respective trees.

$B(\mathbf{t}, r) \in \mathbb{T}_{\mathcal{S}}$ denote the subgraph of $\mathbf{t}$ of vertices within graph distance r from $\rho_{\mathbf{t}}$, viewed as an element of $\mathbb{T}$ and rooted again at $\rho_{\mathbf{t}}$.

Given two rooted finite trees $\mathbf{s}, \mathbf{t} \in \mathbb{T}$, say that $\mathbf{s} \simeq \mathbf{t}$ if, there exists a **root preserving** isomorphism between the two trees viewed as unlabelled graphs. Given two rooted trees $\mathbf{t}, \mathbf{s} \in \mathbb{T}$ (see [62, Equation 2.3.15]), define the distance

$$d_{\mathbb{T}_{\mathcal{S}}}(\mathbf{t}, \mathbf{s}) := \frac{1}{1 + R^*} \quad (3.4)$$

where

$$R^* = \sup\{r : B(\mathbf{t}, r) \simeq B(\mathbf{s}, r), \text{ and } \exists \text{ isomorphism } \mathcal{F}_r \text{ between } B(\mathbf{t}, r) \text{ and } B(\mathbf{s}, r)\}.$$

Next, fix a tree $\mathbf{t} \in \mathbb{T}$ with root $\rho = \rho_{\mathbf{t}}$ and a vertex $v \in \mathbf{t}$ at (graph) distance h from the root. Let $(v_0 = v, v_1, v_2, \dots, v_h = \rho)$ be the unique path from v to ρ . The tree $\mathbf{t}$ can be decomposed as $h + 1$ rooted trees $f_0(v, \mathbf{t}), \dots, f_h(v, \mathbf{t})$, where $f_0(v, \mathbf{t})$ is the tree rooted at v , consisting of all vertices for which there exists a path from the root passing through v , and for $i \geq 1$, $f_i(v, \mathbf{t})$ is the subtree rooted at v_i , consisting of all vertices for which the path from the root passes through v_i but not through v_{i-1} . Call the map $(v, \mathbf{t}) \leadsto \mathbb{T}^{\infty}$ where $v \in \mathbf{t}$, defined via,

$$F(v, \mathbf{t}) = (f_0(v, \mathbf{t}), f_1(v, \mathbf{t}), \dots, f_h(v, \mathbf{t}), \emptyset, \emptyset, \dots),$$

as the fringe decomposition of $\mathbf{t}$ about the vertex v . Call $f_0(v, \mathbf{t})$ the **fringe** of the tree $\mathbf{t}$ at v . For $k \geq 0$, call $F_k(v, \mathbf{t}) = (f_0(v, \mathbf{t}), \dots, f_k(v, \mathbf{t}))$ the **extended fringe** of the tree $\mathbf{t}$ at v truncated at distance k from v on the path to the root (see Figure 3.2). Now consider the space $\mathbb{T}^{\infty}$. The metric in (3.4) can be extended in a straightforward fashion to $\mathbb{T}^{\infty}$.

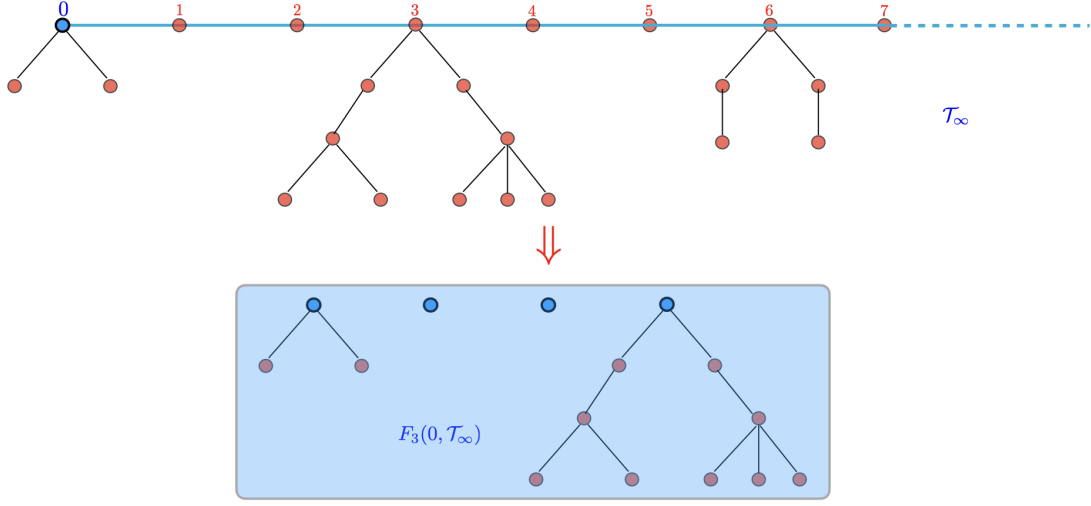


FIGURE 3.3. A sin-tree $\mathcal{T}_\infty$, namely a tree rooted at 0 with a single infinite path to infinity, and the corresponding extended fringe $F_3(0, \mathcal{T}_\infty)$ upto level three about 0.

Next, an element $\omega = (t_0, t_1, \dots) \in \mathbb{T}^\infty$, with $|t_i| \geq 1$ for all $i \geq 0$, can be thought of as a locally finite infinite rooted tree with a **single** path to **infinity** (thus called a **sin-tree** [3]), as follows: identify the sequence of roots of $\{t_i : i \geq 0\}$ with the integer lattice $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, equipped with the natural nearest neighbor edge set, rooted at $\rho = 0$ (see Figure 3.3). Analogous to the definition of extended fringes for finite trees, for any $k \geq 0$ write $F_k(0, \omega) = (t_0, t_1, \dots, t_k)$. Call this the extended fringe of the tree ω at vertex 0, till distance k , on the infinite path from 0. Call $t_0 = F_0(0, \omega)$ the **fringe** of the sin-tree ω . Now suppose $\mathbb{P}$ is a probability measure on $\mathbb{T}^\infty$ such that, for $\mathcal{T} := (t_0(\mathcal{T}), t_1(\mathcal{T}), \dots) \sim \mathbb{P}$, $|t_i(\mathcal{T})| \geq 1$ almost surely (a.s.) $\forall i \geq 0$. Then $\mathcal{T}$ can be thought of as an infinite **random sin-tree**.

Define a matrix $\mathbf{Q} = (\mathbf{Q}(\mathbf{s}, \mathbf{t}) : \mathbf{s}, \mathbf{t} \in \mathbb{T})$ via the following prescription: suppose the root ρ_s in $\mathbf{s}$ has degree $\deg(\rho_s) \geq 1$, and let $(v_1, \dots, v_{\deg(\rho_s)})$ denote its children. For $1 \leq i \leq \deg(\rho_s)$, let $\mathfrak{T}(\mathbf{s}, v_i)$ be the subtree below v_i and rooted at v_i , viewed as an element of $\mathbb{T}$.

Write,

$Q(\mathbf{s}, \mathbf{t}) = 2$

$$\mathbf{Q}(\mathbf{s}, \mathbf{t}) := \sum_{i=1}^{\deg(\rho_s)} \mathbb{1}_{\{d_{\mathbb{T}_\infty}(\mathfrak{T}(\mathbf{s}, v_i), \mathbf{t}) = 0\}}. \quad (3.5)$$

See the figure on the left for an example. In words, $Q(\mathbf{s}, \mathbf{t})$ counts the number of descendant subtrees of the root of $\mathbf{s}$ that are isomorphic to $\mathbf{t}$. If $\deg(\rho_s) = 0$, define $Q(\mathbf{s}, \mathbf{t}) = 0$.

Now consider a sequence,

$$(\bar{\mathbf{t}}_0, \bar{\mathbf{t}}_1, \dots) \subseteq \mathbb{T} \text{ such that } Q(\bar{\mathbf{t}}_i, \bar{\mathbf{t}}_{i-1}) \geq 1 \quad \forall i \geq 1. \quad (3.6)$$

Then there exists a unique sin-tree $\mathcal{T}$ with infinite path indexed by $\mathbb{Z}_+$ such that $\bar{\mathbf{t}}_i$ is the subtree rooted at i for all $i \in \mathbb{Z}_+$. Conversely, it is easy to see, by taking $\bar{\mathbf{t}}_i$ to be the union of (vertices and induced edges) of $t_0, \dots, t_i$ for each $i \in \mathbb{Z}_+$, that every infinite sin-tree has a representation of the form (3.6). Following [3], we call this the *monotone representation* of the sin-tree $\mathcal{T}$.

3.4.1. *Convergence on the space of trees.* Now for any $1 \leq k \leq \infty$, let $\mathcal{M}_{\mathbb{P}}(\mathbb{T}^k)$ denote the space of probability measures on the associated space, metrized using the topology of weak convergence inherited from the corresponding metric on the space $\mathbb{T}^k$, see e.g. [20]. Suppose $\{\mathcal{T}_n\}_{n \geq 1} \subseteq \mathbb{T}$ be a sequence of **finite** rooted random trees on some common probability space (for notational convenience, assume $|\mathcal{T}_n| = n$, all one needs is $|\mathcal{T}_n| \xrightarrow{\text{a.e.}} \infty$). For $n \geq 1$ and for each fixed $k \geq 0$, consider the empirical distribution of fringes up to distance k ($\delta\{\cdot\}$ below represents Dirac mass):

$$\mathfrak{P}_n^k := \frac{1}{n} \sum_{v \in \mathcal{T}_n} \delta\{F_k(v, \mathcal{T}_n)\}. \quad (3.7)$$

Thus $\{\mathfrak{P}_n^k : n \geq 1\}$ can be viewed as random sequence in $\mathcal{M}_{\mathbb{P}}(\mathbb{T}^k)$, with accompanying notions of almost sure convergence and convergence in distribution.

Definition 3.7 (Local weak convergence). *Consider two notions of convergence of $\{\mathcal{T}_n : n \geq 1\}$:*

- (a) *Fix a probability measure ω on $\mathbb{T}$. Say that a sequence of trees $\{\mathcal{T}_n\}_{n \geq 1}$ converges almost surely, in the fringe sense, to ω , if*

$$\mathfrak{P}_n^0 := \frac{1}{n} \sum_{v \in \mathcal{T}_n} \delta\{F_0(v, \mathcal{T}_n)\} \xrightarrow{\text{a.e.}} \omega, \quad \text{as } n \rightarrow \infty.$$

Denote this convergence by $\mathcal{T}_n \xrightarrow{\text{a.s.-fr}} \omega$ as $n \rightarrow \infty$.

- (b) *Say that a sequence of trees $\{\mathcal{T}_n\}_{n \geq 1}$ converges almost surely, in the **extended fringe sense**, to a limiting infinite random sin-tree $\mathcal{T}_\infty$ if for all $k \geq 0$ one has*

$$\mathfrak{P}_n^k \xrightarrow{\text{a.s.}} \mathbb{P}(F_k(0, \mathcal{T}_\infty) \in \cdot), \quad \text{as } n \rightarrow \infty.$$

Denote this convergence by $\mathcal{T}_n \xrightarrow{\text{a.s.-efr}} \mathcal{T}_\infty$ as $n \rightarrow \infty$.

Intuitively, fringe convergence, namely (a) above implies that, if we look at the subtree below a *typical* (i.e. selected uniformly at random) vertex, then the distribution of the corresponding random tree converges in distribution as the network size $n \rightarrow \infty$. Extended fringe convergence implies not just the structure of the neighborhood **below** typical vertices, but the entire local neighborhood, within any finite distance converges. Next note that in the setting of the convergence of (b) above, let $\omega_\infty(\cdot) = \mathbb{P}(F_0(0, \mathcal{T}_\infty) \in \cdot)$ denote the distribution of the fringe of $\mathcal{T}_\infty$ on $\mathbb{T}_{\mathcal{S}}$. Convergence in (b) above clearly implies convergence in notion (a) with $\omega = \omega_\infty(\cdot)$. More surprisingly, if the limiting distribution ω in (a) has a certain ‘stationarity’ property (defined next), *convergence in the fringe sense implies convergence in the extended fringe sense.*

Definition 3.8 (Fringe distribution [3]). *Say that a probability measure ω on $\mathbb{T}$ is a fringe distribution if*

$$\sum_{\mathbf{s}} \omega(\mathbf{s}) \mathbf{Q}(\mathbf{s}, \mathbf{t}) = \omega(\mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{T}.$$

For any fringe distribution ω on $\mathbb{T}_{\mathcal{S}}$, one can uniquely obtain the law ω^{EF} of a random sin-tree $\mathcal{T}$ with monotone decomposition $(\bar{\mathbf{t}}_0(\mathcal{T}), \bar{\mathbf{t}}_1(\mathcal{T}), \dots)$ such that for any $i \in \mathbb{Z}_+$, any $\bar{\mathbf{t}}_0, \bar{\mathbf{t}}_1, \dots$ in $\mathbb{T}_{\mathcal{S}}$,

$$\omega^{EF}((\bar{\mathbf{t}}_0(\mathcal{T}), \bar{\mathbf{t}}_1(\mathcal{T}), \dots, \bar{\mathbf{t}}_i(\mathcal{T})) = (\bar{\mathbf{t}}_0, \mathbf{t}_1, \dots, \mathbf{t}_i)) := \omega(\mathbf{t}_i) \prod_{j=1}^i \mathbf{Q}(\mathbf{t}_i, \mathbf{t}_{i-1}), \quad (3.8)$$

where the product is taken to be one if $i = 0$. The following Lemma follows by adapting the proof of [3, Propositions 10 and 11].

Lemma 3.9. *Suppose a sequence of trees $\{\mathcal{T}_n\}_{n \geq 1}$ converges almost surely, in the fringe sense, to ω . Moreover, suppose that ω is a fringe distribution in the sense of Definition 3.8. Then $\{\mathcal{T}_n\}_{n \geq 1}$ converges almost surely, in the extended fringe sense, to a limiting infinite random sin-tree $\mathcal{T}_\infty$ whose law ω^{EF} is uniquely obtained from ω via (3.8).*

Both notions imply convergence of functionals such as the degree distribution. For example, in notion (a), letting $\mathcal{T}_\omega \sim \omega$ with root denoted by 0 say, convergence in notion (a) in particular implies, for any $\ell \geq 0$,

$$\frac{\#\{v \in \mathcal{T}_n, \deg(v) = \ell + 1\}}{n} \xrightarrow{\text{a.e.}} \mathbb{P}(\deg(0, \mathcal{T}_\omega) = \ell). \quad (3.9)$$

However, both convergences give a lot more information about the asymptotic properties of $\mathcal{T}_n$, beyond its degree distribution, we will see some specific example in the next Section.

3.5. Extensions. The basic technical tools outlined above can be extended to various settings. Let us give one example, and leave other settings such as weighted networks (sometimes called Geometric networks) for the reader (and if lost, consult the first few pages of [6]). Recall that one of the motivations given in the introduction were attributed network models, namely where vertices have types. Assuming the networks under consideration are trees, let us now show how all the concepts explored in the previous section on fringe convergence easily extend. Fix attribute space $\mathcal{S}$ and assume for the rest of the paper that $\mathcal{S}$ is a Polish space with distance metric $d_{\mathcal{S}}$. For $n \geq 1$, let $\mathbb{T}_{n, \mathcal{S}}$ be the space of all rooted trees on n vertices where every vertex has a mark in $\mathcal{S}$. Let $\mathbb{T}_{\mathcal{S}} = \cup_{n=0}^{\infty} \mathbb{T}_{n, \mathcal{S}}$ be the space of all finite rooted marked trees. Here $\mathbb{T}_{0, \mathcal{S}} = \emptyset$ will be used to represent the empty tree (tree on zero vertices). For any $\mathbf{t} \in \mathbb{T}_{\mathcal{S}}$ and $v \in \mathbf{t}$, write $a(v)$ for the corresponding attribute of that vertex. Let $\rho_{\mathbf{t}}$ denote the root of $\mathbf{t}$. For any $r \geq 0$ and $\mathbf{t} \in \mathbb{T}_{\mathcal{S}}$, let $B(\mathbf{t}, r) \in \mathbb{T}_{\mathcal{S}}$ denote the subgraph of $\mathbf{t}$ of vertices within graph distance r from $\rho_{\mathbf{t}}$, viewed as an element of $\mathbb{T}_{\mathcal{S}}$ and rooted again at $\rho_{\mathbf{t}}$.

Given two rooted finite trees $\mathbf{s}, \mathbf{t} \in \mathbb{T}_{\mathcal{S}}$, say that $\mathbf{s} \simeq \mathbf{t}$ if, after ignoring all attribute information, there exists a **root preserving** isomorphism between the two trees viewed as unlabelled graphs. Given two rooted trees $\mathbf{t}, \mathbf{s} \in \mathbb{T}_{\mathcal{S}}$ (adapting [62, Equation 2.3.15]), define the distance

$$d_{\mathbb{T}_{\mathcal{S}}}(\mathbf{t}, \mathbf{s}) := \frac{1}{1 + R^*} \quad (3.10)$$

where

$$R^* = \sup\{r : B(\mathbf{t}, r) \simeq B(\mathbf{s}, r), \text{ and } \exists \text{ isomorphism } \phi_r \text{ between } B(\mathbf{t}, r) \text{ and } B(\mathbf{s}, r) \text{ with } d_{\mathcal{S}}(a(v), a(\phi_r(v))) \leq 1/r \ \forall v \in B(\mathbf{t}, r)\}.$$

Now proceed with all the notions of convergence, local weak convergence, fringe convergence etc as in the previous Subsection.

3.6. Examples. While all of the results described are well known in the literature, we will give references where a passionate reader new to the field might find complete proofs.

3.6.1. Growing random trees. Let us start with an example and describe how the lack of memory property of the exponential distribution transforms this into a more tractable model and then describe the general setting. Consider a sequence of growing random trees $\{\mathcal{T}_k : k \geq 1\}$, initialized at $k = 1$ with $\mathcal{T}_1$ having one vertex, which we will call the root. For $k \geq 1$, having constructed $\mathcal{T}_{k-1}$, $\mathcal{T}_k$ is constructed as follows: a new vertex labelled k enters the system and connects to a randomly selected vertex uniformly at random amongst the available $k - 1$ vertices in $\mathcal{T}_{k-1}$. Suppose we wanted to understand the asymptotics of this model.

The Zen of the exponential distribution: We will now consider an embedding of the above tree process in continuous time, an example of the famous Athreya-Karlin embedding [12] which uses the *lack of memory property* of the exponential distribution. Consider the following *continuous time* branching process $\{BP(t) : t \geq 0\}$ started with a single individual ρ at time $t = 0$ with dynamics:

- (a) Each individual in the system lives forever.
- (b) The offspring distribution of every individual is a Poisson rate one process independent across individuals; rephrasing, each individual in the branching process gives birth to new individuals at rate one.

It turns out, this is a classical and fundamental example of a pure birth process often called the *rate one* Yule process. For any fixed $t \geq 0$, let $|BP(t)|$ denote the size of the branching process at this time and for any $k \geq 1$ define the stopping time $T_k = \inf\{t : |BP(t)| = k\}$. The following fundamental facts are known about the Yule process and its connection to the discrete time tree process described before:

- (a) Using the lack of memory property of the exponential distribution, for any $k \geq 1$, consider the tree $\tilde{\mathcal{T}}_k$ describing parent child relationships in $BP(T_k)$ (see Figures 3.4 and 3.5. Then $\tilde{\mathcal{T}}_k \stackrel{d}{=} \mathcal{T}_k$ and in fact, viewed as growing tree processes, $\{\tilde{\mathcal{T}}_k : k \geq 1\} \stackrel{d}{=} \{\mathcal{T}_k : k \geq 1\}$. This is one example of the Athreya-Karlin embedding of discrete time processes in continuous time processes. The key conceptual point is that, in the dynamics of the continuous time process, there is a *plethora of independence* since every vertex behaves independently and so is much more amenable to probabilistic analysis.
- (b) The process $|BP(t)|$ grows exponentially at rate one in the sense that there exists a rate one exponential random variable W such that $e^{-t}|BP(t)| \rightarrow W$ a.s. and in $\mathbb{L}^2$.

Rationale for fringe convergence of $\{\mathcal{T}_k : k \geq 1\}$: Let us go back to the continuous time process $\{BP(t) : t \geq 0\}$ and suppose for “large t ”, we sample a vertex V_t uniformly at random from $BP(t)$. Let $\mathfrak{B}(V_t) \leq t$ denote the time this vertex was born into the system, so that by time t , if we consider the $\text{Age}(V_t) = t - \mathfrak{B}(V_t)$ of this vertex then for any $a \geq 0$,

$$\mathbb{P}(\text{Age}(V_t) \geq a | BP(t)) = \mathbb{P}(V_t \in BP(t-a) | BP(t)) = \frac{|BP(t-a)|}{|BP(t)|} \rightarrow \exp(-a), \quad (3.11)$$

where in the last assertion, we have used (b). Now since in continuous time all the dynamics is independent, the following seems like a believable guess:

Guess: The fringe distribution of the $\{\mathcal{T}_k : k \geq 1\}$ should converge to the genealogical tree of the branching process BP observed for a random exponential mean one amount of time.

This is true, and in fact the above implies something much deeper.

Theorem 3.10 ([3, 38–40]). *The sequence of random trees $\mathcal{T}_n \xrightarrow{\text{a.s. -efr}} \mathcal{T}_\infty$ where $\mathcal{T}_\infty$ is the random sin-tree whose fringe distribution $\mathfrak{d}$ is constructed as follows:*

- (i) Generate $T \sim \exp(1)$.
- (ii) Generate a continuous time branching process BP with offspring distribution given by a Poisson process of rate one, independent of T .

Let $BP(T)$ genealogical tree of the branching process observed for T units of time then $\mathfrak{d}$ is the distribution on $\mathbb{T}$ of this random finite tree.

Spurred by the above discussion, it turns out much of the above discussion can be extended to much more general tree models. Consider the general random tree model sometimes called *non-uniform random recursive trees* [60]. Fix $n \geq 1$ and an *attachment function* $f : \{0, 1, 2, \dots\} \rightarrow (0, \infty)$. A sequence of random trees $\{\mathcal{T}_k : 1 \leq k \leq n\}$ is grown as follows ($\mathcal{T}_k$ has k vertices labelled by the integers $[k] := \{1, \dots, k\}$). For $k = 1$, $\mathcal{T}_1$ has one vertex, which we call the “root.” For fixed $k \geq 2$, $\mathcal{T}_k$ is

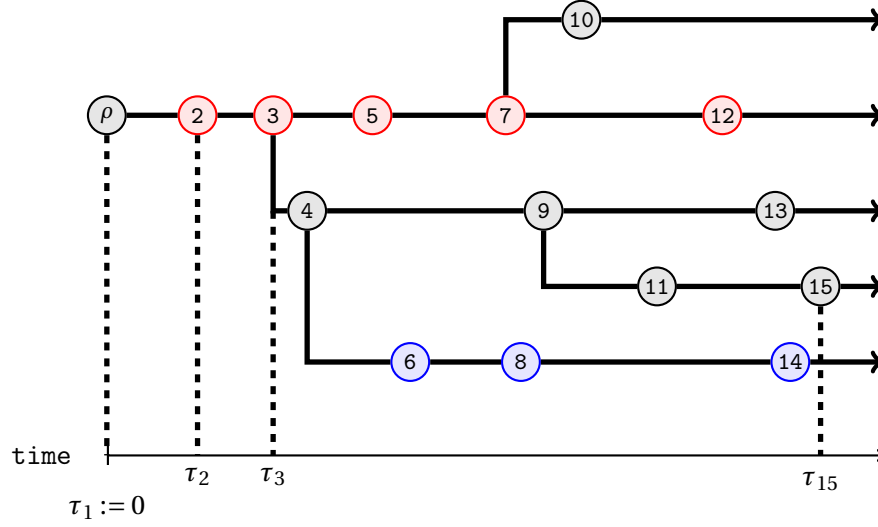


FIGURE 3.4. The process $\text{BP}_\alpha(\cdot)$ in continuous time starting from the root ρ and stopped at τ_{15} .

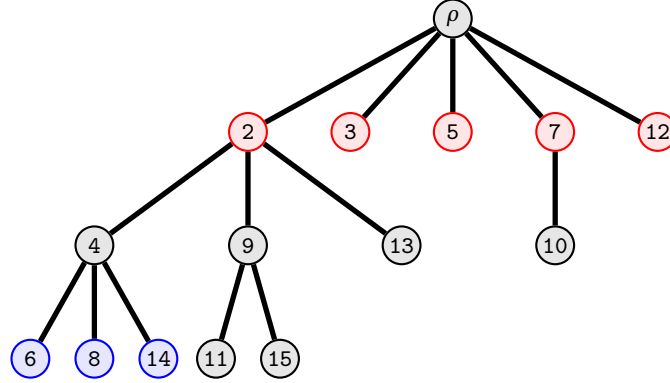


FIGURE 3.5. The corresponding discrete tree containing only the genealogical information of vertices in $\text{BP}_\alpha(\tau_{15})$.

constructed conditional on $\mathcal{T}_{k-1}$ as follows. A new vertex, k , is born into the system and attaches to a previously existing vertex $v \in [k-1]$ with probability proportional to $f(\deg(v))$, where $\deg(v)$ denotes the number of children of v (which is one less than its graph degree in $\mathcal{T}_{k-1}$). Thus,

$$\mathbb{P}(k \text{ attaches to } v \in [k-1] \mid \mathcal{T}_{k-1}) := \frac{f(\deg(v))}{\sum_{u=1}^{k-1} f(\deg(u))}.$$

The vertex that k selects is called the “parent” and the edge is directed from the parent to the new “child” vertex. The case of $f(\cdot) \equiv 1$ corresponds to the above model we described and is now known known under the phrase *random recursive trees* [58]. The specific case of *linear preferential attachment* when f is affine, namely $f(k) = k + 1 + \beta$, $k \geq 0$, for a fixed parameter β was considered in [16] to provide a generative story for heavy tailed degree distributions of real networks.

Then it turns out, that each of these models can be embedded in continuous time in corresponding continuous time branching processes BP_{model} , as we did for the random recursive tree

model. Further, under technical conditions on the attachment function f , each of the models have a (model dependent) constant $\lambda = \lambda_{\text{model}} > 0$ (called the Malthusian rate of growth of this branching process) such that $|\text{BP}_{\text{model}}(t)| \sim W_{\text{model}} e^{\lambda t}$. Thus the above heuristic once again leads to guesses as well as a proof path to show this class of random tree models converge to limiting (model dependent) random $\sin$ -trees. Further discussion is postponed to Section 5.

3.6.2. Networks: Erdős-Rényi and the Configuration model. While all the results here are well known, a description of the history and proofs can be found in [62, Chapter 2]. Recall the intuition developed related to the Erdős-Rényi random graph in Section 3.2. The following should not come as a surprise.

Theorem 3.11. *Fix $\lambda > 0$ and consider the sequence of random graphs $\{\mathcal{G}_n : n \geq 1\}$ with $\mathcal{G}_n \sim \text{ERRG}(n, \lambda/n)$. Let $\mathcal{G}_\infty \sim \text{BP}(\text{Pois}(\lambda))$ denote a Galton-Watson Branching process with $\text{Pois}(\lambda)$ offspring distribution. Then $\mathcal{G}_n \xrightarrow{\text{P-LWC}} \mathcal{G}_\infty$ as $n \rightarrow \infty$.*

This is also a good place to describe functionals that local weak convergence might **not** be able to capture **without more work or conditions**. Consider the sequence of network models $\{\tilde{\mathcal{G}}_n : n \geq 1\}$ via:

- (a) $\tilde{\mathcal{G}}_n$ has vertex set $[2n]$ partitioned as $[1 : n]$ and $[n+1 : 2n]$.
- (b) Generate independent $\text{ERRG}(n, 4/n)$ on each of these two vertices.
- (c) Select $U_n^{(1)}, U_n^{(2)}$ uniformly at random from $[1 : n]$ and $[n+1 : 2n]$ and connect them by a single edge.

Then it is possible to check that $\tilde{\mathcal{G}}_n \xrightarrow{\text{P-LWC}} \mathcal{G}_\infty$ as $n \rightarrow \infty$ as in the above result, however this sequence of networks are **qualitatively** different from the ERRG model; for example with some strictly positive probability as $n \rightarrow \infty$, these networks have two disjoint giant components (components of the same order as the system size $2n$) as $n \rightarrow \infty$. This should not be surprising as local weak convergence is not built to reveal asymptotics for global functionals (such as the connectivity properties of the network). What will be surprising is, with a lot of hard work, one can push this notion to give detailed asymptotics for global functionals in a wide variety of models.

The next example describes a mechanism for generating network models such that (in the limit), the empirical degree distribution converges to a pre-specified pmf $\mathbf{p} := \{p_k : k \geq 0\}$. There are various variants of the following model, known as the *configuration model*, and we will pick one specific construction. For network with vertex set $[n]$:

- (a) Generate $\mathbf{d}_n := \{d_i : i \in [n]\}$ iid with distribution $\mathbf{p}$. Assume $\ell_n = \sum_{i \in [n]} d_i$ is even else add one to d_n .
- (b) Imagine every vertex $i \in [n]$ having d_i half edges, where two half-edges have to be paired to create a full edge. Label ℓ_n half edges in some arbitrary fashion from $1, 2, \dots, \ell_n$.
- (c) Start half-edge labelled one and pair this one of the remaining $\ell_n - 1$ half-edges uniformly at random to form a full edge and remove both of these half-edges; next pick the smallest indexed unpaired half-edge and pair this one of the remaining $\ell_n - 3$ remaining half-edges to form the second full edge and remove both of these from the collection of unpaired half-edges; continue till all half-edges have been paired.

Write $\mathcal{G}_n \sim \text{CM}_n(\mathbf{p})$ for the random network obtained at the end of this process; in principle this will be an *multi-graph* as there might potentially be multiple edges and self-loops.

Intuition for asymptotics: In the large network $n \rightarrow \infty$ limit, suppose we pick a vertex V_n uniformly at random from this graph; then the number of neighbors (namely initial half-edges) of this vertex should have distribution $\mathbf{p}$. Let us follow one of these half-edges of V_n to understand the degree of the vertex whose half-edge connects to this half-edge completing a full edge; conceptually vertices

with degree k (of which there are a density p_k) are k times more likely to complete this edge than a vertex of degree 1. Thus it is not hard to convince oneself that the probability that the degree of the vertex that completes this edge is k should be proportional to kp_k . However after completing this edge, this new vertex has $k - 1$ half-edges to connect to subsequent vertices. This motivates the following definitions.

Definition 3.12 (Unimodular branching processes with input pmf $\mathbf{p}$). *Assume the pmf $\mathbf{p}$ has finite mean $\mu = \sum_{k=0}^{\infty} kp_k < \infty$. Define the size-biased pmf $\mathbf{p}^\circ$ given by:*

$$p_k^\circ = \frac{(k+1)p_{k+1}}{\mu}, \quad k \geq 0.$$

Consider the branching process $\text{BP}^\circ(\mathbf{p}, \mathbf{p}^\circ)$ started from a single root ρ where the root has offspring distribution $\mathbf{p}$ but every subsequent generation has offspring distribution $\mathbf{p}^\circ$.

While the following is well known in the literature, a full proof can once again be found in [62, Chapter 2]. Here Let $D \sim \mathbf{p}$ will denote a generic random variable with distribution given by the degree pmf generating $\text{CM}_n(\mathbf{p})$.

Theorem 3.13. *Assume $\mathbb{E}(D^2) < \infty$. Then the sequence of models $\{\mathcal{G}_n : n \geq 1\}$ with $\mathcal{G}_n \sim \text{CM}_n(\mathbf{p})$ satisfy $\mathcal{G}_n \xrightarrow{\text{P-LWC}} \text{BP}^\circ(\mathbf{p}, \mathbf{p}^\circ)$ as $n \rightarrow \infty$ in the sense of Definition 3.6.*

4. POWER OF LOCAL WEAK CONVERGENCE

4.1. An illustrative example: Spectral distribution of adjacency matrices for trees. Recall that $\mu_{\text{EMP}}^{(n)}$ denoted the empirical distribution of the eigen-values of $\mathcal{T}_n$.

Theorem 4.1 ([19]). *Consider a sequence of trees converging in fringe sense to a random infinite sin-tree. Then there exists a model dependent, non-random, probability measure μ_{model} such that $\mu_{\text{EMP}}^{(n)} \xrightarrow{\text{P}} \mu_{\text{model}}$ as $n \rightarrow \infty$.*

Rationale for why this might be true: Recall that $\mathcal{M}_{\mathbb{P}}(\mathbb{R})$ denotes the space of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For $\mu \in \mathcal{M}_{\mathbb{P}}(\mathbb{R})$, write $\text{supp}(\mu)$ for the support of μ . For complex $z \in \mathbb{C}$, $z \notin \text{supp}(\mu)$, define the Stieltjes transform at z ,

$$s_\mu(z) := \int_{\mathbb{R}} \frac{1}{x - z} d\mu(x). \quad (4.1)$$

One can check that if $z = x + iy$ with $y > 0$ then $|s_\mu(z)| < 1/y$. It seems believable that if one can show, that there is a (model dependent) deterministic function $s_{\text{model}, \infty}(z)$ such that for $z \in \mathbf{A} \subseteq \mathbb{C}$, $s_{\mu_{\text{EMP}}^{(n)}}(z) \xrightarrow{\text{P}} s_{\text{model}, \infty}(z)$, where the domain of convergence $\mathbf{A}$ is “large enough”, then this should at least augur convergence of the empirical distribution to a limit deterministic measure μ_{model} whose Stieltjes transform is given by $s_{\text{model}, \infty}(z)$.

Now for $\mu_{\text{EMP}}^{(n)}$, with Tr denoting the trace of a matrix,

$$s_{\mu_{\text{EMP}}^{(n)}}(z) = \frac{1}{n} \text{Tr}(\mathbf{A}_n - zI)^{-1} := \frac{1}{n} \sum_{v \in \mathcal{T}_n} R_{(vv), \mathcal{T}_n}(z), \quad (4.2)$$

where $R_{(vv), \mathcal{T}_n}(\cdot)$ is sometimes referred to as the *resolvent* of vertex v in the tree $\mathcal{T}_n$. Let $\mathcal{N}(v)$ denote the number of children of v in $\mathcal{T}_n^-(v)$ and denote these as $\{c_i(v) : 1 \leq i \leq \mathcal{N}(v)\}$ and the corresponding subtrees hanging from these vertices by $\mathcal{T}_{n,i}^-(v)$ (viewed as a rooted tree at $c_i(v)$). Let $\rho(v)$ denote the parent of v (namely the unique node on the path to the root of $\mathcal{T}_n$) and let $\mathcal{T}_n^+(v)$ denote the tree rooted at $\rho(v)$ obtained by deleting the edge $(v, \rho(v))$ and the fringe tree below v .

In the setting of trees, it turns out via Schur-decomposition that the resolvent at v in the original tree $\mathcal{T}_n$ can be decomposed as,

$$R_{(vv), \mathcal{T}_n}(z) = \frac{1}{-z + \sum_{i=1}^{\mathcal{N}(v)} R_{(c_i(v)c_i(v)), \mathcal{T}_{n,i}^-(v)}(z) + R_{(\rho(v), \rho(v)), \mathcal{T}_n^+(v)}(z)}. \quad (4.3)$$

Fix $z \in \mathbb{C}$ with imaginary part $\text{Im}(z) > 1$ and iterate the above expansion d times, resulting in a continued fraction type expansion, where we have d terms dependent on the fringe expansion upto distance d and a remainder term that depends on the graph beyond distance d from v . Intuitively it seems clear that, if one has extended fringe convergence as in Def. 3.9, then when $\text{Im}(z) > 1$, this should thus imply that $s_{\mu_{\text{EMP}}^{(n)}}(z) \xrightarrow{P} s_{\text{model}, \infty}(z)$. Converting this idea into a rigorous proof takes a lot of work and can be found in [19].

What did we learn? Conceptually the spectral distribution of the adjacency matrix is a **global** functional and depends on the entire network. However if one has local weak convergence, which a priori only gives information on asymptotics of local functionals, one can exploit it to also establish convergence of global functionals, provided that the effect of ‘far away’ vertices on the functional value of a given vertex diminishes in a quantifiable way.

4.2. Random matrices for general random graph models. Now let us consider the case of general (not-necessarily tree) networks that still converge in the local weak sense to trees. Our goal is to describe a small sub-collection of the foundational results in [24]. We will mainly focus on the configuration model driven by pmf $\mathbf{p}$ (owing to the assumption in Theorem 3.13 we assume $\mathbf{p}$ has finite second moment). We let $\text{BP}(\mathbf{p}^\circ)$ denote a (standard) branching process with offspring distribution $\mathbf{p}^\circ$ while $\text{BP}^\circ(\mathbf{p}, \mathbf{p}^\circ)$ denotes the unimodular branching process as in Def. 3.12; similarly define $D \sim \mathbf{p}, D^\circ \sim \mathbf{p}^\circ$. Let $\mathcal{G}_n \sim \text{CM}_n(\mathbf{p})$ and let $\mathbf{A}_n, \mu_{\text{EMP}}^{(n)}$ as before denote the corresponding adjacency matrices and empirical spectral distribution respectively. The Stieltjes transform as in (4.2) will be our main tool again so let us give a name to the class of such functions, with $\mathbb{C}_+ := \{z \in \mathbb{C}, \text{Im}(z) > 0\}$:

$$\mathcal{H} = \left\{ f : \mathbb{C}_+ \rightarrow \mathbb{C}_+, f \text{ holomorphic, } |f(z)| \leq \frac{1}{\text{Im}(z)} \right\}. \quad (4.4)$$

Paraphrasing a small sub-class of many amazing results:

Theorem 4.2 ([24]). (a) *Let $\mathcal{P}(\mathcal{H})$ denote the class of probability measures on $\mathcal{H}$. There is a unique probability measure $\mathbb{Q} \in \mathcal{P}(\mathcal{H})$ such that $Y \sim \mathbb{Q}$ satisfies the following recursive distributional equation:*

$$Y(z) \stackrel{d}{=} -\frac{1}{z + \sum_{i=1}^{D^\circ} Y_i(z)}$$

where $\mathbf{Y} := \{Y_i(\cdot) : i \geq 1\}$ are iid with distribution $\mathbb{Q}$ and independent of $D^\circ \sim \mathbf{p}^\circ$.

(b) *Let $D \sim \mathbf{p}$ independent of $\mathbf{Y}$ and define the random element $X \in \mathcal{H}$ via,*

$$X(z) = -\frac{1}{z + \sum_{i=1}^D Y_i(z)},$$

Then $\mu_{\text{EMP}}^{(n)} \xrightarrow{P} \mu_{\infty, \mathbf{p}}$ where $\mu_{\infty, \mathbf{p}}$ is the unique probability measure on $\mathcal{P}(\mathbb{R})$ with Stieltjes transform $s_{\infty, \mathbf{p}}(z) = \mathbb{E}(X(z))$.

Illustrative example: Fix $k \geq 3$ and consider the degree distribution $\mathbf{p} \sim \delta_k$, this then results in the random k -regular graph, which converges in the LWC sense to a rooted tree where the root has degree k and all the remaining individuals have degree $k-1$. Suppose, as a guess, one searches for deterministic solutions $Y(\cdot)$ in (a) above. This results in $Y(z) = -(z + (k-1)Y(z))^{-1}$. Solving, one finds that this is the Stieltjes transform of the semi-circular law with radius $2\sqrt{k-1}$. Using this

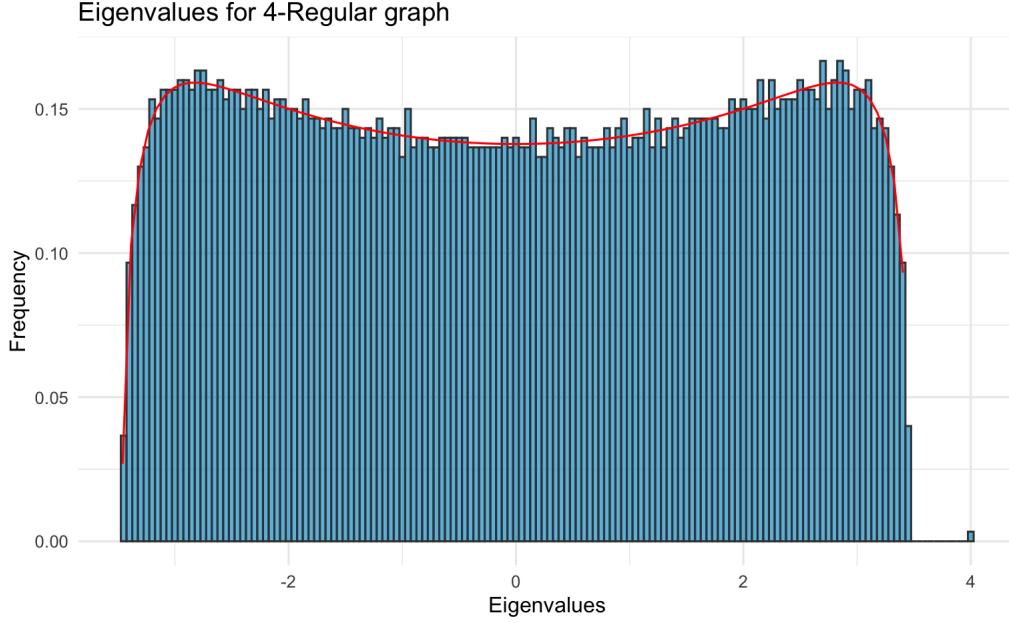


FIGURE 4.1. A simulation with $k = 4$ and network size $n = 6000$ showing the empirical spectral distribution with the theoretical limit.

in (b) shows that the empirical spectral distribution of random k -regular graph converges to the probability measure with density

$$f(x) = \frac{k}{2\pi} \frac{\sqrt{4(k-1) - x^2}}{k^2 - x^2}, \quad x \in [-2\sqrt{k-1}, 2\sqrt{k-1}].$$

Rationale for why this might be true: There are four steps in the proof in [24]:

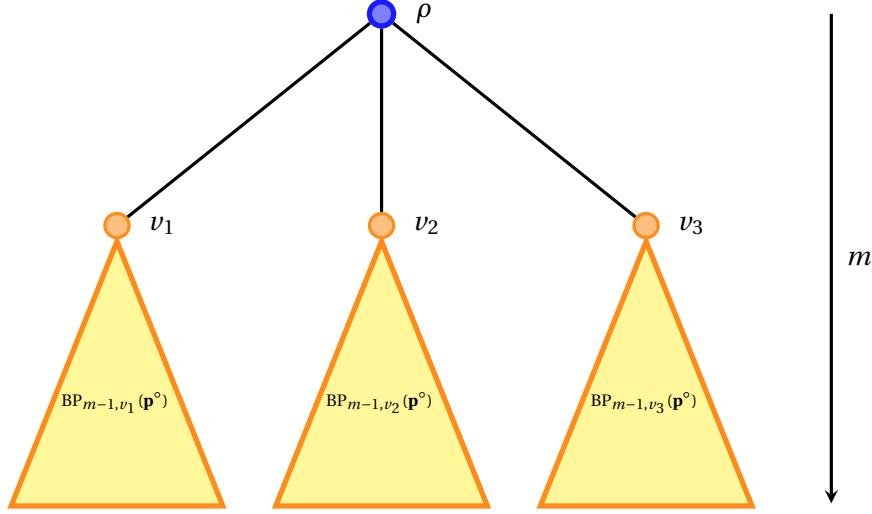
- (i) **Local weak convergence implies convergence:** Using general functional analytic techniques, the authors are able to show that local weak convergence (not even necessarily for trees) with some minor technical conditions automatically implies convergence of the Stieltjes transform with the limit Stieltjes transform $s_\infty(z) = \lim_{n \rightarrow \infty} \mathbb{E}(R_{V_n, V_n}(z))$, where $R_{V_n, V_n}(\cdot)$ denotes the resolvent of a uniformly selected vertex V_n .
- (ii) **Specific analysis when one has convergence to BP° :** Now let us consider the case of the configuration model where one has convergence to BP° . Let us start with an additional simplification: let us try to understand what happens to the adjacency matrix of the standard branching process $\text{BP}_m(\mathbf{p}^\circ)$ grown to m levels say, with adjacency matrix $\mathbf{A}_m$, in particular what happens to the resolvent $Y_\rho^m(z) := R_{\rho, \rho}(z) = \text{Tr}(\mathbf{A}_m - z\mathbf{I})_{\rho, \rho}^{-1}$. Let $D_\rho^\circ \sim \mathbf{p}$ denote the number of children of the root and write these as $\{v_1, v_2, \dots, v_{D_\rho^\circ}\}$. Note that the subtrees below each of these vertices constitute independent Branching processes run upto $m-1$ generations, see Figure 4.2.

Let $\mathbf{A}_{m-1}^{v_i}$ denote the adjacency matrices of each of these subtrees and let

$$Y_{v_i}^{m-1}(z) = \text{Tr}(\mathbf{A}_{m-1}^{v_i} - z\mathbf{I})_{v_i, v_i}^{-1}$$

An application of Schur-decomposition as in the previous Section results in the equation:

$$Y_\rho^m(z) = -\frac{1}{z + \sum_{i=1}^{D_\rho^\circ} Y_{v_i}^{m-1}(z)}. \quad (4.5)$$

FIGURE 4.2. Figure of $\text{BP}^o(\mathbf{p}^o)$ with root degree $D_\rho^o = 3$

- (iii) **Recursive distributional equations and uniqueness:** Consider the map $\Psi : \mathcal{P}(\mathcal{H}) \rightarrow \mathcal{P}(\mathcal{H})$ defined as follows: For $\mathbb{Q} \in \mathcal{P}(\mathcal{H})$ define $\Psi(\mathbb{Q})$ to be the distribution of the random variable $\tilde{Z}$ obtained as follows:
- (a) Generate $\mathbf{Z} = \{Z_i(\cdot) : i \geq 1\}$ iid with distribution $\mathbb{Q}$ and independent of $D^o \sim \mathbf{p}^o$.
 - (b) Let $\tilde{Z}(z) = -(z + \sum_{i=1}^{D^o} Z_i(z))^{-1}$.
- Say that $\mathbb{Q}$ is a solution of the recursive distribution equation if $\Psi(\mathbb{Q}) = \mathbb{Q}$. Suppose we were able to show that this recursive distributional equation has a **unique** solution. Then (4.5) should (after some work) imply that as $m \rightarrow \infty$, $Y^m(\cdot)$ converges to this unique solution as $m \rightarrow \infty$. The authors in [24] are able to prove this uniqueness via showing that this mapping is a contraction and then using Banach fixed point theorem.
- (iv) **Relating this to CM_n :** Redoing the Schur decomposition analysis but now to understand $R_{V_n, V_n}(\cdot)$ for a randomly selected vertex in CM_n , and connecting it to the above analysis of $\text{BP}^o(\mathbf{p}, \mathbf{p}^o)$ via local weak convergence as described before, it should now be believable that $R_{V_n, V_n}(z) \approx X(z)$ for large n where $X(z)$ satisfies the distributional equation in (b) of the Theorem.

What did we learn? Here the local computations not just proved the existence of the limit (of the empirical spectral distribution) but owing to the recursive construction of the limit using independent and identically distributed components (here, independent branching processes hanging off the root), this resulted in explicit information about the limit via *computations* carried out on the limit object. This phenomenon related to local weak convergence was termed the *objective method* in [6] where local weak convergence allows one to:

- (a) Prove asymptotics for functionals of interest in the large network limit;
- (b) Obtain information on the limit via recursive computations performed on the limit object.

4.3. Interacting particle systems on random graphs. Recall the motivation in Section 2.3. Once again our goal will be to understand properties of the Gibbs distribution on a sequence of configuration model driven graphs $\mathcal{G}_n \sim \text{CM}_n(\mathbf{p})$ as $n \rightarrow \infty$. Let $\mu_{\mathcal{G}_n}(\cdot)$ denote the probability measure in (2.4). Recall that in the previous sections, the first step was to relate such questions to a functional (the Stieltjes transform for example) that we could use to pave the way for understanding

asymptotics. In this setting, the normalizing factor $Z_n(\beta, B)$, which is a function of the inverse temperature β and external field B seems like a natural candidate. Since the number of potential configurations (terms in the summand) is exponential, the following object:

$$\phi_n(\beta, B) = \frac{1}{n} \log Z_n(\beta, B), \quad (4.6)$$

seems like a natural starting point. Further note that appropriate derivatives of this function give us important information, for example, the mean magnetization is given by,

$$\frac{\partial \phi_n(\beta, B)}{\partial B} = \sum_{\mathbf{x} \in \{-1, 1\}^n} \frac{\sum_{i \in [n]} x_i}{n} \mu_{\mathcal{G}_n}(\mathbf{x}) = m_{\mathcal{G}_n}(\beta, B). \quad (4.7)$$

Thus we try to understanding asymptotics of this object as $n \rightarrow \infty$. Our initial goal in this Section is to describe the foundational results in [31] describing these asymptotics in the *ferromagnetic* ($\beta \geq 0$) regime. Since the model is “symmetric” with respect to B (in the sense that $\phi_n(\beta, B) = \phi_n(\beta, -B)$), throughout this Section we will assume that $B > 0$. We will now describe the foundational result in [31], the reader should contrast the result and the preceding setup with Theorem 4.2. First we need some notation. Throughout this Section we assume $\beta \geq 0$. Let $\tanh, \operatorname{atanh}$ denote the tan hyperbolic and inverse tan hyperbolic functions respectively. Define the function,

$$\mathfrak{E}(\beta, h) := \operatorname{atanh}[\tanh(\beta) \tanh(h)] \quad (4.8)$$

Consider the functional $\Psi_{\beta, B} : \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$ from the space of probability measures on $\mathbb{R}_+$ to itself given as the distribution of the random variable X obtained via:

- (a) Generate $\{Y_i : i \geq 1\} \sim_{iid} \mathbb{P}$ independent of $D^\circ \sim \mathbf{p}^\circ$.
- (b) Let $X = B + \sum_{i=1}^{D^\circ-1} \mathfrak{E}(\beta, Y_i)$.

Theorem 4.3. [31] *Consider the Ising model as in (2.4) on $\operatorname{CM}_n(\mathbf{p})$ with finite second moments and with $\beta \geq 0, B \geq 0$. Then,*

- (a) *For $\beta \geq 0$ the operation Ψ is monotone increasing (under the stochastic ordering operation) and there is a unique fixed point $\tilde{\mathbb{P}}_{\beta, B}$ supported on $[0, \infty)$ satisfying $\tilde{\mathbb{P}}_{\beta, B} = \Psi(\tilde{\mathbb{P}}_{\beta, B})$ i.e. for $Y, \mathbf{Y} := \{Y_i : i \geq 1\} \sim_{iid} \tilde{\mathbb{P}}_{\beta, B}$ independent of D° ,*

$$Y \stackrel{d}{=} B + \sum_{i=1}^{D^\circ-1} \mathfrak{E}(\beta, Y_i).$$

Call this the distribution of “local fields”.

- (b) *Let $D \sim \mathbf{p}$ independent of $\mathbf{Y}$. Then $\phi_n(\beta, B) \xrightarrow{P} \phi_\infty(\beta, B)$ where*

$$\begin{aligned} \phi_\infty(\beta, B) = & \frac{\mu}{2} \log \cosh(\beta) - \frac{\mu}{2} \mathbb{E}[\log[1 + \tanh(\beta) \tanh(Y_1) \tanh(Y_2)]] \\ & + \mathbb{E} \log \left\{ e^B \prod_{i=1}^D [1 + \tanh(\beta) \tanh(Y_i)] + e^{-B} \prod_{i=1}^D [1 - \tanh(\beta) \tanh(Y_i)] \right\}. \end{aligned}$$

Rationale for why this might be true: The proof of this result is significantly more complicated so we will only be able to describe a small aspect of the proof. We will write $\langle v, f \rangle = \sum_x f(x) v(x)$ for the expectation operation of a function with respect to a measure on a finite set. There are four major ideas in the proof:

- (a) **Monotonicity in the Ferromagnetic regime:** In the ferromagnetic regime, it turns out that there are fundamental monotonicity properties of the model that one can take advantage. Two important threads in this direction are:

- (i) **Monotonicity with respect to local magnetic fields:** For a graph $\mathcal{G}$ and vertex dependent non-negative constants $\mathbf{B} = \{B_u : u \in V(\mathcal{G})\}$ (called local field) consider a generalized version of the model in (2.4) on a graph via:

$$\mu_{\mathbf{B},\beta}(\mathbf{x}) = \frac{1}{Z(\beta, \mathbf{B})} \exp \left(\beta \sum_{(u,v) \in E(\mathcal{G})} x_u x_v + \sum_{u \in V} B_u x_u \right) \quad (4.9)$$

For two local field parameters $\mathbf{B}, \mathbf{B}'$ with $0 \leq B_u \leq B'_u$ for all $u \in \mathcal{G}$ and inverse temperature parameters $0 \leq \beta \leq \beta'$ Griffith's inequality says that for any subset $U \subset V(\mathcal{G})$, $0 \leq \langle \mu_{\beta, \mathbf{B}}, \prod_{i \in U} x_i \rangle \leq \langle \mu_{\beta', \mathbf{B}'}, \prod_{i \in U} x_i \rangle$.

- (ii) **Monotonicity with respect to boundary conditions:** To reduce notational overhead, we will suppress dependence on $\beta, \mathbf{B}$ of the measures in question unless required. Fix a connected graph $\mathcal{G}$ and a strict subset of vertices $U \subset \mathcal{G}$ and let the boundary ∂U denote the collection of vertices in U that have connections to at least one vertex in $\mathcal{G} \setminus U$. Suppose one was interested in the marginal distribution of the Ising model on the subgraph spanned by vertices in U . Then intuitively there are two Ising type models that “lower” and “upper” bound this distribution, the Ising model on U with free ($\mu^{U,0}$) and all + ($\mu^{U,+}$) boundary conditions respectively:

$$\begin{aligned} \mu^{U,0}(\mathbf{x}_U) &= \frac{1}{Z^{U,0}(\beta, \mathbf{B})} \exp \left(\beta \sum_{(u,v) \in E(U)} x_u x_v + \sum_{u \in U} B_u x_u \right), \\ \mu^{U,+}(\mathbf{x}_U) &= \frac{1}{Z^{U,+}(\beta, \mathbf{B})} \exp \left(\beta \sum_{(u,v) \in E(U)} x_u x_v + \sum_{u \in U} B_u x_u \right) \mathbb{1}_{\{\mathbf{x}_{\partial U} \equiv +\mathbf{1}\}}, \end{aligned}$$

i.e. when attempting to compute the marginal distribution of the Ising model on U , the first measure assumes no influence from the outside (in the ferromagnetic regime one would imagine some sort of “positive influence” from the boundary) while the second measure assumes maximal positive influence from the boundary in the sense that all these vertices are set to have spins +1.

- (b) **Connecting this to local weak convergence on trees and vanishing impact of boundary conditions:** While we are interested in ϕ_n , (4.7) already shows how *derivatives* of this object can be related to averages. After significant amount of work, what is shown in [31] is that to prove the main result, it is enough to show that for each $\beta \geq 0$,

$$\frac{\partial \phi_n(\beta, B)}{\partial \beta} \xrightarrow{P} \frac{\partial \phi_\infty(\beta, B)}{\partial \beta}.$$

Writing $\mu_{\mathcal{G}_n}$ for the Ising model on the graph $\mathcal{G}_n$, direct computation shows that,

$$\frac{\partial \phi_n(\beta, B)}{\partial \beta} = \frac{1}{n} \sum_{v \in [n]} \sum_{u \sim v} \langle \mu_{\mathcal{G}_n}, x_u x_v \rangle = \mathbb{E}_n \left(\sum_{u \in \mathcal{N}(V_n)} \langle \mu_{\mathcal{G}_n}, x_u x_{V_n} \rangle \right),$$

where $\mathbb{E}_n$ denotes expectation conditional on the graph $\mathcal{G}_n$ and thus the only randomization is over the choice of the random vertex V_n sampled uniformly from the graph. Now fix any $t \geq 1$ and as before let $B_{\mathcal{G}_n}(V_n, t)$ denote the ball of radius t around V_n . Conditional on V_n , let $\mu^{B_{\mathcal{G}_n}(V_n, t), 0}, \mu^{B_{\mathcal{G}_n}(V_n, t), +}$ denote the Ising models with free and + boundary conditions on this neighborhood and let $\mathbb{E}_n^{0,t}, \mathbb{E}_n^{+,t}$ denote expectations of functionals over choices of V_n . Then the monotonicity described in the previous step implies for each t ,

$$\mathbb{E}_n^{0,t} \left(\sum_{u \in \mathcal{N}(V_n)} \langle \mu^{B_{\mathcal{G}_n}(V_n, t), 0}, x_u x_{V_n} \rangle \right) \leq \frac{\partial \phi_n(\beta, B)}{\partial \beta} \leq \mathbb{E}_n^{+,t} \left(\sum_{u \in \mathcal{N}(V_n)} \langle \mu^{B_{\mathcal{G}_n}(V_n, t), +}, x_u x_{V_n} \rangle \right).$$

Now local weak convergence implies that as $n \rightarrow \infty$, the left hand side and right side converge to the following: Consider $\text{BP}(\rho, t) := \text{BP}^\circ(\mathbf{p}, \mathbf{p}^\circ)$ grown upto generation t from root ρ . Consider the Ising model on this with the same parameters with free and + boundaries and let $\mathbb{E}_\infty^{0,t}, \mathbb{E}_\infty^{+,t}$ denote expectations over the branching process measure. Then the left hand side should converge to

$$\mathbb{E}_n^{0,t} \left(\sum_{u \in \mathcal{N}(V_n)} \langle \mu^{B_{g_n}(V_n, t), 0}, x_u x_{V_n} \rangle \right) \xrightarrow{P} \mathbb{E}_\infty^{0,t} \left(\sum_{u \in \mathcal{N}(\rho)} \langle \mu^{\text{BP}(\rho, t), 0}, x_u x_\rho \rangle \right)$$

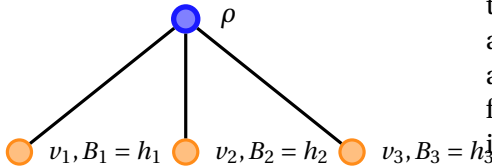
and similar expressions for the right hand side. Further it is shown that as $t \rightarrow \infty$, the effect of the boundary conditions on the above expectations vanish (a highly consequential finding). An example of quantification of such a result is as follows: consider the branching process $\text{BP}^\circ(\rho, m)$ as above upto generation m and the corresponding Ising model μ^m say with corresponding free and + measures $\mu^{m,0}, \mu^{m,+}$. Suppose for some $t < m$ we are interested in the marginal distribution $\mu_U^{m,0}, \mu_U^{m,+}$ of the Ising model on a subset of vertices $U \subset \text{BP}^\circ(\rho, t)$, in particular the impact of the above boundary conditions. Then [31, Theorem 4.2] shows that (under the assumptions of the Theorem) there are constants M, C such that the expected total variation distance satisfies,

$$\mathbb{E}(\|\mu_U^{m,0} - \mu_U^{m,+}\|_{TV}) \leq \frac{\mathbb{E}(C^{|\text{BP}(t)|})}{m-t}.$$

- (c) **Pruning argument for the Ising model on trees and the origin of local fields:** This sequence of arguments leads to trying to understand the Ising model on $\text{BP}^\circ(\rho, m)$ for “large m ” (where the boundary conditions are inconsequential) and in particular asymptotics of expectations of the sum taken over products of root vertex spin with its neighbors. Now we are back at Figure 4.2, the Ising model on this tree, and in particular the *marginal distribution of the root and its neighbors*. If we had a tractable description for this, we are in business. This is where another fundamental finding related to Ising models for trees in [31] comes to the rescue. Their general result says that the marginal distribution of the root and its neighbors can be *recursively* constructed as follows:

- (i) For each neighbor v_i of the root, consider the Ising model $\mu^{\text{BP}_{m-1, v_i}}$ on the corresponding subtree $\text{BP}_{m-1, v_i}(\mathbf{p}^\circ)$ (this includes the vertex v_i). Let $\langle \mu^{\text{BP}_{m-1, v_i}}, x_{v_i} \rangle$ denote the mean magnetization of the vertex v_i in this model.
- (ii) Then the marginal distribution on the root and its neighbors is again an Ising model with the same inverse temperature parameter β and where the root has local field B while each of its neighbors has local fields $h_i = \text{atanh}(\langle \mu^{\text{BP}_{m-1, v_i}}, x_{v_i} \rangle)$.

The distribution of these local field random variables can be recursively computed and this is what leads to the main theorem.



What did we learn? This application in particular shows that it sometimes takes a lot of work to connect local asymptotics of models to asymptotics of global functionals, in particular, to quantify the impact of vertices “far” from a typical vertex which do not have significant impact in terms of the functional being considered. Further we arrived at these “local fields” $\{h_i : 1 \leq i \leq D_\rho^\circ\}$ which quantified the impact of the (in the large $m \rightarrow \infty$ limit, infinite) subtrees below these children on the distribution of the

root. In the next application, we will see similar local fields arising out of, at first sight, a calculation that does not make any sense!

4.4. Random assignment problem. Recall the motivation and setup of the random assignment problem in Section 2.5. Under moment conditions (and assumption of continuity) on the edge cost distributions, it turns out that the limit depends only on the nature of the density at zero, so we will assume that the edge costs are *iid* exponential random variables; further it will be convenient to take these to have mean n (or rate $1/n$) variables. Before the nail in the coffin on this problem by Aldous in [4, 7], [36] had found a lower bound of 1.51 on the limit constant while [28] had found an upper bound of 1.94.

Theorem 4.4 ([7]). *Consider the random assignment problem where edge costs are exponential mean n random variables. Let A_n denote the cost of the minimum assignment problem. Then:*

- (a) *Let $\Xi = (\xi_1, \xi_2, \dots)$ be the points of a rate one Poisson point process and consider the recursive distributional equation given by*

$$X \stackrel{d}{=} \min_{1 \leq i < \infty} (\xi_i - X_i),$$

where $X, \{X_i : i \geq 1\}$ are i.i.d. with the same distribution and independent of Ξ . Then the only solution to this distributional equation is the logistic distribution $X \sim f_X(\cdot)$ where the density is given by,

$$f_X(x) = \frac{1}{(e^{x/2} + e^{-x/2})^2}, \quad -\infty < x < \infty.$$

- (b) *Let $X_1, X_2 \sim f_X(\cdot)$ and independent. Then*

$$\frac{1}{n} \mathbb{E}(A_n) \rightarrow \int_0^\infty x \mathbb{P}(X_1 + X_2 > x) dx = \frac{\pi^2}{6}.$$

Rationale for why this might be true:

- (a) **Asymptotics around a random job:** Writing $\pi_{opt}^{(n)}$ for the optimal assignment, and switching from $c_{i,j}$ to $c(i, j)$ to ease the load on subscripts, in the finite n problem, note that by symmetry we see that

$$\frac{1}{n} \mathbb{E}(A_n) = \mathbb{E} \left(c(V_n, \pi_{opt}^{(n)}(V_n)) \right).$$

Thus once again this problem has been transformed into something related to what happens for a typical vertex, in this case the cost of the edge that vertex V_n is matched to in the optimal assignment.

- (b) **Local weak convergence for Geometric graphs:** Let us think of the assignment problem as jobs being assigned to machines to facilitate explanation and by symmetry let $V_n = 1$. Recall that $\Xi = (\xi_1, \xi_2, \dots)$ denotes a Poisson point process as above. Let $\{Y_i : i \geq 1\}$ be an iid sequence of exponential mean one random variables. Now the least cost amongst possible machines to be assigned to this job is $\xi_{11}^{(n)} := \min_{j \in [n]} c_{1j} \stackrel{d}{=} \xi_1$ using properties of the exponential distribution. Using lack of memory property of the exponential distribution, the second smallest cost has distribution

$$\xi_{12}^{(n)} \stackrel{d}{=} \xi_1 + \frac{n}{n-1} Y_2 \approx \xi_2.$$

Similarly the 3rd least cost of possible machine assignments to job 1 has distribution,

$$\xi_{12}^{(n)} \stackrel{d}{=} \xi_{12}^{(n)} + \frac{n}{n-2} Y_3 \approx \xi_3.$$

Rigorous justification of the above facts proceeds via the Rényi representation of exponential order statistics. Thus, in the large n -limit, if we think of costs as edge lengths and explore the graph from job 1, then distances to machines are approximately distributed as a rate one Poisson point process. Now from each of these machines if we continue to explore the jobs closest to these machines (sequentially “unfolding” the geometric structure of costs from a single job),

once again in the large n -limit they have Poisson rate one edge lengths, independent across machines. This suggests the following natural limit object.

Definition 4.5 (Poisson weighted infinite tree (PWIT) [4]). *Consider the infinite locally finite random tree constructed as follows: start with a single root ρ and attach infinite number of children of this root where the children are assigned lengths according to a rate one Poisson process. Call this generation one. Now recursively repeat the process, namely assuming we have constructed the process upto generation k , for each vertex v in any generation k independently and identically repeat this process namely each of these individuals gives birth to individuals in generation $k + 1$ with edge lengths according to a rate one Poisson process Ξ_v , independent across individuals. Call the random rooted tree with edge lengths the Poisson weighted infinite Tree (PWIT) and denote this by $\mathcal{T}_\infty$. Let $\mathbf{W}$ denote the collection of edge lengths and $\mathbf{W}(e)$ for the weight of a specific edge.*

- (c) **Assignments as matchings and the first approach that fails:** Recall how we arrived at the PWIT, this was via unfolding the cost structure around a typical job in the finite n problem where we go from jobs in one generation to machines in the next generation and then to jobs in the generation after etc. Now any specific assignment of jobs to machines can, at least approximately (for large n) be thought of as a “matching” namely a collection of edges such that every vertex is part of exactly one edge. We can also define matchings $\mathcal{M}$ on the PWIT, namely a collection of edges in $\mathcal{T}_\infty$ such that every vertex is incident on exactly one edge. Further, whilst at the far end of one’s imagination, one can at least hope that something like the following is true:

There is an “optimal” matching on the PWIT such that as $n \rightarrow \infty$, $\mathbb{E}(c_n(1, \pi_{opt}^{(n)}(1)))$ converges to $\mathbb{E}(\mathbf{W}(\rho, \mathcal{M}_{opt}(\rho)))$ namely the expected weight of the edge containing the root.

Now suppose we try the natural first approach to construct a matching on the PWIT which is the greedy matching approach $\mathcal{M}_{greedy}$: match ρ with the vertex that is closest to it (which is at a distance of $\exp(1)$); then for all other children of ρ match them with their closest child etc. Then it is obvious that under this scheme $\mathbb{E}(\mathbf{W}(\rho, \mathcal{M}_{greedy})) = 1$, however as mentioned above, the lower bound for the limit constant before the result of Aldous was 1.51. Thus this cannot arise from the optimal matching. Conceptually it turns out that “good” matchings have to satisfy a spatial invariance property, in the sense that, going back to the definition of the PWIT, if we moved the root of the PWIT from the original root to one of its neighbors and then relabelled vertices, the matching should in some sense be invariant. It turns out that the greedy matching does not satisfy this property. In proving the existence of the limit constant, Aldous in [4] showed that $n^{-1} \mathbb{E}(A_n)$ converges to the $\inf \mathbb{E}(\mathbf{W}(\rho, \mathcal{M}(\rho)))$, where the infimum is taken over all spatially invariant matchings on the PWIT.

- (d) $\infty - \infty =$ **Magic. Origin of the Logistic distribution:** While trying to give an outline of the full and technical proof in [7] is beyond the scope of this article, let us describe one “magical” explanation (see [7, Sec 4.2]) that shows the origin of the Logistic distribution above. Label the children of vertex ρ from $\{1, 2, \dots\}$ and let $\mathcal{T}_{j,\infty}$ denote the subtree rooted at j which by construction has the same distribution as the PWIT. Let $C(\mathcal{T}_\infty)$ denote the weight of the optimal matching on the PWIT (which is obviously ∞ but let us keep going). Let (this is [7, Eqn 14]),

$$X_\rho = \text{cost of optimal matching on } \mathcal{T}_\infty - \text{cost of optimal matching on } \mathcal{T}_\infty \setminus \{\rho\},$$

and define similar variables for each of the children of ρ namely $\{X_j : j \geq 1\}$ are the corresponding random variables for the subtrees $\mathcal{T}_{j,\infty}$. By construction, if these random variables make sense in any universe, then they have the same distribution. For simplicity, on a set $\mathcal{U}$ write

$C(\mathcal{U})$ to denote the cost of the optimal matching on this set. Then note that

$$X_\rho = \min_{j \geq 1} \left[W(\rho, j) + C(\mathcal{T}_{j,\infty} \setminus \{j\}) + \sum_{i \neq j} C(\mathcal{T}_{i,\infty}) \right] - \sum_{j \geq 1} C(\mathcal{T}_{i,\infty}) = \min_{j \geq 1} \left[W(\rho, j) - X_j \right].$$

This is exactly the same recursion as in (a) of the Theorem leading to the logistic distribution. Further if j_{opt} is the child of ρ that minimizes the right hand side then the matching of the root corresponds to the edge (ρ, j_{opt}) . In [7] Aldous found a magical scheme to make sense of the above “calculations” and give a complete proof for the asymptotics of the limit constant.

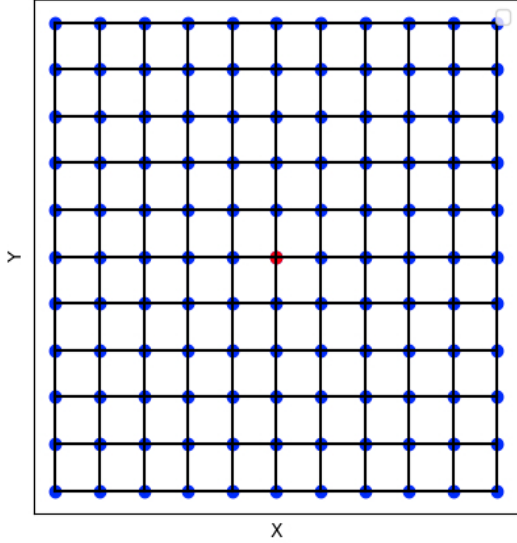


FIGURE 4.3. Standard integer lattice with the origin marked red.

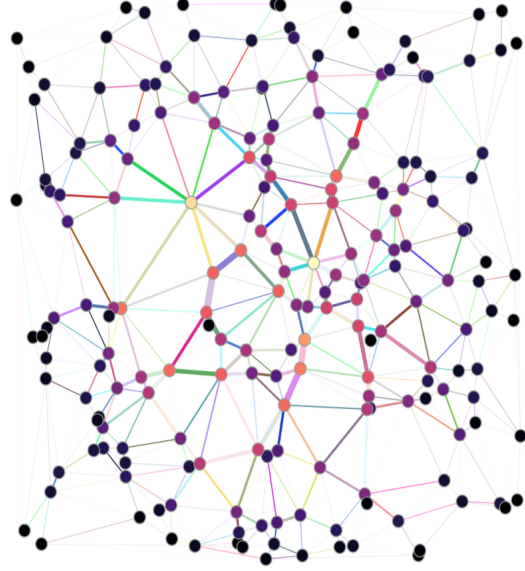


FIGURE 4.4. Random Triangulation of part of the plane resulting in a planar graph where each edge has degree 3.

4.5. Recurrence of random walks on planar graphs. The previous section described the importance of “spatial versions” of local weak convergence, where edge lengths played an important role. We conclude this section with a remarkable result of Benjamini and Schramm in [17] that in fact was another genesis for the notion of local weak convergence in the early 2000s. To set the stage, recall that one of the classical results says that, if one considers the two dimensional integer lattice then the random walk is recurrent, namely starting from any location, it returns to that location eventually with probability one. A natural question is: what if one considers other planar geometries, namely infinite graphs deterministic or random that can be embedded in space. Examples of two such geometries in (part of) the plane are given in Figures 4.3-4.4. We will now paraphrase the fundamental result in [17], the proof idea is significantly different from the rest of this review so we refer the interested reader to this beautiful paper for details regarding the proof.

Theorem 4.6 ([17]). *Let $\{\mathcal{G}_n : n \geq 1\}$ be a sequence of (potentially random) finite planar connected graphs with maximal degrees uniformly bounded by a constant M . Suppose $\mathcal{G}_n \xrightarrow{\text{LWC}} \mathcal{G}_\infty^*$ as $n \rightarrow \infty$. Then with probability one, the random walk on $\mathcal{G}_\infty^*$ is recurrent.*

Once again what is amazing is that recurrence of the limit object which is in principle a completely global property is modulated by local weak convergence.

5. LOCAL WEAK CONVERGENCE FOR DYNAMIC NETWORK MODELS

The previous Sections largely dealt with network models on *static* networks, namely where the underlying topology of the network does not change. The goal here is to describe dynamic models, and for the sake of clarity we will focus on various families of growing random tree models, some of whose motivation was given in Section 2 and fundamentals related to fringe convergence (Section 3.4) were described in Section 3.6.1. To keep this paper to manageable length, we will give a proof idea only in the first subsection and only paraphrase results in the remaining sections giving references to where the interested reader can find more details.

5.1. Local weak convergence and PageRank. Recall PageRank from Definition 2.1. Introduced by Google to rank web pages [52], PageRank is arguably one of the most effective centrality measures on complex networks and is a main factor behind the initial success of Google as a search engine. When compared to simpler centrality measures like degree, PageRank incorporates the effect of geometry of the network around a vertex beyond finite radius neighborhoods, and this should likely make it a more sensitive notion of centrality. This observation is elucidated by the following representation of PageRank, which we describe for tree networks for simplicity (an analogous representation holds for non-tree networks). For a tree network $\mathcal{G}_n$ with n vertices (viewed as a directed graph with edges pointing from offspring to parent), and $v \in \mathcal{G}_n$, let $P_l(v, n)$ denote the number of *directed* paths of length l that end at v in $\mathcal{G}_n$. It is easy to check from Definition 2.1 that the PageRank scores have the explicit formulae for any vertex v ,

$$\mathfrak{R}_{v,c} = \frac{(1-c)}{n} \left(1 + \sum_{l=1}^{\infty} c^l P_l(v, n) \right). \quad (5.1)$$

For the sequel, it will be easier to formulate results in terms of the *graph normalized* PageRank scores $\{R_{v,c} : v \in \mathcal{G}_n\} = \{n\mathfrak{R}_{v,c}(n) : v \in \mathcal{G}_n\}$.

A natural question that one could ask is whether PageRank and degree truly differ at the *extremal* (or large deviations) level. Namely, how do the identities of the most central vertices in the network quantified via degree and PageRank compare? Along this line of enquiry, a popular belief is the so-called *power-law hypothesis*, which conjectures that for real world networks with a power-law degree distribution, the PageRank also has a power-law distribution with the same exponent as the degree. This has been shown to hold in several *static* network models (no time evolution) like the directed configuration model [27, 51] and the inhomogeneous random digraph [46, 51]. For dynamic random graphs, the picture was recently shown to be very different. For linear preferential attachment models, [15] showed that the PageRank distribution has heavier tails than degree. This phenomenon has been extended to a variety of other dynamic random graph models in [10, 13]. Strikingly, for the random recursive tree described in Section 3.6.1, the limiting degree distribution has an exponentially decaying tail while the limiting PageRank distribution has a power-law tail with exponent $1/c$.

To prove such results, local weak convergence serves as a crucial tool. At a high level, the idea is sketched as follows.

(a) **Enter local weak convergence:** Observe from (5.1) that the PageRank of a vertex v depends only on its *in-component*, namely the subgraph spanned by vertices with a directed path to v . Moreover, although the PageRank of v depends on the whole in-component of v , the contribution of a vertex u to the PageRank of v decays exponentially with its distance from v . From these observations, it is not too hard to believe that if the in-component of a uniformly chosen vertex in the network converges in a (directed) local weak sense to a rooted limiting random tree $\mathcal{G}_\infty$, then the limiting PageRank distribution (limit of the empirical distribution of PageRanks of vertices in $\mathcal{G}_n$ as $n \rightarrow \infty$) corresponds to the law of the root PageRank in $\mathcal{G}_\infty$. This was rigorously verified in [15, 35].

(b) **PageRank on static graphs:** The distribution of the root PageRank in $\mathcal{G}_\infty$ is qualitatively different for static and dynamic random graphs. For static graphs, the (directed) local weak limit corresponds to the progeny tree of a Galton-Watson Branching process. This is not a surprise, given our discussion in Section 3.6.2. This implies that the root PageRank $\mathcal{R}_\phi$ in the limiting tree satisfies the *recursive distributional fixed point equation*

$$\mathcal{R}_\phi \stackrel{d}{=} c \sum_{i=1}^{\mathcal{D}} \mathcal{R}_i + (1 - c), \quad (5.2)$$

where $\mathcal{D}$ is the (in-)degree of the root and $\{\mathcal{R}_i\}$ are iid having the same distribution as $\mathcal{R}$. This equation can then be analyzed via renewal theory to establish that

$$0 < \liminf_{x \rightarrow \infty} \frac{\mathbb{P}[\mathcal{R}_\phi > x]}{\mathbb{P}[\mathcal{D} > x]} \leq \limsup_{x \rightarrow \infty} \frac{\mathbb{P}[\mathcal{R}_\phi > x]}{\mathbb{P}[\mathcal{D} > x]} < \infty.$$

A detailed proof can be found in [41]. This line of argument verifies the power law hypothesis for a variety of static random graphs.

(c) **PageRank on dynamic graphs:** For dynamic graphs, as previously indicated, the local weak limits turn out to be more exotic. For a concrete example, consider the *non-uniform random recursive tree* introduced in Section 3.6.1. As previously discussed, the discrete network process can be embedded in a *continuous time* branching processes $\{\text{BP}(t) : t \geq 0\}$ where each individual in the population reproduces (independently across individuals) at times given by a point process $\zeta(\cdot)$ on $\mathbb{R}_+$ with intensity measure $\mu(\cdot)$ and newly born individuals start reproducing according to an independent copy of the same point process. In the point process $\zeta(\cdot)$, the time gap between the k -th and $(k+1)$ -th birth has distribution $\exp(f(k))$, where f is the attachment function governing the network evolution. The directed local weak limit is then given by $\text{BP}(T_\lambda)$, where $T_\lambda \sim \exp(\lambda)$ is independent of $\text{BP}(\cdot)$ and λ is the *Malthusian rate* of growth of the branching process, given analytically (under some regularity assumptions) as the unique root of the equation

$$\int_0^\infty e^{-\lambda t} \mu(dt) = 1.$$

See [55] for a detailed treatment. Observe that different vertices are born at different times and hence, at any time t , the distribution of the number of offspring of each existing child of the root differ across children. It is this *temporal inhomogeneity* that is behind the fundamentally different behavior of PageRank in dynamic networks as one cannot exploit a recursive distributional equation analogous to (5.2). However, there is a different way to look at things in terms of *percolation* on branching processes described below.

Definition 5.1 (Percolation on BP). *Fix a damping factor $c \in (0, 1)$. For any $t \geq 0$, write $\text{BP}^c(t)$ for the connected cluster of the root (which is also a tree) when we retain each edge $e \in \text{BP}(t)$ with probability c and delete with probability $(1 - c)$, independently across edges. Write $\{\text{BP}^c(t) : t \geq 0\}$ for the corresponding non-decreasing rooted tree valued Markov process where children born to vertices in the connected cluster of the root are retained with probability c at their time of birth.*

By (5.1), it then follows that

$$\mathcal{R}_\phi \stackrel{d}{=} (1 - c) \mathbb{E} \left(|\text{BP}^c(T_\lambda)| \mid \text{BP}(T_\lambda) \right). \quad (5.3)$$

It turns out that $\text{BP}^c(\cdot)$ is a continuous time branching process in its own right with Malthusian rate λ^c (say). Then, heuristically, recalling that $|\text{BP}(t)| \sim e^{\lambda t}$, as $x \rightarrow \infty$,

$$\Pr[\mathcal{R}_\phi > x] \sim \Pr[e^{\lambda^c T_\lambda} > x] \sim \Pr \left[T_\lambda > \frac{1}{\lambda^c} \log x \right] = x^{-\lambda/\lambda^c}.$$

This idea has been rigorously laid out in a number of cases. We present a result from [15] summarizing this for the linear preferential attachment case.

Theorem 5.2. *Consider the linear preferential attachment tree process $\{\mathcal{G}_n : n \geq 0\}$ with attachment function $f(k) = k + 1 + \beta$, $k \geq 0$, for some fixed $\beta > 0$. The in-degree and PageRank of a uniformly chosen vertex V_n in $\mathcal{G}_n$ jointly converge in distribution:*

$$(D_{V_n}, R_{V_n}) \xrightarrow{d} (\mathcal{D}, \mathcal{R}_\phi) \text{ as } n \rightarrow \infty.$$

Moreover, there exists $C > 0$ such that as $k \rightarrow \infty$,

$$\Pr[\mathcal{D} \geq k] = Ck^{-2-\beta} (1 + O(k^{-1})),$$

and positive constants C_1, C_2 such that for any $r \geq 1$,

$$C_1 r^{-(2+\beta)/(1+(1+\beta)c)} \leq \Pr[\mathcal{R}_\phi \geq r] \leq C_2 r^{-(2+\beta)/(1+(1+\beta)c)}.$$

More examples are discussed in the next few Subsections.

We remark here that, although the above discussion was presented for tree networks, the ideas extend naturally to *locally tree-like* networks. For such networks, the local weak limits are trees and the arguments sketched in (b) and (c) above apply to the limiting objects. Local weak limits for non-tree linear preferential attachment models are described in [15, 18, 34], and limits for collapsed branching processes are obtained in [14].

5.2. Attributed network models. Next recall the nodal attribute model defined in Section 2.1. The following paraphrases some of the main results in [10].

Theorem 5.3 ([10]). *Consider the linear ($\gamma \equiv 1$) setting of the model with finite attribute space and assume the propensity kernel $\kappa > 0$ (i.e. every entry of this matrix is strictly positive). Then:*

- (a) *While the model cannot be embedded directly as a continuous time, multitype, branching process, its evolution can be analyzed using stochastic approximation techniques to show that the sequence of trees converges to a limiting infinite sin-tree with nodal types.*
- (b) *This result gives information on joint distribution asymptotics of types and degrees, showing that tail exponents of the limiting degree distribution of different types can depend on the type.*
- (c) *The asymptotic limit also gives information about the PageRank distribution, in particular showing that the extremal behavior of the PageRank scores do not depend on types.*
- (d) *These limit results and constructions related to these results also give information on the behavior of various network sampling algorithms. In particular, it is shown in some settings with rare minority vertices that sampling from the graph with probability proportional to PageRank scores, which can be accomplished using local exploration schemes, has a quantifiably higher chance of sampling minority vertices in comparison to uniform and degree-based sampling schemes.*

5.3. Co-evolving networks. One major frontier, especially for developing rigorous understanding of proposed models, are the so-called co-evolutionary (or adaptive) networks, where specific dynamics (e.g. random walk explorations) on the network influence the structure of the network, which in turn influences the dynamics; thus both modalities (dynamics on the network and the network itself) co-evolve [11, 37, 56, 57]. Motivated by the growth of social networks, there has been significant interest in trying to understand the influence of processes such as search engines or influence ranking mechanisms in the growth of networks. A number of papers [21, 26, 53] have explored the dynamic evolution of networks through new nodes first exploring neighborhoods of randomly selected vertices before deciding on whom to connect. We now describe a specific class of such models.

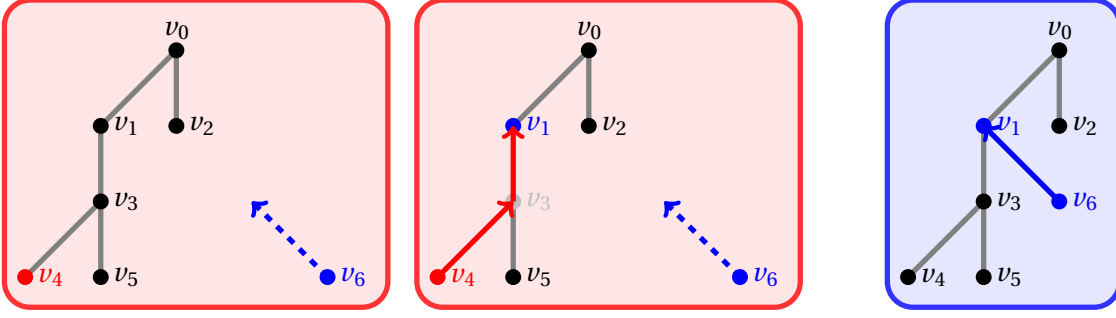


FIGURE 5.1. v_6 is a new incoming vertex, and selects v_4 to start exploring the network, with sampled number of exploration steps $Z_6 = 2$.

Fix a probability mass function $\mathbf{p} := \{p_k : k \geq 0\}$ on $\mathbb{Z}_+$. For the rest of the paper, let $\mathbf{Z} = \{Z_1, Z_2, \dots\}$ be an *i.i.d* sequence with distribution $\mathbf{p}$. We now describe the recursive construction of a sequence of random trees $\{\mathcal{T}_n : n \geq 1\}$, always rooted at vertex $\{v_0\}$, with edges pointed from descendants to their parents. Start with two vertices $\{v_0, v_1\}$, with $\mathcal{T}_1$ a rooted tree at $\{v_0\}$, an oriented edge from v_1 to v_0 . Assume for some $n \geq 1$, we have constructed $\mathcal{T}_n$. Then to construct $\mathcal{T}_{n+1}$:

- (a) New vertex $\{v_{n+1}\}$ enters the system at time $n + 1$.
- (b) This new vertex selects a vertex V_n , *uniformly at random*, amongst the existing vertices $\mathcal{V}(\mathcal{T}_n) = \{v_0, \dots, v_n\}$.
- (c) Let $\mathcal{P}(v_0, V_n)$ denote the path from the root to this vertex. This new vertex traverses up this path for a random length Z_{n+1} and attaches to the terminal vertex. If the graph distance to the root, $\text{dist}(v_0, V_n) \leq Z_{n+1}$ then this new vertex attaches to the root v_0 .

It turns out, the setting where the pmf $\mathbf{p}$ satisfies $p_0 + p_1 = 1$ is identical to the linear preferential attachment model. So to state our main results, we will assume that $p_0 + p_1 < 1$ and further $\mathbb{E}(Z) < \infty$. To get some intuition consider the simulation figures below (network size $n = 30,000$) and $\mathbf{Z}$ having a Geometric distribution with parameter $p = .35, .65$ respectively.

When $p = .35$, this implies $\mathbb{E}(Z) > 1$ namely the exploration walk tends to explore further up the path to the root, whereas when $p = .65$ then $\mathbb{E}(Z) \leq 1$, the walk is more local in nature. It turns out $\mathbb{E}(Z) = 1$ is the “phase transition” point; we paraphrase a few results from [13].

Theorem 5.4 ([13]). *For the above network model:*

- (a) When $\mathbb{E}(Z) \leq 1$ then the sequence of random trees converges to a limiting infinite random sin-tree $\mathcal{T}_n \xrightarrow{\text{a.s.}} \mathcal{T}_\infty$ from which one can derive information about various quantities including asymptotics for the degree distribution and the corresponding degree exponents. This has deep connections with large deviations and quasi-stationary distributions of random walks on $\mathbb{Z}$.
- (b) When $\mathbb{E}(Z) > 1$ then $\{\mathcal{T}_n : n \geq 1\}$ converges in the fringe sense (Def 3.7 (a)) but not in the extended fringe sense. Further there is condensation at the root in the sense that there is a model dependent strictly positive constant $\gamma(\mathbf{p}) > 0$ such that the root degree $\deg(\rho, n)$ in $\mathcal{T}_n$ satisfies, $\deg(\rho, n)/n \xrightarrow{P} \gamma(\mathbf{p})$ as $n \rightarrow \infty$.
- (c) We noted before that the power-law hypothesis (comparing PageRank and degree tail behavior) holds for several static graphs and fails to hold for several dynamic graphs. The model under consideration ‘interpolates’ these two regimes in the following sense. When $\mathbb{E}[Z] \leq 1$, then (under some regularity assumptions), there exist $R \geq 1$ and $s(\mathbf{p}) \in (0, 1)$ such that the limiting degree and

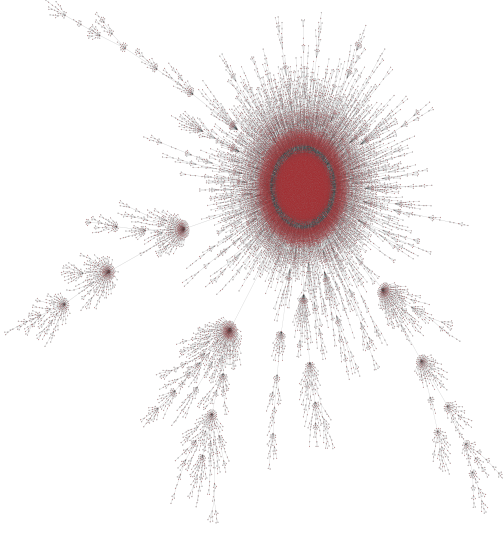


FIGURE 5.2. $\mathbf{Z} \sim$ Geometric distribution with $p = .35$

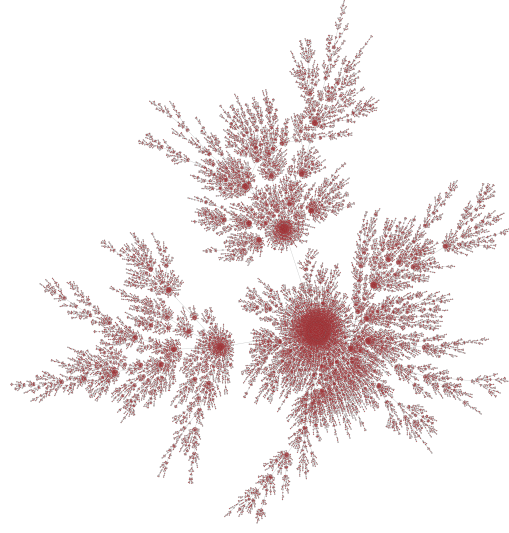


FIGURE 5.3. $\mathbf{Z} \sim$ Geometric distribution with $p = .65$

PageRank distributions satisfy:

$$\lim_{k \rightarrow \infty} \frac{\log(\mathbb{P}(\mathcal{D} \geq k))}{\log k} = -R,$$

$$\lim_{r \rightarrow \infty} \frac{\log(\mathbb{P}(\mathcal{R}_\phi \geq r))}{\log r} = \begin{cases} -R & \text{for } c \in (0, s(\mathbf{p})] \text{ with } c < 1, \\ -\frac{1}{cf(1/c)} & \text{for } c \in (s(\mathbf{p}), 1). \end{cases}$$

Hence, by varying the damping factor, one transitions from the regime where the power-law hypothesis holds to one where PageRank has strictly heavier tails than the degree distribution. Moreover, one can verify that $\frac{1}{s(\mathbf{p})f(1/s(\mathbf{p}))} = R$, and thus this transition is continuous.

6. CONCLUSION AND FURTHER READING

As mentioned in the introduction, this paper is not meant to be a survey and rather was meant to be a gentle introduction to junior researchers in the use of local weak convergence in various problems of probabilistic combinatorics; still we apologize to the many wonderful researchers who have contributed to this methodology and its use at the frontiers of the field that we were not able to cite. Our goal now is to give starting points for such readers to take next steps in their exploration of this vast field and get closer to the edge of research and the vast unexplored terrain of the unknown.

For such researchers, perhaps the most comprehensive survey can be found in [62, Chapter 2] which outlines all the myriad fundamentals and their extensions, especially in the context of unweighted graphs, in great detail. In the geometric setting, for the fundamentals and their use in probabilistic combinatorial optimization, [6] is a phenomenal resource. We also urge readers to peruse [3] which laid the genesis for many of these concepts via considering the setting of trees and which also explains and connects these topics to the more classical findings related to stable age distribution theory of continuous time branching processes by Jagers and Nerman.

For statistical physics inspired models such as the Ising model on sparse random graphs, the two summer school lecture notes [30, 61] are wonderful resources for rigorous theory, while [48] gives

an overview of the intuition, especially from statistical physics for what “should happen”. One major tool that seemed to show up repeatedly in Section 4 and Section 5 were recursive distributional equations and [8] gives a wide-ranging overview of rigorous theory of such constructs.

Finally for random matrix theory and its connections to local weak convergence, the various lecture notes of Charles Bordenave e.g. [23] are a great starting point, while for planar graphs, the St. Flour lecture notes of Asaf Nachmias [50] provide a great diving board into the ramifications of one of the classics in this entire field, namely the work of Benjamini and Schramm [17].

ACKNOWLEDGEMENT

Banerjee was supported in part by the NSF CAREER award DMS-2141621. Bhamidi was supported in part by NSF DMS-2113662. Banerjee, Bhamidi and Young were partially funded by NSF RTG grant DMS-2134107. We thank David Aldous, Remco van der Hofstad, Mariana Olvera-Cravioto and Allan Sly for many insights over the years.

REFERENCES

- [1] *A different kind of mind* K.R. Parthasarathy in conversation with B.V. Rajarama Bhat, Bhavana: The mathematics magazine **3** (2019).
- [2] E. Abbe, *Community detection and stochastic block models: recent developments*, Journal of Machine Learning Research **18** (2018), no. 177, 1–86.
- [3] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, The Annals of Applied Probability (1991), 228–266.
- [4] D. Aldous, *Asymptotic fringe distributions for general families of random trees*, The Annals of Applied Probability (1991), 228–266.
- [5] D. Aldous and J. A. Fill, *Reversible Markov chains and random walks on graphs*, 2002. Unfinished monograph, recompiled 2014, available at <http://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [6] D. Aldous and J. M. Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, 2004, pp. 1–72.
- [7] D. J. Aldous, *The $\zeta(2)$ limit in the random assignment problem*, Random Structures & Algorithms **18** (2001), no. 4, 381–418.
- [8] D. J. Aldous and A. Bandyopadhyay, *A survey of max-type recursive distributional equations*, The Annals of Applied Probability **15** (2005), no. 2, 1047–1110.
- [9] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, 2005.
- [10] N. Antunes, S. Banerjee, S. Bhamidi, and V. Pipiras, *Attribute network models, stochastic approximation, and network sampling and ranking algorithms*, arXiv preprint arXiv:2304.08565 (2023).
- [11] T. Aoki, L. E. Rocha, and T. Gross, *Temporal and structural heterogeneities emerging in adaptive temporal networks*, Physical Review E **93** (2016), no. 4, 040301.
- [12] K. B. Athreya and S. Karlin, *Embedding of urn schemes into continuous time markov branching processes and related limit theorems*, The Annals of Mathematical Statistics **39** (1968), no. 6, 1801–1817.
- [13] S. Banerjee, S. Bhamidi, and X. Huang, *Co-evolving dynamic networks*, arXiv preprint arXiv:2203.11877 (2022).
- [14] S. Banerjee, P. Deka, and M. Olvera-Cravioto, *Local weak limits for collapsed branching processes with random out-degrees*, arXiv preprint arXiv:2302.00562 (2023).
- [15] S. Banerjee and M. Olvera-Cravioto, *Pagerank asymptotics on directed preferential attachment networks*, The Annals of Applied Probability **32** (2022), no. 4, 3060–3084.
- [16] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, Science **286** (1999), no. 5439, 509–512.
- [17] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Selected Works of Oded Schramm, 2011, pp. 533–545.
- [18] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi, *Asymptotic behavior and distributional limits of preferential attachment graphs*, The Annals of Probability **42** (2014), no. 1, 1–40.
- [19] S. Bhamidi, S. N. Evans, and A. Sen, *Spectra of large random trees*, Journal of Theoretical Probability **25** (2012), no. 3, 613–654.
- [20] P. Billingsley, *Convergence of probability measures*, John Wiley & Sons, 2013.

- [21] A. Blum, T. H. Chan, and M. R. Rwebangira, *A random-surfer web-graph model*, 2006 Proceedings of the Third Workshop on Analytic Algorithmics and Combinatorics (ANALCO), 2006, pp. 238–246.
- [22] B. e. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, *The degree sequence of a scale-free random graph process*, Random Structures & Algorithms **18** (2001), no. 3, 279–290.
- [23] C. Bordenave, *Spectrum of random graphs*, Advanced topics in random matrices **53** (2016), 91–150.
- [24] C. Bordenave and M. Lelarge, *Resolvent of large random graphs*, Random Structures & Algorithms **37** (2010), no. 3, 332–352.
- [25] C. Bordenave, M. Lelarge, and L. Massoulié, *Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs*, 2015 IEEE 56th Annual Symposium on Foundations of Computer Science, 2015, pp. 1347–1357.
- [26] P. Chebolu and P. Melsted, *Pagerank and the random surfer model.*, SODA, 2008, pp. 1010–1018.
- [27] N. Chen, N. Litvak, and M. Olvera-Cravioto, *Generalized pagerank on directed configuration networks*, Random Structures & Algorithms **51** (2017), no. 2, 237–274.
- [28] D. Coppersmith and G. B. Sorkin, *Constructive bounds and exact expectations for the random assignment problem*, Random Structures & Algorithms **15** (1999), no. 2, 113–144.
- [29] D. J. Daley, D. Vere-Jones, et al., *An introduction to the theory of point processes: volume i: elementary theory and methods*, Springer, 2003.
- [30] A. Dembo and A. Montanari, *Gibbs measures and phase transitions on sparse random graphs*, Brazilian Journal of Probability and Statistics **24** (2010), no. 2, 137–211.
- [31] A. Dembo and A. Montanari, *Ising models on locally tree-like graphs*, The Annals of Applied Probability (2010), 565–592.
- [32] L. Espín-Noboa, C. Wagner, F. Karimi, and K. Lerman, *Towards quantifying sampling bias in network inference*, Companion proceedings of the the web conference 2018, 2018, pp. 1277–1285.
- [33] L. Espín-Noboa, C. Wagner, M. Strohmaier, and F. Karimi, *Inequality and inequity in network-based ranking and recommendation algorithms*, Scientific reports **12** (2022), no. 1, 1–14.
- [34] A. Garavaglia, R. S. Hazra, R. van der Hofstad, and R. Ray, *Universality of the local limit of preferential attachment models*, arXiv preprint arXiv:2212.05551 (2022).
- [35] A. Garavaglia, R. van der Hofstad, and N. Litvak, *Local weak convergence for pagerank*, The Annals of Applied Probability **30** (2020), no. 1, 40–79.
- [36] M. X. Goemans and M. S. Kodialam, *A lower bound on the expected cost of an optimal assignment*, Mathematics of Operations Research **18** (1993), no. 2, 267–274.
- [37] T. Gross and B. Blasius, *Adaptive coevolutionary networks: a review*, Journal of the Royal Society Interface **5** (2008), no. 20, 259–271.
- [38] P. Jagers, *Branching processes with biological applications*, Wiley-Interscience [John Wiley & Sons], London-New York-Sydney, 1975. Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics. MR0488341
- [39] P. Jagers and O. Nerman, *The growth and composition of branching populations*, Adv. in Appl. Probab. **16** (1984), no. 2, 221–259. MR742953 (86j:60193)
- [40] P. Jagers and O. Nerman, *Limit theorems for sums determined by branching and other exponentially growing processes*, Stochastic Process. Appl. **17** (1984), no. 1, 47–71. MR738768 (86b:60138)
- [41] P. R. Jelenković and M. Olvera-Cravioto, *Information ranking and power laws on trees*, Advances in Applied Probability **42** (2010), no. 4, 1057–1093.
- [42] J. Jordan, *Geometric preferential attachment in non-uniform metric spaces*, Electronic Journal of Probability **18** (2013), 1–15.
- [43] O. Kallenberg et al., *Random measures, theory and applications*, Vol. 1, Springer, 2017.
- [44] F. Karimi, M. Génois, C. Wagner, P. Singer, and M. Strohmaier, *Homophily influences ranking of minorities in social networks*, Scientific Reports **8** (2018), no. 1, 11077.
- [45] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, *Spectral redemption in clustering sparse networks*, Proceedings of the National Academy of Sciences **110** (2013), no. 52, 20935–20940.
- [46] J. Lee and M. Olvera-Cravioto, *Pagerank on inhomogeneous random digraphs*, Stochastic Processes and their Applications **130** (2020), no. 4, 2312–2348.
- [47] M. L. Mehta, *Random matrices*, Elsevier, 2004.
- [48] M. Mezard and A. Montanari, *Information, physics, and computation*, Oxford University Press, 2009.
- [49] M. Mézard and G. Parisi, *On the solution of the random link matching problems*, Journal de Physique **48** (1987), no. 9, 1451–1459.

- [50] A. Nachmias, *Planar maps, random walks and circle packing: École d'été de probabilités de saint-flour xlviii-2018*, Springer Nature, 2020.
- [51] M. Olvera-Cravioto, *Pagerank's behavior under degree correlations*, Annals of Applied Probability **3** (2021), no. 1, 1403–1442.
- [52] L. Page, S. Brin, R. Motwani, and T. Winograd, *The pagerank citation ranking: Bringing order to the web.*, Stanford InfoLab, 1999.
- [53] G. Pandurangan, P. Raghavan, and E. Upfal, *Using pagerank to characterize web structure*, International computing and combinatorics conference, 2002, pp. 330–339.
- [54] K. R. Parthasarathy, *Probability measures on metric spaces*, Vol. 352, American Mathematical Soc., 2005.
- [55] A. Rudas, B. Tóth, and B. Valkó, *Random trees and general branching processes*, Random Structures & Algorithms **31** (2007), no. 2, 186–202.
- [56] H. Sayama, I. Pestov, J. Schmidt, B. J. Bush, C. Wong, J. Yamanoi, and T. Gross, *Modeling complex systems with adaptive networks*, Computers & Mathematics with Applications **65** (2013), no. 10, 1645–1664.
- [57] H. Sayama and R. Sinatra, *Social diffusion and global drift on networks*, Physical Review E **91** (2015), no. 3, 032809.
- [58] R. T. Smythe and H. M. Mahmoud, *A survey of recursive trees*, Theory of Probability and Mathematical Statistics **51** (1995), 1–28.
- [59] J. M. Steele, *Probability theory and combinatorial optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 69, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. MR1422018
- [60] J. Szymański, *On a nonuniform random recursive tree*, North-holland mathematics studies, 1987, pp. 297–306.
- [61] R. van der Hofstad, *Stochastic processes on random graphs*, Lecture notes for the 47th Summer School in Probability Saint-Flour **2017** (2017).
- [62] R. Van Der Hofstad, *Random graphs and complex networks, Volume ii*, Cambridge university press, 2023.
- [63] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*, Vol. 47, Cambridge university press, 2018.
- [64] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, Vol. 48, Cambridge university press, 2019.