The Objective Method: Probabilistic Combinatorial Optimization and Local Weak Convergence

David Aldous
Department of Statistics
University of California Berkeley
Berkeley, CA, 94720-3860
aldous@stat.berkeley.edu

J. Michael Steele
Department of Statistics
Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6302
steele@wharton.upenn.edu

December 19, 2002

1 Introduction

This survey describes a general approach to a class of problems that arise in combinatorial probability and combinatorial optimization. Formally, the method is part of weak convergence theory, but in concrete problems the method has a flavor of its own. A characteristic element of the method is that it often calls for one to introduce a new, infinite, probabilistic *object* whose *local properties* inform us about the *limiting properties* of a sequence of finite problems.

The name *objective method* hopes to underscore the value of shifting ones attention to the new, large random object with fixed distributional properties and way from the sequence of objects with changing distributions. The new object always provides us with some new information on the asymptotic behavior of the original sequence, and, in the happiest cases, the constants associated with the infinite object even permit us to find the elusive limit constants for that sequence.

1.1 A Motivating Example: the Assignment Problem

The assignment problem for the $n \times n$ cost matrix (c_{ij}) is the task of determining the permutation $\pi : [n] \to [n]$ that minimizes the total cost $c_{1,\pi(1)} + c_{2,\pi(2)} + \cdots + c_{n,\pi(n)}$ of assigning one column to each row. This problem arises in many applied contexts, and it has spawned an extensive algorithmic literature. Here we are more specifically concerned with the value of the objective function

$$A_n = \min_{\pi} \sum_{i=1}^{n} c_{i,\pi(i)},$$

where the costs c_{ij}, $1 \le i, j \le n$, are assumed to be independent and identically distributed random variables.

Investigation of A_n seems to have begun in 1962, when Kurtzberg [42] used heuristic methods like greedy matching to obtain upper bounds on $E[A_n]$ for uniformly distributed costs. Kurtzberg's bounds were of order $\log n$, and these were not improved until 1979 when Walkup [64] showed that $E[A_n]$ is bounded independently of n, which was quite a surprising discovery at the time. Eight years later, Karp [38] introduced a new approach to the estimation of $E[A_n]$ that was based on linear programming. Karp exploited the explicit bases that were known to be optimal for the assignment problem, and he obtained the elegant bound $E[A_n] \le 2$. Inspired in part by Karp's result, Dyer, Frieze and McDiarmid [19] developed a general bound for the objective function of linear programming problems with random costs, and they were able to recapture Karp's bound without recourse to special bases. A probabilist's interpretation of the Dyer-Frieze-McDiarmid inequality forms the basis of Chapter 4 of Steele [60] where one can find further information on the early history of the assignment problem with random costs.

A new period in the development of the random assignment problem began in 1987 with the fascinating article of Mézard and Parisi [50] which offered a non-rigorous statistical mechanical argument for the assertion that

$$\lim_{n\to\infty} E[A_n] = \frac{\pi^2}{6} = \zeta(2). \tag{1.1}$$

The desire for a rigorous proof of this limit has influenced much of the subsequent work on the assignment problem, and the critical first step is simply to show that actually $E[A_n]$ converges as $n \to \infty$. Convergence would follow immediately if one could show that $E[A_n]$ is nondecreasing, but it is still not known if $E[A_n]$ is monotone. Nevertheless, in 1992 Aldous [3] used the objective method to show that $E[A_n]$ does indeed converge.

In 1998 Parisi [51] added further interest to the random assignment problem when he advanced the remarkable conjecture that for independent exponentially distributed random cost c_{ij} with mean 1, one has an *exact formula*:

$$E[A_n] = \sum_{k=1}^{n} k^{-2}. \tag{1.2}$$

A proof of Parisi's conjecture would immediately confirm the $\zeta(2)$ limit formula; moreover, there are good reasons to believe the truth of this stronger assertion. Specifically, Alm and Sorkin [9] proved the conjecture for all values of n up to $n = 5$. Unfortunately, even for such small values, the proofs are not easy, and Alm and Sorkin had to find their way through a swamp of special cases with the help of automated theorem proving computations. The conjectured formula for $E[A_n]$ remains a tantalizing challenge.

Nevertheless, progress on the limit value for $E[A_n]$ has been more definitive. In 2001 — by means of the objective method — Aldous [4] finally proved the $\zeta(2)$ limit formula that Mézard and Parisi [50] first brought to light in 1987.

1.2 A Stalking Horse: the Partial Matching Problem

One of the aims of this survey is to show how the objective method helps to solve problems such as the determination of the limiting value of $E[A_n]$, but the assignment problem itself is too burdened with individual nuance for it to serve as our basic guide. For this reason, we introduce a new problem, the *maximal partial matching problem*.

This problem does not have the long history of the random assignment problem, but it is especially well suited for illustrating the objective method. In particular, it builds on the theory of random trees, and the limit theory of such trees provides the prototype for the objective method. Also, the maximal partial matching problem leads inexorably to the notion of a distributional identity. Such identities have always had a role in probability theory, but, along with allied notions like self-similarity and subadditivity, distributional identities are now a topic of emerging importance.

The relevant identities for the maximal partial matching problem are also much more tractable than the corresponding identities for the random assignment problem, yet many of the same tools come into play. In particular, we will find that one often does well to begin by guessing the solution of a distributional identity. After one has a good guess, then classic tools like Kolmogorov's consistency theorem and contemporary tricks like "coupling from the past" can be used to confirm the existence, uniqueness, and stability of the solutions of the distributional identity.

1.3 Organization of the Survey

The maximal partial matching problem provides our introductory case study, but before it can be addressed we need to deal with some foundational issues. In particular, Section 2 develops the notion of *local weak convergence*. Formally, this is nothing more than the weak convergence of probability measures on a certain metric space of "rooted geometric graphs," but the attending intuition differs substantially from the classic weak convergence theory. Section 2 also introduces the "standard construction," which is a general recipe for building rooted geometric graphs. This construction turns out to have a subtle, yet pervasive, influence on the applications of the objective method.

After we have dealt with the essential metric space formalities, we see what local weak convergence tells us about the *simplest* model for random trees. In particular, we review Grimmett's lemma, and we develop an understanding of the convergence of the large and the small parts of a random tree that has been cut at a randomly chosen edge. This analysis provides us with the essential distribution theory for random trees that is needed later.

The theory of the maximum partial matching problem developed in Section 3 is concrete and self-contained. Nevertheless, it faces most of the issues that one meets in more complex applications of the objective method, and it offers the best introduction we know to the essential ideas.

In Section 4 we introduce the mean-field model of distance which is a physically motivated probability model designed to gain insight into problems for point processes in $\mathbb{R}^d$. In this model the distribution of some of inter-point distances of $\mathbb{R}^d$ are captured precisely while other (hopefully less essential) inter-point distance distributions are distorted in comparison to $\mathbb{R}^d$. The mean-field model leads us to the PWIT, or Poisson weighted infinite tree, which is arguably the most important infinite tree model. To illustrate the close connection of the theory of the PWIT and the objective method, we give a reasonably detailed proof of the $\zeta(3)$ theorem of Frieze.

The relationship of the PWIT to problems of combinatorial optimization is continued in Section 5 by developing the limit theory for the minimum cost C_n of a perfect matching of the complete graph K_n with independent edge weights ξ_e having a common distribution F. We provide a reasonably detailed (and hopefully friendly) sketch of the fact that $E[C_n]$ converges to $\zeta(2)/2$ as $n \to \infty$ when F is the exponential distribution with mean one; this result is the direct analog of the $\zeta(2)$ limit theorem for $E[A_n]$. Understandably, some non-trivial details must be omitted from the sketch, but the section should still provide a useful introduction to Aldous's proof of the $\zeta(2)$ limit theorem.

All of the problems in Sections 3 – 5 call on distance models with substantial intrinsic independence; for example, in Section 3 we have independent weights on every edge, and we face the infinite Poisson-Galton-Watson tree in the limit, while in Sections 4 and 5, we have the tempered independence that one inherits from the mean-field model, and we face the PWIT in the limit. The problems of Section 6 are different; they deal directly with inter-point distances in $\mathbb{R}^d$ rather than with edge weights that are blessed *ex cathedra* with independence properties, and in the limit we often need to deal with the Poisson process.

The probability theory of Euclidean combinatorial optimization has grown quite large, and in Section 6 we provide very few proofs. Nevertheless, we hope to provide a useful up-date of the survey of Steele [60]. In particular, we address progress on the minimal spanning tree problem, including a recent weak law of large numbers developed in Penrose and Yukich [53] which has close ties to the objective method. The section also contrasts the objective method and the subadditive method that has served for years as a principal workhorse in the probability theory of Euclidean combinatorial optimization. In the closing subsection, we break into entirely new territory and describe a remarkable new result of Benjamini and Schramm on the recurrence properties of the local weak

limits of planar graphs.

In Section 7, the last section, we first summarize some of the circumstances that seem to be needed for the objective method to be successful. We then develop the background of an attractive conjecture on the independence number of a random regular graph. This problem is used in turn to illustrate several of the basic challenges that appear to be critical to the deeper development of the objective method.

Finally, we should note that even though our main intention is to provide a survey and a tutorial, this exposition also contains new results. In particular, the material in Section 3 on the maximum partial matching problem is new, including the basic limit theorem (Theorem 1) and the theorem that determines the limit constants (Theorem 2). Another new result is the Convergence Theorem for Minimal Spanning Trees (Theorem 6). This theorem both generalizes and simplifies much earlier work; it is also applied several times during the course of the survey.

2 Geometric Graphs and Local Weak Convergence

Before the theory of local weak convergence can be brought face-to-face with the concrete problems of combinatorial optimization, one is forced to introduce an appropriate complete metric space. This introduction has been made several times before on a purely *ad hoc* basis, but now there is enough cumulative experience to suggest a general framework that should suffice for most applications. After describing this framework, we will give it a quick test run by discussing a prototypical result from the limit theory of random trees — Grimmett's lemma on the convergence of rooted Cayley trees.

2.1 Geometric Graphs

If $G = (V, E)$ is a graph with a finite or countable vertex set V and a corresponding edge set E, then any function $\ell : E \to (0, \infty]$ can be used to define a distance between vertices of G. Specifically, for any pair of vertices u and v, one just takes the distance between u and v to be the infimum over all paths between u and v of the sum of the lengths of the edges in the path.

Definition 1 (Geometric Graphs and the Two Classes $\mathcal{G}$ and $\mathcal{G}_\star$)
If $G = (V, E)$ is a connected, undirected graph with a countable or infinite vertex set V and if ℓ is an edge length function that makes G **locally finite** *in the sense that for each vertex v and each real $\varrho < \infty$ the number of vertices within distance ϱ from v is finite, then G is called a* **geometric graph**. *When there is also a distinguished vertex v, we say that G is a* **rooted geometric graph** *with* root v. *The set of geometric graphs will be denoted by $\mathcal{G}$, and the set of rooted geometric graphs will be denoted by $\mathcal{G}_\star$.*

2.2 $\mathcal{G}_\star$ as a Metric Space

The set $\mathcal{G}_\star$ of rooted geometric graphs provides our basic workspace, but before honest work can begin we need to say what we mean for a sequence G_n of elements of $\mathcal{G}_\star$ to converge to a G in $\mathcal{G}_\star$. The idea one wants to capture is that for large n, the rooted geometric graph G_n looks very much like G in an arbitrarily large neighborhood of the root of G.

To formalize this idea, we first recall that an isomorphism between graphs $G = (V, E)$ and $G' = (V', E')$ is a bijection $\phi : V \to V'$ such that $(\phi(u), \phi(v)) \in E'$ if and only if $(u, v) \in E$. Also, given any such isomorphism, one can extend the domain of ϕ to E simply by defining $\phi(e)$ to be $(\phi(u), \phi(v))$ for each $e = (u, v) \in E$.

Finally, we say that two geometric graphs $G = (V, E)$ and $G' = (V', E')$ are isomorphic provided that (1) they are isomorphic as ordinary graphs and (2) there is a graph isomorphism ϕ between G and G' that also preserves edge lengths (so $\ell'(\phi(e)) = \ell(e)$ for all $e \in E$). In the case of two *rooted* geometric graphs $G = (V, E)$ and $G' = (V', E')$, we will say they are isomorphic provided that there is a graph isomorphism ϕ that preserves edges lengths and that also maps the root of G to the root of G'.

Next we consider a special rooted geometric graph that one may view intuitively as the "neighborhood of radius ϱ about the root" of the rooted geometric graph G. Specifically, for any $\varrho > 0$ we let $N_\varrho(G)$ denote the graph whose vertex set $V_\varrho(G)$ is the set of vertices of G that are at a distance of at most ϱ from the root of G and whose edge set consists of just those edges of G that have both vertices in $V_\varrho(G)$, where, as before, the distance between any two vertices u and v in G is taken to be the infimum over all paths between u and v of the sum of the lengths of the edges in the path. We again view $N_\varrho(G)$ as an element of $\mathcal{G}_\star$ whose the edge length function and root are just those of G. Finally, we say that $\varrho > 0$ is a *continuity point* of G if no vertex of G is exactly at a distance ϱ from the root of G.

Definition 2 (Convergence in $\mathcal{G}_\star$) *We say that G_n* **converges** *to G_∞ in $\mathcal{G}_\star$ provided that for each continuity point ϱ of G_∞ there is an $n_0 = n_0(\varrho, G_\infty)$ such that for all $n \geq n_0$ there exists a isomorphism $\gamma_{n,\varrho}$ from the rooted geometric graph $N_\varrho(G_\infty)$ to the rooted geometric graph $N_\varrho(G_n)$ such that for each edge e of $N_\varrho(G_\infty)$ the length of $\gamma_{n,\varrho}(e)$ converges to the length of e as $n \to \infty$.*

With a little work, one can show that this definition determines a topology that makes $\mathcal{G}_\star$ into a complete separable metric space. As a consequence, all of the usual tools of weak convergence theory apply to sequences of probability measures on $\mathcal{G}_\star$, and we may safely use the conventional notation and write

$$\mu_n \xrightarrow{d} \mu \quad \text{to mean that} \quad \int_{\mathcal{G}_\star} f \, d\mu_n \to \int_{\mathcal{G}_\star} f \, d\mu$$

for each bounded continuous function $f : \mathcal{G}_\star \to \mathbb{R}$.

2.3 Local Weak Convergence

The topology on the metric space $\mathcal{G}_\star$ turns out to give weak convergence of probability measures on $\mathcal{G}_\star$ 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 $\mathcal{G}_\star$ never involves any rescaling, and the special role of the neighborhoods $N_\varrho(G)$ means that convergence in $\mathcal{G}_\star$ only informs us about behavior in the neighborhood of the root.

In practical terms, this means that weak convergence in $\mathcal{G}_\star$ can tell us about local features such as the degree of the root, the length of the longest edge incident to the root, and so on; yet it cannot convey detailed information on a global quantity such as the length of the longest path. To underscore this difference, one sometimes speaks of weak convergence in $\mathcal{G}_\star$ as *local weak convergence*.

2.4 The Standard Construction

Most of the random processes considered here are associated with a standard construction that has some lasting consequences, even though it may seem rather bland at first. The construction begins with a probability measure that is concentrated on the subset of $\mathcal{G}$ consisting of geometric graphs with exactly n vertices. We then consider a random element $\mathbf{G}_n$ of $\mathcal{G}$ that is chosen according to this measure, and we choose a vertex X_n at random according to the uniform distribution on the n vertices of $\mathbf{G}_n$. We then make $\mathbf{G}_n$ into a random rooted geometric graph $\mathbf{G}_n[X_n]$ by distinguishing X_n as the root vertex. The distribution of $\mathbf{G}_n[X_n]$ is then a probability measure on the set of n-vertex elements in the set of rooted geometric graphs. Finally, if a sequence $\{\mathbf{G}_n[X_n] : n = 1, 2, ...\}$ of such $\mathcal{G}_\star$-valued random variables converges weakly in $\mathcal{G}_\star$ to a $\mathcal{G}_\star$-valued random variable $\mathbf{G}_\infty$, then we say that the distribution of $\mathbf{G}_\infty$ is obtained by the *standard construction*.

One might think that virtually any measure on $\mathcal{G}_\star$ might be obtained by the standard construction, but the measures given by the standard construction are not unconstrained. Later we will find that they must satisfy a modest symmetry property that we call *involution invariance*. This property may sometimes be used to rule out what would otherwise be a tempting candidate for a limiting object.

2.5 A Prototype: The Limit of Uniform Random Trees

A classical formula of Cayley tells us that if V is a set of n distinguishable elements, then the set $\mathcal{S}_n$ of rooted trees with vertex set V has cardinality n^{n-1}. If T_n denotes a tree that is chosen from $\mathcal{S}_n$ according the uniform distribution, then one can show that T_n converges in distribution to a random variable T_∞ that takes values in $\mathcal{G}_\star$. Remarkably, one can characterize the distribution of T_∞ by a direct construction that relies on the classical theory of branching processes.

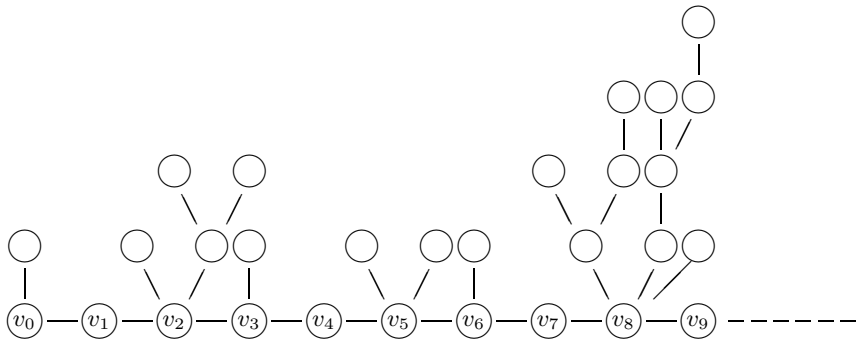

Figure 1: The "Skeleton Tree" T_∞ — also known as $\mathrm{PGW}^\infty(1)$. Here, one should note that only the root v_0 of T_∞ is labelled; the chimerical notations $v_1, v_2, ...$ are only given to focus the eye on the unique path from the root to infinity. The existence of such a path is one of the most fundamental features of $\mathrm{PGW}^\infty(1)$.

The Infinite "Skeleton Tree" T_∞ — or, $\mathrm{PGW}^\infty(1)$

To begin, we consider a Galton-Watson branching process with one progenitor and an offspring distribution that has the Poisson distribution with mean one. With probability one, any such branching process has a finite number of vertices, and the finite trees generated in this way are said to have the *Poisson Galton-Watson distribution* with mean one. This distribution on the finite elements of $\mathcal{G}_\star$ is denoted by PGW(1).

Now consider an infinite sequence of independent PGW(1) distributed trees $T_0, T_1, T_2, ...$, and let $v_0, v_1, v_2, \ldots$ denote their roots. Finally, to make this collection of trees into one infinite rooted tree, we add all of the edges $\{(v_i, v_{i+1}),\ 0 \leq i < \infty\}$, and we declare v_0 the root of the new infinite tree. The tree T_∞ that one builds this way is said to have the $\mathrm{PGW}^\infty(1)$ distribution, and it is illustrated in Figure 1.

One might imagine that T_∞ could be an undistinguished member of the multitude of infinite trees that anyone might paste together from finite trees, but this intuition is faulty. As the next lemma explains, the skeleton tree T_∞ has an inevitable — and highly distinguished — role in the theory of random trees. We should note that in the next lemma, and throughout the survey, a graph without its own edge length function is given one by taking the length of each edge to be 1.

Lemma 1 (Grimmett's Lemma) *The randomly rooted uniformly distributed labelled tree T_n on n vertices converges weakly in $\mathcal{G}_\star$ to the "skeleton tree" T_∞;*

that is, one has

$$T_n \xrightarrow{d} T_\infty \quad \textit{as } n \to \infty.$$

This result was first formalized and proved in Grimmett[26], and further proofs and generalizations are given by Devroye [18] and Aldous [2]. In some respects, Lemma 1 is the most fundamental fact about random trees under the model of uniform distribution. Grimmett's Lemma also has an interesting *self-generalizing* quality, and shortly we will see that the convergence in distribution of T_n actually implies that certain special parts of T_n must converge to corresponding special parts of T_∞.

This *convergence of parts* seems to be a recurring theme in the theory of local weak convergence. In particular, it provides one key to the appearance of the distributional identities that have had a crucial role in some of the most interesting applications of the objective method.

Special Parts of the Skeleton Tree $\mathrm{PGW}^\infty(1)$

Our first observation about the skeleton tree of Figure 1 is that it has a pleasantly recursive structure. From the root v_0 there is just one edge (v_0, v_1) that is part of the unique infinite path $v_0 \to v_1 \to \cdots$ of T_∞, and if we delete that edge we find two subtrees — one finite and one infinite. If T_∞^{small} denotes the finite tree rooted at v_0 and T_∞^{big} denotes the infinite tree rooted at v_1, then we see from the definition of T^∞ that we can identify their distributions

$$T_\infty^{small} \stackrel{d}{=} \mathrm{PGW}(1) \quad \text{and} \quad T_\infty^{big} \stackrel{d}{=} \mathrm{PGW}^\infty(1).$$

This decomposition and the definition of weak convergence of probability measures on $\mathcal{G}_\star$ promptly suggest that one may give a more detailed interpretation of Grimmett's Lemma.

Convergence of the Corresponding Parts

Consider again the randomly rooted uniform tree T_n on n vertices, and let r denote the root of T_n. Any child c of r determines an edge $e = (r, c)$, and, when e is deleted from T, we let $T_{n,c}$ denote the remaining component of T_n that contains c. We again view $T_{n,c}$ as a rooted geometric weighted graph by taking its root to be c.

Now, we let c^* denote a child of r that is chosen at random from the set of all children of the root for which the cardinality of $T_{n,c}$ is maximal; in fact, for large n, one finds that with high probability there is a unique maximal subtree. If we remove the edge (r, c^*) from T_n we get two subtrees that we now label T_n^{small} and T_n^{big} according to their cardinality (and with any tie being broken by a coin flip). Finally, we view T_n^{small} and T_n^{big} as rooted trees by letting the end point of the deleted edge determine the new roots.

Next, by our observations on the structure of the skeleton tree, we can give a more detailed view of the convergence that is guaranteed by Grimmett's Lemma. Specifically, Grimmett's Lemma automatically implies both the convergence of

the small component

$$T_n^{small} \xrightarrow{d} T_\infty^{small} \stackrel{d}{=} \mathrm{PGW}(1)$$

and convergence of the large component

$$T_n^{big} \xrightarrow{d} T_\infty^{big} \stackrel{d}{=} \mathrm{PGW}^\infty(1).$$

Moreover, Grimmett's Lemma even implies the joint convergence

$$\left(T_n^{small}, T_n^{big}\right) \xrightarrow{d} \left(T_\infty^{small}, T_\infty^{big}\right),$$

where the two processes $T_\infty^{small} \stackrel{d}{=} \mathrm{PGW}(1)$ and $T_\infty^{big} \stackrel{d}{=} \mathrm{PGW}^\infty(1)$ are *independent*. This independence is often useful, and in the Section 3 its contribution is essential.

Beyond the Skeleton Trees

Many models for random trees have been introduced in combinatorics and the theory of algorithms ([47], [63]), and for the majority of these models one finds that there is a natural weak limit. Moreover, the analysis of Aldous [2] shows that for a large class of these models the limit object is a random rooted tree which shares many of the qualitative features of the Skeleton Tree given in Grimmett's Lemma.

Specifically, one finds that each of these limit objects has a unique infinite path from the root, and, as we saw in our discussion of Grimmett's Lemma, when one removes this skeleton one finds a sequence of finite trees that have a useful description. In general, these finite trees are no longer independent and identically distributed; instead, this sequence trees is more closely described as a hidden Markov process. In fortunate circumstances, the Markov structure that one finds can serve as a useful substitute for the independence that we use here.

3 Maximal Weight Partial Matching on Random Trees

Like most methods, the objective method is best understood through the examination of a concrete example, and, for all the reasons mentioned in the introduction, we believe that the maximal weight partial matching problem is the right place to begin. It is tractable enough to be developed completely, yet rich enough to demonstrate the structural features of the objective method that are important in more complex problems such as the random assignment problem.

3.1 Weighted Matchings of Graphs in General

For any graph $G = (V, E)$ a *partial matching* S is simply a subset of the set of edges of G such that no pair of edges of S share a common vertex. Any function $w : E(G) \to \mathbb{R}$ may be viewed as a weight function on the edge set of G, and the *weight of the partial matching* S is defined simply as

$$w(S) \stackrel{\text{def}}{=} \sum_{e \in S} w(e).$$

If S^* is a partial matching of G such that $w(S^*)$ is equal to the supremum of $w(S)$ over all partial matchings of G, then S^* is called a *maximal weight partial matching* of G. Such matchings are important in many problems of combinatorial optimization, and methods for computing maximum weight partial matchings have an important role in the theory of algorithms (cf. Lovász and Plummer [46]).

3.2 Our Case: Random Trees with Random Edge Weights

To each fixed finite rooted tree T and to each edge e of T, we now associate a random variable ξ_e. Moreover, we assume that the ensemble $\{\xi_e : e \in T\}$ is independent and that ξ_e has distribution F for each $e \in T$. We view $e \mapsto \xi_e$ as a weight function on T, and a T with this weight function will be called an *F-weighted rooted tree*. Finally, for each F-weighted rooted tree T we let $M(T)$ denote the maximum weight over all partial matchings of T.

In particular, if T_n denotes an F-weighted rooted tree that is chosen at random according to the uniform distribution on the set of n^{n-1} rooted trees with n-vertices, then we write M_n as shorthand for $M(T_n)$. The random variable M_n is therefore the maximum weight of a partial matching of a random n-vertex F-weighted rooted tree, and it will be at the center of our attention for the rest of this section.

3.3 Two Obvious Guesses: One Right, One Wrong

When the edge weight distribution F has a finite mean, one immediately suspects that there is an asymptotic relationship of the form

$$E[M_n] \sim \gamma n \quad \text{as } n \to \infty$$

where $\gamma = \gamma(F)$ is a constant that depends only on F. What one further suspects is that even for the nicest choices of F the calculation of γ might be an almost impossible task. By comparison, one can list perhaps a dozen problems where subadditive methods yield a similar asymptotic relationship, yet the limit constants have gone unknown for decades.

In the partial matching problem we face a happier circumstance. For continuous F the objective method not only yields a proof of the asymptotic relation, it also provides a concrete characterization of the limit constant. This characterization is not necessarily simple, but at least in the leading case of exponentially

distributed edge weights it does lead to an explicit integral representation for γ that can be calculated numerically.

Our Assumptions on the Distribution F

The intuition that leads one to guess that $E(M_n)/n$ converges does not impose any constraint on the edge-weight distribution F, except the trivial constraint that the expectation of $\xi_e \sim F$ should be well defined. Nevertheless, here we will always assume that F is continuous and that $P(\xi_e \geq 0) = 1$; these assumption guarantee that the maximal weight partial matching exists and is unique with probability one.

We do not believe that the continuity of F is needed for the convergence of $E(M_n)/n$, but, without this assumption, the possibility of multiple optima would forced us to face many irritating complications. Since our main intention here is to demonstrate the fundamental features of the objective method in the simplest realistic light, the issue of discontinuous F is best left for another time.

3.4 Not Your Grandfather's Recursion

The first thought of anyone interested in the asymptotics of $E[M_n]$ is to look for a relation between $E[M_n]$ and the earlier expectations $E[M_i]$, $1 \leq i < n$. Here we will also hunt for a recurrence relation, but what we find differs radically from the recursions one commonly meets in discrete mathematics.

If we remove an edge e from the edge set of T_n, then T_n is broken into two connected components. We can then view these components as rooted trees where we take the old vertices of e to be our new roots. Next, we label these trees $T_n^{small}(e)$ and $T_n^{big}(e)$ according to their cardinality (with any tie being broken by taking the labels at random). This process is similar to the discussion of the refinement of Grimmett's Lemma, except that here the cut edge e can be any edge of T_n.

Now, we consider whether or not the edge e is in the maximal partial matching of T_n. If one does not use the edge e, then the maximum weight of a partial matching of T_n is also equal to

$$M\big(T_n^{small}(e)\big) + M\big(T_n^{big}(e)\big). \tag{3.3}$$

To go deeper we need some additional notation. If T is any weighted rooted tree, we define $B(T)$ by letting $M(T) - B(T)$ denote the maximum weight of a partial matching of T where the root is *not* in an edge of the partial matching. By optimality of $M(T)$ we see that $B(T) \geq 0$, and we think of $B(T)$ as the *bonus* one gets from the *option* of being allowed to use edges that meet the root. With this notation, we see that the maximal weight of a partial matching that is *required* to use e is given by the sum

$$\xi_e + \big\{\, M(T_n^{small}(e)) - B(T_n^{small})(e)\big\} + \big\{M(T_n^{big}(e)) - B(T_n^{big})(e)\big\}. \tag{3.4}$$

When we compare the weights (3.3) and (3.4), we see that with probability one the edge e is in the maximum weight partial matching *if and only if*

$$\xi_e > B\big(T_n^{small}(e)\,\big) + B\big(T_n^{big}(e)\,\big). \tag{3.5}$$

This inclusion criterion naturally gives us a nice way to write M_n as a sum over the edges of T_n. If we use $\mathbf{1}(A)$ to denote the indicator function for an event A, then the inclusion criterion (3.5), tells us that

$$M_n = \sum_{e \in T_n} \xi_e \mathbf{1}\Big(\xi_e > B(T_n^{small}(e)\,) + B(T_n^{big}(e)\,) \Big). \tag{3.6}$$

Now, if $\mathbf{e}$ denotes an edge chosen uniformly from the edge set of T_n, we see from the sum (3.6) that the expectation $E[M_n]$ can be written as

$$E[M_n] = (n-1)E\big[\xi_{\mathbf{e}} \mathbf{1}\big(\, \xi_{\mathbf{e}} > B(T_n^{small}(\mathbf{e})\,) + B(T_n^{big}(\mathbf{e})\,)\,\big)\big].$$

Finally, since the distribution of $\xi_{\mathbf{e}}$ does not depend on $\mathbf{e}$ and since $\xi_{\mathbf{e}}$ is independent of $B(T_n^{small}(\mathbf{e}))$ and $B(T_n^{big}(\mathbf{e}))$, the last equation may be written a bit more crisply as

$$E[M_n] = (n-1)E\big[\, \xi \mathbf{1}\big(\, \xi > B(T_n^{small}(\mathbf{e})\,) + B(T_n^{big}(\mathbf{e})\,)\,\big)\,\big], \tag{3.7}$$

where we understand that $\xi \stackrel{d}{=} F$ and that ξ is independent of the pair $\big(\, B(T_n^{small}(\mathbf{e})\,),\, B(T_n^{big}(\mathbf{e})\,)\,\big)$.

3.5 A Direct and Intuitive Plan

The representation for $E[M_n]$ given by equation (3.7) is hardly a conventional recursion, yet it is still encouraging. The first factor $n-1$ hints loudly at the gross behavior one expects from $E[M_n]$. Moreover, from our discussion of the refinement of Grimmett's Lemma, one can guess that two tree processes $T_n^{small}(\mathbf{e})$ and $T_n^{big}(\mathbf{e})$ will converge in distribution, and this strongly suggests that the second factor of formula (3.7) will converge to a constant as $n \to \infty$.

Certainly, one should worry about the fact that $B(\cdot)$ is not a bounded continuous function on $\mathcal{G}_\star$; after all, $B(\cdot)$ is not even well defined on all of $\mathcal{G}_\star$. Nevertheless, $B(\cdot)$ is well defined on the finite graphs of $\mathcal{G}_\star$, and any optimist surely suspects it to nice enough for the random variables $B(T_n^{small}(\mathbf{e}))$ and $B(T_n^{big}(\mathbf{e}))$ to inherit convergence in distribution from the weak convergence of $T_n^{small}(\mathbf{e})$ and $T_n^{big}(\mathbf{e})$.

To refine these suspicions, we first note that $T_n^{small}(\mathbf{e})$ and $T_n^{big}(\mathbf{e})$ differ from the trees T_n^{small} and T_n^{big} in our discussion of Grimmett's Lemma only in that the roots of $T_n^{small}(\mathbf{e})$ and $T_n^{big}(\mathbf{e})$ are determined by the random edge $\mathbf{e}$ while the roots of T_n^{small} and T_n^{big} were determined by a random uniform choice from the respective vertex sets.

As Figure 2 suggests, the root r of $T_n^{small}(\mathbf{e})$ is not quite uniformly distributed on the vertex set of T_n when $\mathbf{e}$ is chosen at uniformly at random from the edge set of T_n. In fact, if n is odd, then there is always one vertex of T_n that has probability 0 of being the root of $T_n^{small}(\mathbf{e})$, and, when n is even, then there are always two vertices that have probability $1/(2(n-1))$ of being the root of $T_n^{small}(\mathbf{e})$. In both cases, all of the other vertices have probability $1/(n-1)$ of

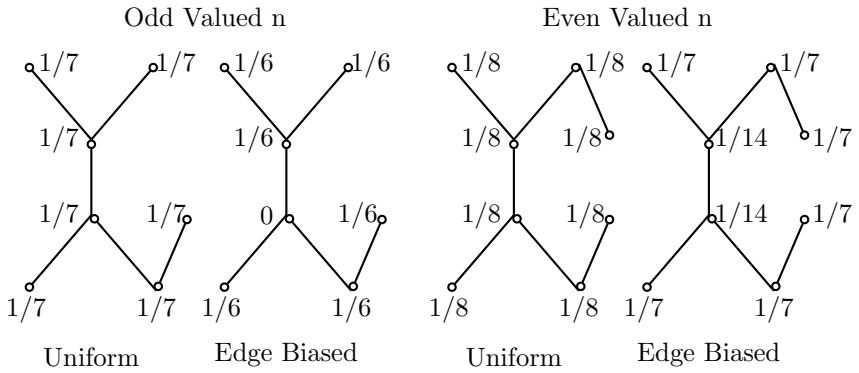

Figure 2: In this figure, the value next to a vertex is the probability of that vertex being the root of T_n^{small}. One should note that the uniform choice of an edge typically leads to a non-uniform distribution on the vertices.

being the root of $T_n^{small}(\mathbf{e})$, so, even though r is not uniformly distributed on the vertex set of T_n, it is *almost* uniformly distributed. In fact, the total variation distance between the distribution of r and the uniform distribution is always bounded by $1/2n$, and for odd values of n this bound is exact.

From this observation and a traditional coupling argument, one finds that there is only a small change in the distribution of the bonus when the omitted edge is used to determine the root; specifically, for all $n \geq 1$ and all $x \in \mathbb{R}$ we have the inequality

$$\left| P\big(B(T_n^{small}(\mathbf{e})) \leq x \big) - P\big(B(T_n^{small}) \leq x \big) \right| \leq \frac{1}{2n}. \tag{3.8}$$

A parallel analysis of the root of $T_n^{big}(\mathbf{e})$ yields an identical bound for the distributions of the bonuses of the big tree components, and together these bounds tell us that if the distribution F of ξ is not too long-tailed, then the asymptotic behavior of the right hand side of our exact formula (3.7) for $E[M_n]$ will not be changed if we replace the edge biased trees $T_n^{small}(\mathbf{e})$ and $T_n^{big}(\mathbf{e})$ by their randomly rooted cousins T_n^{big} and T_n^{small} whose asymptotic behavior is well understood.

We now see that along the path to proving that $E[M_n]/n$ converges, we have the natural intermediate task of showing that the bonuses $B(T_n^{small})$ and $B(T_n^{big})$ converge in distribution. Further, if we hope to have a serious chance of calculating the value of the limit of $E[M_n]/n$, we also need a concrete characterization of the limiting distributions of $B(T_n^{small})$ and $B(T_n^{big})$.

3.6 Characterization of the Limit of $B(T_n^{small})$

From our discussion of Grimmett's Lemma, we know that T_n^{small} converges in distribution to a tree T with the Poisson Galton-Watson distribution PGW(1).

When we view T as an F-weighted tree, then the bonus $B(T)$ is well defined since T is almost surely finite. Moreover, one can check directly from the definition of weak convergence in $\mathcal{G}_\star$ that we have

$$B(T_n^{small}) \xrightarrow{d} B(T), \tag{3.9}$$

so the real question is whether there is an effective way to characterize the distribution of $B(T)$. This is where the notion of a distributional identity enters the picture.

The basic idea is that the recursive definition of the tree T should translate into a useful self-referencing identity for the distribution of $B(T)$; we just need to put this idea to work. If C denotes the set of children of the root of T, then for each $i \in C$ there is a subtree T^i of T that is determined by the descendants of i, and we may view T^i as a rooted tree with root i. If ξ_i denotes the weight of the edge from the root of T to the child i, then the maximum weight of a matching on T that *does not* meet the root is given by the sum

$$\sum_{i \in C} M(T^i);$$

so, to get an identity for $B(T)$ we just need to calculate the maximum weight of a partial matching that *does* meet the root.

If $j \in C$, then the maximum weight for a partial matching that uses the edge from the root of T to j is equal to

$$\xi_j + \big\{M(T^j) - B(T^j)\big\} + \sum_{i \in C,\, i \neq j} M(T^i) = \xi_j - B(T^j) + \sum_{i \in C} M(T^i),$$

so the maximal weight for a partial matching that *does* include the root is equal to

$$\max_{j \in C} \big\{\xi_j - B(T^j)\big\} + \sum_{i \in C} M(T^i).$$

We now see that the difference $B(T)$ between the overall maximum and the constrained maximum $\sum_{i \in C} M(T^i)$ is given by

$$B(T) = \max\big\{\, 0,\, \xi_i - B(T^i) :\ i \in C\big\}. \tag{3.10}$$

Finally, the cardinality of C is Poisson with mean one, $T \stackrel{d}{=} T^i$ for all $i \in C$, and all of the random variables on the right hand side of the identity (3.10) are independent, so we have before us a very powerful constraint on the distribution of $B(T)$. In fact, we can check that any random variable that satisfies a distributional recursion like that defined by equation (3.10) must be equal in distribution to $B(T)$.

Formalization

To formalize this assertion, we let F and G denote any two distributions, and we define $D_F(G)$ to be the distribution of the random variable

$$\max\{0,\ \xi_i - Y_i : 1 \leq i \leq N\} \tag{3.11}$$

where random variables in the collection $\{N, Y_i, \xi_i : i = 1, 2, ...\}$ are independent, N has the Poisson(1) distribution, and we have $\xi_i \sim F$ and $Y_i \sim G$ for all $1 \leq i < \infty$. The next proposition tells us that one may characterize the distribution of the bonus $B(T)$ as a fixed-point of the mapping $D_F(\cdot)$.

Proposition 1 *If T is an F-weighted PGW(1) tree where F is continuous and $F([0, \infty)) = 1$, then the distribution*

$$G(x) = P\big(B(T) \leq x\big)$$

is the unique solution of the fixed-point equation

$$D_F(G) = G. \tag{3.12}$$

Since the discussion leading up to this proposition shows that the distribution of $B(T)$ is a solution of the fixed-point equation, we see that the characterization will be complete if we show that any solution of equation (3.12) must be the distribution of $B(T)$ for some F-weighted PGW(1) tree T. This will be done in a way that is particularly probabilistic.

A Uniqueness Proof by a Probabilistic Construction

Let G denote a solution of the fixed-point equation $D_F(G) = G$, so our task is to show that G is the distribution function of $B(T)$ where T is an F-weighted PGW(1) tree. Our plan is simply to use G to construct such a T, and Figure 3 suggests how we start.

In the first step, we just consider the tree consisting of a root and a set C of N children. We then attach a random variable Y_i to each child $i \in C$ where the Y_i are independent and $Y_i \sim G$. We then *define* Y by the identity

$$Y = \max\{0, \xi_i - Y_i : i \in C\},$$

where the ξ_i are independent and $\xi_i \sim F$. We describe this process by saying that the root has been *expanded*.

Next, take any unexpanded vertex v (at the second step this would be any of the elements of C) and expand v in the same way that we expanded the root. Specifically, we give v a set $C(v)$ of children with cardinality $N(v) = |C(v)|$ that is independent Poisson(1), attach a random variable $Y_{v,i}$ to each child $i \in C(v)$ where the $Y_{v,i}$ are independent and $Y_{v,i} \sim G$, and then define Y_v by the identity

$$Y_v = \max\{0, \xi_{v,i} - Y_{v,i} : i \in C\}, \tag{3.13}$$

where as before the $\xi_{v,i}$ are independent and $\xi_{v,i} \sim F$.

When we expand the root, the fixed-point property of G guarantees that Y has the distribution G. Later, when we expand v, the values of Y_v and Y will be changed, but their distribution will not be changed. Thus, to complete the construction, we simply repeat the process of expanding any unexpanded vertex until no such vertices remain. The tree we construct in this way is precisely a Galton-Watson tree with the Poisson(1) off-spring distribution, so,

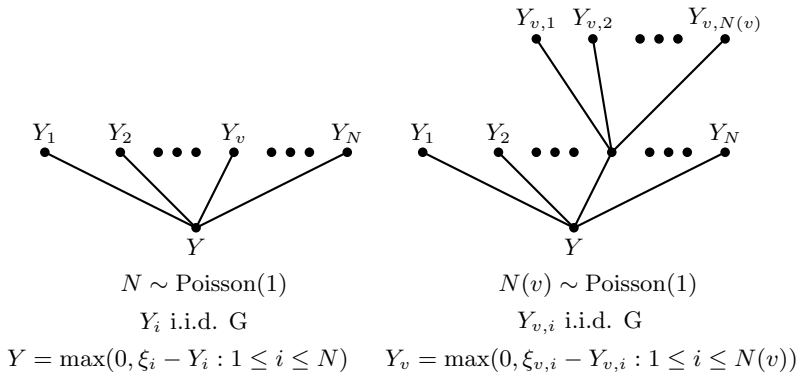

Figure 3: Expansion of the root and expansion of a descendant vertex.

with probability one, the expansion process stops after finitely many steps. Moreover, even though each vertex expansion changes the values of the random variables that are attached to the vertices along the path back to the root, all of the distributions will be unchanged.

By this construction, we see that for any leaf ℓ of the resulting tree T the value Y_ℓ associated to ℓ by the value rule (3.13) is zero. The bonus $B(T_\ell)$ is also zero for any leaf, and the recursion (3.10) for the bonus is the same as the recursion that defines the associated values Y_v, $v \in T$. We therefore find that $B(T) = Y$, and consequently we see that any solution G of the fixed-point equation $D_F(G) = G$ is equal to the distribution of $B(T)$ for an F-weighted PGW(1) tree T.

An Analytical Question

This proof of Proposition 1 is quite natural given the way that we came to the fixed-point equation $D_F(G) = G$, but the uniqueness question for this equation is in fact a purely analytical problem. It seems natural to ask if the uniqueness of the solution of $D_F(G) = G$ may be proved directly without recourse to the probabilistic interpretation of G, but we do not address that question here. For the moment the more pressing question is whether we might also be able to characterize $B(T_n^{big})$ by a distributional identity like the one we used to characterize $B(T_n^{small})$.

3.7 Characterization of the Limit of $B(T_n^{big})$

To be sure, one can begin just as before; specifically, we can note from our discussion of Grimmett's Lemma that T_n^{big} converges weakly in $\mathcal{G}_\star$ to a random tree, although this time the limit is T_∞, the skeletal tree with distribution

$\mathrm{PGW}^{\infty}(1)$. We would dearly love to conclude that

$$B(T_n^{big}) \xrightarrow{d} B(T_\infty), \tag{3.14}$$

but as soon as one writes this equation trouble appears.

Since T_∞ is almost surely infinite, one sees that the total weight of the maximal partial matching of T_∞ is infinite. For the same reason, the total weight of the maximal partial matching that does not meet the root of T_∞ is also infinite. The bottom line is that the bonus $B(T_\infty)$, which is nominally the difference of these two quantities, is not well defined.

Nevertheless, one should not lose heart. If indeed there is a finite random variable Z such that

$$B(T_n^{big}) \xrightarrow{d} Z, \tag{3.15}$$

then we still may be able to use our intuition about the undefined quantity $B(T_\infty)$ to help us to understand Z. Moreover, if we successfully guess a distributional identity that characterizes Z, we may even be able to use the stability of that characterization to build an honest proof of the conjectured convergence (3.15).

Guessing the Distributional Identity for Z

As physicists sometimes do, we cross our fingers and calculate with $B(T_\infty)$ even though it represents the difference of two infinite quantities. Since we only want to guess an identity for Z that will be subsequently justified by a rigorous argument, there is no cost to proceeding heuristically for the moment.

If we label T_∞ as we did in Figure 1 and if we remove the edge (v_0, v_1), then we obtain two subtrees with roots v_0 and v_1. If T denotes the subtree with root v_0, then T has the PGW(1) distribution, and, if T'_∞ denotes the subtree with root v_1, then T'_∞ has the $\mathrm{PGW}^{\infty}(1)$ distribution. By the "definition" of $B(T_\infty)$, the optimal partial matching of T_∞ that *does not match the root* v_0 has weight $M(T_\infty) - B(T_\infty)$, and our first task is to get an alternative representation for this quantity.

To begin, we let C denote the set of all of the children of v_0 *except* for the special child v_1 that is on the path from v_0 to infinity in T_∞. For each $i \in C$ we view the vertex i and its descendants as a subtree T^i of T, and we take i as the root of T^i. With this notation, we can write the weight of the optimal matching on T_∞ that does not match the root v_0 to a second vertex and obtain the bookkeeping identity

$$M(T_\infty) - B(T_\infty) = M(T'_\infty) + \sum_{i \in C} M(T^i). \tag{3.16}$$

The next step is to find an appropriate expression for the maximal weight of a partial matching of T_∞ that *does* meet the root. We first note the maximal weight of a partial matching of T_∞ that contains the edge (v_0, j) with $j \in C$ is given by

$$\xi_j + \left\{M(T^j) - B(T^j)\right\} + M(T'_\infty) + \sum_{i \in C, i \neq j} M(T^i)$$

which we may simplify to

$$\big\{\, \xi_j - B(T^j) \,\big\} + M(T'_\infty) + \sum_{i\in C} M(T^i).$$

Now, if ξ denotes the weight of the edge (v_0, v_1), then the maximal weight of a partial matching of T_∞ that contains the edge (v_0, v_1) can be written as

$$\sum_{i\in C} M(T^i) + \xi + \big\{M(T'_\infty) - B(T'_\infty)\big\},$$

so, choosing the best of all of the possibilities, we see that $M(T_\infty)$ equals

$$\sum_{i\in C} M(T^i) + M(T'_\infty) + \max\big\{\, 0,\, \max_{j\in C}\{\xi_j - B(T^j)\},\, \xi - B(T'_\infty) \,\big\}.$$

When we subtract the value of $M(T_\infty) - B(T_\infty)$ given by equation (3.16), we find

$$B(T_\infty) = \max\big\{0,\, \max_{j\in C}\{\xi_j - B(T^j)\},\, \xi - B(T'_\infty)\big\}.$$

We know $\max_{j\in C}\{0,\, \xi_j - B(T^j)\}$ has the same distribution as $B(T)$ by the basic distributional identity (3.10) for the bonus, and we also know that T'_∞ and T_∞ have the same distribution, so at last we have a nice, simple, and — lamentably heuristic — identity for $B(T_\infty)$:

$$B(T_\infty) \;\stackrel{d}{=}\; \max\big\{B(T),\, \xi - B(T_\infty)\big\}.$$

Conversion to an Honest Proposition

For any given continuous distribution F, we know that the distribution G of the bonus $B(T)$ for the F-weighted PGW(1) tree T can be characterized as the unique solution of the fixed-point equation $D_F(G) = G$, and our heuristic derivation suggests that we can characterize the limiting distribution of $B(T_n^{big})$ in a way that is perfectly analogous. Specifically, given any distribution H and independent random variables

$$Z \sim H, \quad Y \sim G, \quad \text{and} \quad \xi \sim F,$$

we let $\widetilde{D}_F(H)$ denote the distribution of $\max\{Y,\, \xi - Z\}$, and we consider the fixed-point equation

$$\widetilde{D}_F(H) = H.$$

This time we do not know *a priori* that there is a solution to this equation, but we do expect the limiting distribution of $B(T_n^{big})$ to exist, and for large n we also expect $B(T_n^{big})$ to be much like $B(T_\infty)$. Therefore, we suspect that the limiting distribution of $B(T_n^{big})$ will satisfy $\widetilde{D}_F(H) = H$. The next proposition confirms these suspicions while also confirming the crucial fact that the solution of the fixed-point equation $\widetilde{D}_F(H) = H$ is unique.

Proposition 2 *For any continuous distribution F with support on $[0,\infty)$, the fixed-point equation*

$$\widetilde{D}_F(H) = H \tag{3.17}$$

has a unique solution H; moreover, one has

$$\lim_{n\to\infty} P\big(\, B\big(T_n^{big}\big) \le x \,\big) = H(x) \quad \textit{for all } x \in \mathbb{R}.$$

Before proving this proposition, we will first show that it really does complete the program with which we began.

3.8 The Limit Theorem for Maximum Weight Partial Matchings

Our whole discussion of $E[M_n]$ has been motivated by the formula

$$E[M_n] = (n-1)E\big[\, \xi\mathbf{1}\big(\, \xi > B(T_n^{small}(\mathbf{e})\,) + B(T_n^{big}(\mathbf{e})\,)\,\big)\,\big], \tag{3.18}$$

and now with the help of Proposition 1 and Proposition 2 we are ready to take the limit. If we assume that ξ has a finite expectation, then the integrands in formula (3.18) are uniformly integrable and they converge in distribution to $\xi\mathbf{1}(\xi > Y + Z)$, so without further concern we can take the limit under the expectation to obtain the main theorem of this section.

Theorem 1 *If F is a continuous distribution with support on $[0,\infty)$ and a finite mean, then the maximum weight M_n of an F-weighted random tree on n vertices satisfies*

$$\lim_{n\to\infty} \frac{1}{n} E[M_n] = E[\xi\mathbf{1}(\xi > Y + Z))] \tag{3.19}$$

where the random variables ξ, Y, and Z are independent,

$$\xi \sim F, \quad Y \sim G, \quad \textit{and} \quad Z \sim H,$$

and the distributions G and H are the unique solutions of the fixed-point equations

$$D_F(G) = G \quad \textit{and} \quad \widetilde{D}_F(H) = H. \tag{3.20}$$

The expression $E[\xi\mathbf{1}(\xi > Y + Z))]$ for the limit (3.19) in Theorem 1 always provides some structural insight into the value of the limit, but, since the distributions of Y an Z are only defined through the solutions of two fixed-point equations, one might think that Theorem 1 is still rather abstract. The next result tells us that in one of the cases that matters most, the limit $E[\xi\mathbf{1}(\xi > Y + Z))]$ may be calculated explicitly.

Theorem 2 *If $F(x) = 1 - e^{-x}$ for $x \ge 0$, then the maximum weight M_n of a partial matching on an F-weighted random tree on n vertices satisfies*

$$\lim_{n\to\infty} \frac{1}{n} E[M_n] = \int_0^\infty \int_0^s \; c(e^{-y} - be^{-s}) \exp(-ce^{-y} - ce^{-(s-y)}) se^{-s} \; dy\, ds,$$

where the constants b and c are defined by equations (3.23) and (3.24) of Lemma 2. Moreover, by numerical integration one finds

$$\lim_{n\to\infty} \frac{1}{n} E[M_n] = 0.239583....$$

To prove Theorem 2 we just need to solve two fixed-point equations, and the lack-of-memory property of the exponential distribution makes this easier than anyone has any right to expect. Later we will find several other situations like this one where distributional identities lead us to calculations that are mysteriously pleasant.

Lemma 2 *If $F(x) = 1 - e^{-x}$ for $x \geq 0$, then the solutions G and H of the fixed-point equations (3.20) are*

$$G(y) = \exp(-ce^{-y}) \quad \textit{for } y \geq 0 \tag{3.21}$$

$$\textit{and} \quad H(z) = (1 - be^{-z})\ \exp(-ce^{-z}) \quad \textit{for } z \geq 0, \tag{3.22}$$

where b and c are related by

$$b = c^2/(c^2 + 2c - 1) \tag{3.23}$$

and c is the unique strictly positive solution of

$$c^2 + e^{-c} = 1. \tag{3.24}$$

Numerically, one finds $b = 0.543353...$ and $c = 0.714556....$

To prove Lemma 2, we first consider $Y \sim G$ along with an independent pair of independent sequences $\{Y_i : i = 1, 2, ...\}$ and $\{\xi_i : i = 1, 2, ...\}$ with $Y_i \sim G$ and $\xi_i \sim F$. In terms of these variables, the equation $D_F(G) = G$ simply means

$$P(Y \leq y) \quad = P\big(\xi_i - Y_i \leq y \text{ for all } 1 \leq i \leq N\,\big) \quad = \exp\big(-P(\xi - Y > y)\big),$$

where the second identity follows by conditioning on the independent Poisson variable N. Now, if we take logarithms and apply the lack-of-memory property of the exponential, then we find

$$\log P(Y \leq y) = -P(\xi > y + Y) = -e^{-y} P(\xi > Y), \tag{3.25}$$

and this identity gives us the explicit formula (3.21) when we set c equal to $P(\xi > Y)$.

Next, to get the determining equation (3.24) we only need to calculate $c = P(\xi > Y)$ with help from the distribution (3.21) of Y; specifically, we find

$$c = P(\xi > Y) = \int_0^\infty e^{-y} P(Y \leq y) dy = \int_0^\infty \exp(-ce^{-y})\ e^{-y} dy$$
$$= \int_0^1 \exp(-cz)\ dz = c^{-1}(1 - e^{-c}).$$

Finally, to get the distribution of Z, we use the defining relation (3.17) and the lack-of-memory property of the exponential to see for $z > 0$ that

$$\begin{aligned} P(Z \le z) =& P(Y \le z)P(\xi - Z \le z) \\ =& \exp(-ce^{-z})\ P(\xi \le z + Z) = \exp(-ce^{-z})\ \big\{1 - e^{-z}P(\xi > Z)\big\}, \end{aligned}$$

so if we set $b = P(\xi > Z)$, then $P(Z \le z) = (1 - be^{-z})\ \exp(-ce^{-z})$, just as we hoped. Now, all that remains is to express $b = P(\xi > Z)$ as a function of $c = P(\xi > Y)$, and we begin by noting

$$\begin{aligned} b = P(\xi > Z) &= \int_0^\infty P(Z < z)e^{-z}\, dz \\ &= \int_0^\infty (1 - be^{-z})\ \exp(-ce^{-z})\ e^{-z}dz = \frac{c-b}{c^2} + \frac{b - c + bc}{c^2}e^{-c}. \end{aligned}$$

Now, if we use the identity (3.24) to replace e^{-c} by $1 - c^2$, we see that the last equation may be solved for b to obtain the promised relation (3.23). Finally, we note that the uniqueness of c follows from the convexity of the map $x \mapsto x^2 + e^{-x}$.

3.9 Closing the Loop: Another Probabilistic Solution of a Fixed-Point Equation

To complete our investigation of the maximal partial matching problem, we only need to prove Proposition 2. Specifically, we need to show that the fixed-point equation

$$\widetilde{D}_F(H) = H$$

has a unique solution, and we need to show that this solution is the limiting distribution for the bonuses $B\big(T_n^{big}\big)$.

Existence Follows from General Theory

The existence of a solution to the equation $\widetilde{D}_F(H) = H$ is almost obvious once one draws a connection between the fixed-point equation and the theory of Markov chains. Specifically, if we set

$$K(z, A) = P\big(\max\{\xi - z, Y\} \in A\big)$$

where $\xi \sim F$ and $Y \sim G$ are independent, then $K(x, A)$ defines a Markov transition kernel on $\mathbb{R}^+$. One then sees at that a distribution is a fixed point of $\widetilde{D}_F$ if and only if it is a stationary distribution for the Markov chain with kernel K.

Finally, since the kernel K has the Feller property and since it is dominated by the distribution of $\max\{\xi, Y\}$, the general theory of Markov chains on $\mathbb{R}^+$ tells us that there is a stationary distribution for K (cf. Meyn and Tweedie [49], Theorem 2.1.2).

Uniqueness — Observed and Confirmed by Coupling

While we might also be able to show the uniqueness of the solution of the fixed-point equation with help from Markov process theory, there is another approach via coupling that is particularly informative. Moreover, the argument anticipates a useful stability result that gives us just the tool we need to prove that $B(T_n^{big})$ converges in distribution.

To begin, we assume that the variables $\{\, \xi_m,\, Y_m : 1 \le m \le \infty \,\}$ are independent with $\xi_m \sim F$ and $Y_m \sim G$ for all m. Next, simultaneously for each $z \ge 0$, we construct a Markov chain $Z_m(z)$, $m = 1, 2, ...$ by taking $Z_0(z) = z$ and by setting

$$Z_m(z) = \max\{\, \xi_m - Z_{m-1}(z),\, Y_m \,\} \quad \text{for } m \ge 1.$$

Now, for all z and z' in $\mathbb{R}^+$, we see that

$$Z_m(z) \ge Z_m(z') \quad \text{implies} \quad Z_{m+1}(z) \le Z_{m+1}(z'),$$

and from this anti-monotone property we see that $Z_m(z) \ge Z_m(0)$ for all even values of m. Therefore, if we let

$$T = \min\{\, m \text{ odd } : Z_m(0) = Y_m \,\},$$

then we see that $Z_T(z) = Y_T$ for all z. As a consequence, we see that T is a *coupling time* in the exceptionally strong sense that

$$Z_m(z) = Z_m(0) \quad \text{for all} \quad z \ge 0 \quad \text{and all} \quad m \ge T; \tag{3.26}$$

moreover, the definition of $Z_m(0)$ gives us the simple bound

$$P(Z_m(0) = Y_m \,|\, Z_{m-1}(0) = z) \ge P(\xi_m < Y_m) \quad \text{for all } z \ge 0.$$

Now, if we set $\varrho = 1 - P(\xi_m < Y_m)$ then we can easily check that $\varrho < 1$. To see why this is so, we first note that in a tree $\mathcal{T}$ that has exactly one edge with weight ξ' the bonus $B(\mathcal{T})$ is equal to ξ'. Now, if $\mathcal{T}$ is a PGW(1) tree, then the probability that $\mathcal{T}$ has exactly one edge is equal to e^{-1}, so for $Y \stackrel{d}{=} B(\mathcal{T})$ we have

$$P(Y \in \cdot) \ge e^{-1} P(\xi' \in \cdot).$$

Now, ξ is independent of Y and has a continuous distribution, so we have the bound

$$P(Y \le \xi) \ge e^{-2} P(\xi' \le \xi) = \frac{1}{2e}.$$

Thus, we find $\varrho \le 1 - 1/2e < 1$, so the bound

$$P(T \ge 2k+1) \le \varrho^k \quad \text{for all } k \ge 0$$

gives us more than we need to show that T is finite with probability one.

Now, if H is the distribution function for a stationary measure for the kernel K and if $Z \sim H$, then by stationarity we have $Z_n(Z) \sim H$ for all $n \ge 1$. Therefore, by the bound on the coupling time, we find for all $x \in \mathbb{R}$ that

$$|H(x) - P(Z_m(0) \le x)| \le \varrho^k \quad \text{for all } m \ge 2k+1,$$

so we see that there is at most one stationary distribution.

A Coupling Connection

Here we should note that this simultaneous coupling argument shares several elements in common with the "coupling from the past" technique that has become a popular tool in the theory of Markov chain Monte Carlo (cf. Propp and Wilson [54]). In particular, the argument given above shows that for *any* random initial value X we have

$$|P(Z_m(X) \in A) - P(Z \in A)| \leq \varrho^k \quad \text{for all } n \geq 2k+1;$$

and, remarkably enough, one can even choose X so that it depends on the elements of $\{\xi_m, Y_m : m \geq 1\}$.

3.10 From Coupling to Stability — Thence to Convergence

If one repeats this coupling argument while relaxing the condition on the distribution of the variables Y_m so that they are only required to have a distribution that is close to G, one is led to bounds on distributions that can be viewed the solutions of an "approximate" fixed-point equation. Any such stability result must face some notational clutter because many quantities are constantly changing, but the next lemma should be easy to parse if one keeps in mind that n is an index that increases as the distribution of Y_m^n comes closer to G. Also, in the lemma we choose the time index m so that we can focus our attention on X_0^n, a quantity that one may view as the dependent variable at the "time" 0 when the "approximation quality" equals n.

Lemma 3 *Fix the integer $k \geq 1$ and suppose for each n we have non-negative random variables $\{X_m^n : -2k+1 \leq m \leq 0\}$ that are linked to the variables $\{Y_m^n : -2k+2 \leq m \leq 0\}$ by the relation*

$$X_m^n = \max\{\, \xi_m^n - X_{m-1}^n,\, Y_m^n \,\} \quad \textit{for all } m = -2k+2, -2k+3, \ldots, 0,$$

where the random variables $\{\xi_m^n : -2k+2 \leq m \leq 0\}$ are independent, have distribution F, and are independent from

$$X_{-2k+1}^n \quad \textit{and} \quad \{Y_m^n : -2k+2 \leq m \leq 0\}.$$

Further, suppose that one has

$$(Y_m^n : -2k+2 \leq m \leq 0) \;\xrightarrow{d}\; (Y_m : -2k+2 \leq m \leq 0) \;\; \textit{as } n \to \infty,$$

where the random variables $\{Y_m : m = 0, -1, -2, ...\}$ are independent and have distribution G. If H is the unique solution of

$$\widetilde{D}_F(H) = H,$$

then we have the bound

$$\limsup_{n\to\infty} |\, P(X_0^n \le x) - H(x)\,| \le \varrho^k \quad \textit{for all } x \in \mathbb{R} \textit{ and } k = 1, 2, ...$$

where ϱ is given by

$$\varrho = \iint_{y<x} dF(x)\, dG(y) < 1.$$

From Stability to Convergence

First, we let T_∞ denote an F-weighted PGW$^\infty(1)$ tree, and let $v_0, v_1, v_2, \ldots$ denote the sequence of vertices along the unique path to infinity from the root v_0 according to the definition of PGW$^\infty(1)$ and the familiar Figure 1. Next, we let $k \ge 1$ be an arbitrary integer, and we note from the local weak convergence of T_n to T_∞ that there exists an integer $N(k, \omega)$ so that for all $n \ge N(k, \omega)$ there is a path π with $2k$ vertices and the following properties:

(1) the first vertex of π is the root v_0^n of T_n

(2) if π is written as $v_0^n \to v_1^n \to v_2^n \to \cdots \to v_{2k-1}^n$ then the deletion of the edges of π creates a forest of subtrees T_i^n, $i = 1, 2, ..., 2k-1$, of T_n where T_i^n is rooted at v_i^n, and

(3) we have the weak convergence relation

$$(\, T_i^n \,:\, 0 \le i \le 2k-2\,) \;\xrightarrow{d}\; (\, T_i^\infty \,:\, 0 \le i \le 2k-2\,)$$

where all of the righthand components T_i^∞, $1 \le i \le 2k-2$, are independent PGW(1) trees.

Next, for each $n \ge 1$ and each $1 \le i < 2k$, we let R_i^n denote the subtree of T_n that is rooted at v_i^n that one obtains by removing the single edge (v_{i-1}^n, v_i^n) from the path π. We think of R_i^n as a "remainder subtree" that contains everything "past" v_{i-1}^n and to complete the picture we write $R_0^n = T_n$.

Finally, if we let ξ_i^n denote the weight on edge (v_i^n, v_{i+1}^n), then we may use *exactly* the same argument that gave us the heuristic derivation of the fixed-point equation $\widetilde{D}_G(H) = H$ to show that

$$B(R_i^n) = \max\big\{\, \xi_i^n - B(R_{i+1}^n),\, B(T_i^n)\,\big\} \quad \text{for all } \; 0 \le i \le 2k-2. \tag{3.27}$$

In this case the argument is also rigorous since all of the trees involved are finite, and $B(\cdot)$ is well defined at each stage.

Finally, we note that the recursion (3.27) sets up the direct application of Lemma 3. We only need to take $m = -i$, to find that the lemma gives us

$$\limsup_{n\to\infty} |P(B(T_n) \le x) - H(x)| \le \varrho^k.$$

Since $\varrho < 1$ and since k is arbitrary, the last inequality is more than we need to complete the proof of Proposition 2.

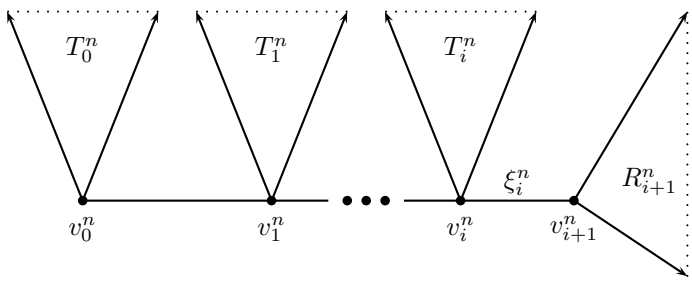

Figure 4: The Residual R_{i+1}^n.

3.11 Looking Back: Perspective on a Case Study

When one looks back on the maximal partial matching problem, there are several themes that seem to be worth recording for future use. First of all, the inclusion criterion (3.5),

$$\xi_e > B\big(T_n^{small}(e)\,\big) + B\big(T_n^{big}(e)\,\big),$$

helped drive the analysis almost from the beginning. In particular, this necessary and sufficient condition for the edge e to be in the maximal partial matching gave us the succinct representation (3.7) for $E[M_n]$, which in turn encouraged us to study the limit theory of $B(T_n^{small}(e)\,)$ and $B(T_n^{big}(e)\,)$.

The asymptotic analysis of $B\big(T_n^{small}(e)\,\big)$ was relatively straightforward, but the analysis of $B(T_n^{big}(e)\,)$ forced us to work harder. The key to its behavior turned out to be the random variable recursion (3.27),

$$B(R_i^n) = \max\{\,\xi_i^n - B(R_{i+1}^n),\, B(T_i^n)\,\} \quad \text{for all } \; 0 \le i \le 2k-2,$$

but this equation might have left us empty handed if we had not first studied its infinite analog. For the analog, we found there was a unique solution, and this was encouraging. Such equations often have "stabilized" versions, and this principle guided us to the proof of the weak convergence of $B(T_n^{big}(e)\,)$.

One is unlikely to find too many problems that may be solved with a step-by-step application of the method used here, but we will soon see several other examples where most of the same elements are present. Certainly, the heuristic derivation and rigorous solution of a distributional identity emerges as a basic theme. Sometimes this identity is followed by stabilization, and at other times it is exploited to provide an analog of an inclusion rule similar to that provided by equation (3.5). Typically, the new inclusion rule would apply only in a more elaborate model that one constructs with help from the distributional identity.

We will see from the analysis of more complicated problems that the solution of the maximal partial matching problem is a hybrid; it is initiated and organized by the objective method, but it is completed with help from a traditional stochastic recursion. Nevertheless, the maximal partial matching problem offers persuasive evidence of the effectiveness of the objective method, and it provides a quick introduction to several of the method's main themes. In the next three sections, we hope to make this case more complete and more inclusive.

4 The Mean-Field Model of Distance

In physics, and increasingly often in probability theory, one meets geometric problems where one tries to gain insight by building an analytic model that only captures *part* of the structure of the original set of inter-point distances. For example, instead of studying a random variable that depends on all of the inter-point distances of a point process in $\mathbb{R}^d$, one may directly model just the distances from a given point to its successive neighbors and then study an appropriate analog of the original variable.

By preserving the distributions of the most essential inter-point distances while distorting the distribution of less important distances, one hopes to achieve a significant simplification while still gaining some insight into the phenomena originally of interest. Models that aim for this kind of simplification are commonly called *mean field* models following an earlier tradition in physics where complicated fields were replaced by simpler ones that were calibrated to match the average behavior of the more complicated fields.

In this section, we will describe several problems of probabilistic combinatorial optimization where mean-field models have been developed. These models have a natural kinship with the objective method, and local weak convergence turns out to be a most natural tool.

4.1 From Poisson Points in $\mathbb{R}^d$ to a Simple Distance Model

Consider for a moment a Poisson process $\mathcal{P}$ on $\mathbb{R}^d$ that has uniform intensity $1/v_d$ where v_d is the volume of the unit ball in $\mathbb{R}^d$. For such a process the expected number of points in the unit ball is equal to one, and, if we look at the successive distances from the origin to the points of $\mathcal{P}$, we have a new point process

$$0 < \Delta_1 < \Delta_2 < \Delta_3 < \dots . \tag{4.28}$$

This process is again Poisson, although now we have nonconstant intensity, and the expected number of points in $[0, x]$ equal to x^d.

In a world where one is more concerned about the distances to one's nearest neighbors than about other distances, this simple reduction suggests a promising line of inquiry. Can one assign random distances to the edges of a complete graph K_n with n vertices so that for large n the successive distances from a point r in K_n will mimic the distances that one finds for the sequence of neighbors of 0 in the d-dimensional Poisson process $\mathcal{P}$?

A Finite Distance Model: The d-Exponential K_n

The answer to this question is affirmative, and it is easy to describe. For each edge e of K_n, we simply introduce an independent random variable ξ_e such that

$$\xi_e \stackrel{d}{=} n^{1/d}\Delta_1, \text{ or equivalently, } P(\xi_e \geq n^{-1/d}x) = e^{-x^d} \text{ for } x \geq 0. \tag{4.29}$$

Now, if we fix a vertex r of K_n and consider the distance from this root r to its nearest neighbor, next nearest neighbor, and so on, then one finds a set of $n-1$ increasing distances

$$0 < D_1^n < D_2^n < \cdots < D_{n-1}^n.$$

The key feature of the distribution (4.29) is that one may now easily prove that

$$(D_i^n : \ i \geq 1) \ \stackrel{d}{\longrightarrow} \ (\Delta_i : \ i \geq 1) \quad \text{as } n \to \infty. \tag{4.30}$$

This useful result yields some geometric insight, but it is possible to go much further. One does not need long to guess that the distributional limit (4.30) may be strengthened to prove the *local weak convergence* of the finite distance models K_n. That is, one hopes to find that K_n satisfies the kind of limit theorem that we found for T_n in Section 2.

Where to Start

Anytime one hopes to prove a weak limit theorem in $\mathcal{G}_\star$, the battle begins by searching out a reasonably explicit candidate for the *limit object*. Here for example, one should minimally note that any reasonable candidate for the local weak limit of K_n must have infinite degree at every vertex. Moreover, from the motivating limit (4.30) we also know that in the candidate object the distances from a vertex to its neighbors must behave like the successive values of the Poisson process (4.28).

After these easy observations, some deeper reflection is needed. In due course, one finds that the description of the candidate is almost complete; only one further observation is needed. Any candidate for the limit of graphs K_n must be a *tree* — despite the fact that K_n is more full of cycles than any other graph!

4.2 The Poisson Weighted Infinite Tree — or, the PWIT

Even with a good limit object in mind, one needs to confirm that the candidate exists as a clearly defined process. Here this is easily done by an inductive procedure. One begins with a single root vertex that we denote by r. This root vertex is then given an *infinite* number of children, and the edges from the root to the children are assigned lengths according to a realization of a Poisson process $(\xi_i^r : 1 \leq i < \infty)$ with mean function x^d. The children of the root are said to form *generation one*.

Now, recursively , each vertex v of generation k is given an infinite number of children, and the edges to these children of v are again assigned lengths

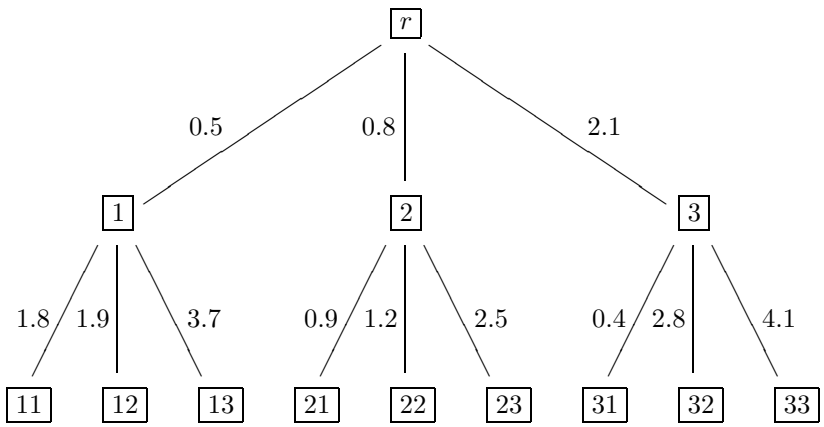

Figure 5: Part of a realization of a PWIT that shows just the first three children of each vertex; one needs to keep in mind that in the full PWIT each vertex has an infinite number of children. Here the values written next to the edges that descend from a fixed vertex are intended to reflect an independent realization of a Poisson process with mean function x^d. Finally, the labels of the vertices in this figure are only for convenience; in a PWIT the vertices are unlabelled except for the root.

according to an independent realization of a Poisson process $(\xi_i^v : 1 \leq i < \infty)$ with mean function x^d. This procedure is then continued *ad infinitum*.

The resulting rooted infinite tree T is then a well defined random variable with values in $\mathcal{G}_\star$. The tree T is said to be a *Poisson weighted infinite tree*, or, a PWIT. Still, the nickname intends no disrespect. PWITs have an inevitable role in the limit theory of trees, and the next theorem tells us that they are precisely the limits of appropriately weighted complete graphs on n vertices.

Theorem 3 (Convergence Theorem for K_n) *If K_n is a randomly rooted complete graph on n vertices with independent edge lengths ξ_e that satisfy*

$$P(\xi_e \geq n^{-1/d}x) = e^{-x^d} \quad \textit{for all } x \geq 0,$$

then we have the local weak convergence

$$K_n \xrightarrow{d} T \quad \textit{as } n \to \infty.$$

where T is a PWIT with edge weight mean function x^d.

This theorem from Aldous [3] permits one to study optimization problems for K_n in the same way that Grimmett's Lemma permits one to study optimization

problems for the uniform random tree T_n. Nevertheless, there is one noteworthy distinction; the exponential weights on K_n are tied to the Poisson processes that generate the PWIT, but the F-weights for the edges of T_n may be chosen with great freedom.

We will not repeat the proof of Theorem 3 here, but we should indicate why the cycles of K_n disappear in the limit. If we introduce a random variable that measures the length of the shortest cycle containing the root,

$$C_n = \min_C \Big\{ \sum_{e \in C} \xi_e : C \text{ is a cycle of } K_n \text{ that contains the root } r \Big\},$$

then the key issue is that one can prove that

$$\lim_{n \to \infty} P(C_n \leq \ell) = 0 \quad \text{for all } \ell > 0. \tag{4.31}$$

Intuitively this result says that in the limit there is no cycle of finite length that contains the root, and, with some attention to detail, this leads to a proof that the weak limit of K_n in $\mathcal{G}_\star$ must be a random tree.

The Dimension Parameter and a Voluntary Restriction

For the rest of this section, we will restrict our attention to the case when $d = 1$, and the Poisson process (4.28) of successive distances $(\Delta_i : 1 \leq i < \infty)$ reduces to a plain vanilla Poisson process with constant rate 1. Nevertheless, we should emphasize that this restriction is voluntary; it is only taken to shorten and simplify our exposition.

For all of the problems studied in this section and the next, there are parallel results even for noninteger $0 < d < \infty$, and, in most cases, one will find complete proofs of these results in the references. Moreover, one should note that for the problems studied here, the logic of the analysis is identical for all values of $0 < d < \infty$, even though the explicit calculations often depend on the specific value. These calculations are almost always simplest for $d = 1$ where one has access to the lack-of-memory property of the exponential.

4.3 The Cut-off Components of a Weighted Graph and a PWIT

For a moment, we consider a perfectly general weighted graph G, and we let G_s denote the subgraph G_s that one obtains from G by deleting all of the edges of length s or greater. Also, if v is a vertex of G we let $c(G, v; s)$ denote the connected component of G_s that contains v.

Now, if T is a dimension-one PWIT with root r we may view the component $c(T, v; s)$ as a rooted tree with root r, and, rather pleasantly, this tree can identified with a Galton-Watson branching process with a single progenitor and an offspring distribution that is Poisson with mean s. In other words, for all $s > 0$ we have

$$c(T, r; s) \stackrel{d}{=} \mathrm{PGW}(s),$$

and this simple observation has several useful consequences.

In particular, if $q(s)$ denotes the probability that the component $c(T, r; s)$ is *infinite*, then classical formulas for the extinction probability of a branching process (such as those of Harris [30], p.7) tell us that we have $q(s) = 0$ when $0 \leq s \leq 1$, and otherwise we find that $q(s)$ is the unique strictly positive solution of

$$1 - q(s) = \exp(-sq(s)) \quad \text{when } s > 1. \tag{4.32}$$

By inverting this relationship, one also sees that the value of s for which we have extinction probability $0 < q < 1$ is given by

$$s(q) = -q^{-1} \log(1 - q), \tag{4.33}$$

a formula that will soon come in handy.

4.4 The Minimum Spanning Forests of an Infinite Graph

We want to define an analog of the minimal spanning tree for the PWIT and other infinite graphs, but some care must be taken. Naive attempts to mimic the usual greedy algorithms certainly will not work; in most cases, they can never get started.

Instead, we are guided by a well-known criterion for an edge $e = (v_1, v_2)$ with length s to be in the MST of a finite graph G with distinct edge weights. For such a graph, e is in the MST of G if and only if there does not exist a path in G from v_1 to v_2 such that each edge of the path has weight less than s. This criterion is easy to prove, and it is occasionally useful in correctness proofs for MST algorithms, such as the five discussed in Section 4.4 of Matoušek and Neštřil [48].

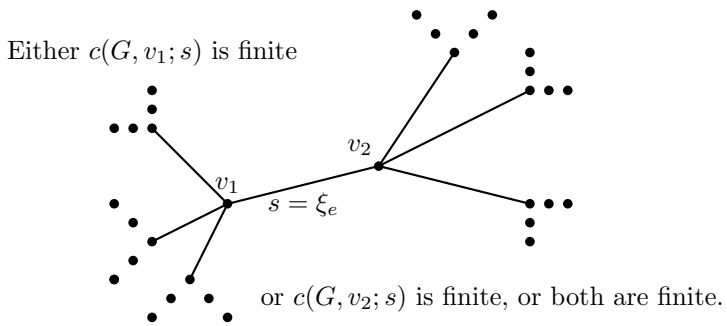

Figure 6: The edge $e = (v_1, v_2)$ of G is in the minimal spanning forest if and only if at least one of the trees $c(G, v_1; s)$ and $c(G, v_2; s)$ is finite when $s = \xi_e$.

Definition 3 (Minimal Spanning Forest) *The* minimal spanning forest *of an infinite graph G that has all distinct edge lengths is the subgraph $MSF(G)$ of G with the same vertex set as G and with an edge set that contains each edge $e = (v_1, v_2)$ of G for which*

(1) $c(G, v_1; s)$ and $c(G, v_2; s)$ are disjoint, and

(2) $c(G, v_1; s)$ and $c(G, v_2; s)$ are not both infinite

when s is taken to be the length of the edge $e = (v_1, v_2) \in G$.

One can easily check that the graph $MSF(G)$ defined this way will have no cycles, and, since it trivially spans the vertex set of G, we see that $MSF(G)$ is certainly a *spanning forest*. Nevertheless, some work is needed to see why $MSF(G)$ is the right analog to the MST of a finite graph and to justify the name *minimal* spanning forest.

A first step, one should note that all of the connected components of $MSF(G)$ are infinite. To see this, suppose to the contrary that G_0 is a finite component of $MSF(G)$ of minimal cardinality. Next, consider the shortest edge $e = (v_1, v_2)$ between G_0 and its complement. Since one cannot get out of G_0 without using an edge length greater than s, we see that $c(G, v_1; s)$ is contained in G_0 and hence that $c(G, v_1; s)$ is finite. From the definition of the $MSF(G)$, we then have $e \in MSF(G)$, and this contradicts the minimality of the component G_0.

4.5 The Average Length Per Vertex of the MSF of a PWIT

If T is a PWIT with root r then the sum of the edges of $MSF(T)$ that are incident to r can be written as

$$S_{\mathrm{MSF}}(T) = \sum_{e:r\in e} \xi_e \mathbf{1}(e \in MSF(T)),$$

and, if we think of each vertex as owning half of the length of each edge that meets the vertex, then it is reasonable to think of

$$A = \frac{1}{2} E[S_{\mathrm{MSF}}(T)] \tag{4.34}$$

as an *average length per vertex* of the minimal spanning forest of T.

One might guess that A could be almost impossible to compute, but such a view underestimates the natural grace of the PWIT. The next lemma and its proof show that the PWIT supports some almost magical possibilities.

Lemma 4 *The average cost A per vertex for the MSF of a PWIT satisfies*

$$A = \zeta(3) = \sum_{k=1}^{\infty} k^{-3}.$$

To prove the lemma, we first recall that if one conditions a Poisson process to have a point at s, then the set of complementary points is again a Poisson process. Thus, if we condition on the existence of an edge e of length s from the root to some child, then the subtrees obtained by cutting that edge are again independent PGW(s) trees.

The probability that at least one of these is finite is equal to $1 - q^2(s)$, so this is also the probability that e is an edge in the minimal spanning forest and we find

$$A = \frac{1}{2} \int_0^{\infty} s(1 - q^2(s))\, ds. \tag{4.35}$$

Calculation of this integral is quite a pleasing experience. We first integrate by parts and then we use the defining equation (4.32) for $q(s)$ to find

$$A = \frac{1}{2} \int_0^{\infty} s^2 q(s) q'(s)\, ds = \frac{1}{2} \int_1^{\infty} s^2 q(s) q'(s)\, ds = \frac{1}{2} \int_0^1 \frac{\log^2(1-q)}{q}\, dq,$$

so the substitution $u = -\log(1-q)$ lets us finish with

$$A = \frac{1}{2} \int_0^{\infty} u^2 \frac{e^{-u}}{1-e^{-u}}\, du = \frac{1}{2} \int_0^{\infty} u^2 \sum_{k=1}^{\infty} e^{-ku}\, du = \sum_{k=1}^{\infty} \frac{1}{k^3} = \zeta(3).$$

4.6 The Connection to Frieze's $\zeta(3)$ Theorem

The emergence of $\zeta(3)$ as the average cost per vertex of the minimal spanning forest of a PWIT is at least a little mysterious. Still, there is a related precedent, and when one sees $\zeta(3)$ in the context of trees, the famous $\zeta(3)$ theorem of Frieze [21] immediately comes to mind.

Theorem 4 *Let each edge e of the complete graph on n vertices be assigned an independent cost ξ_e with distribution F where F has a finite variance and where $F'(0)$ exists and is nonzero. If $C_n^{MST}(F)$ denotes the minimal cost of a spanning tree of this graph, then as $n \to \infty$, we have*

$$C_n^{MST}(F) \xrightarrow{p} \zeta(3)/F'(0) \quad \textit{and} \quad E[C_n^{MST}(F)] \to \zeta(3)/F'(0). \tag{4.36}$$

By a simple coupling argument, Steele [57] showed that the convergence in probability still holds without any moment assumptions, and later by a martingale argument Frieze and McDiarmid [22] showed that convergence in probability can be strengthened to almost sure convergence.

Nevertheless, what matters here are simply the exponential costs, and to make the connection with our earlier calculations explicit, we note that Frieze's theorem implies that if the edge weight distribution F_n is exponential with mean n, then we also have

$$\lim_{n\to\infty} \frac{1}{n} E[C_n^{MST}(F_n)] = \zeta(3). \tag{4.37}$$

This limit says the average cost per vertex of the minimal spanning tree of the appropriately weighted complete graph tends to the same value as the expected average cost per vertex of the PWIT with $d = 1$. Conceivably, the reappearance of $\zeta(3)$ is only a coincidence, but it is not. Quite to the contrary, the limit (4.37) is a consequence of the fact that $A = \zeta(3)$.

The Local Weak Convergence Link

As a special case of Theorem 3, we know that for the complete graph K_n on n vertices with edge lengths that are independent and exponentially distributed with mean n, one has the local weak convergence

$$K_n \xrightarrow{d} T \quad \text{as } n \to \infty \tag{4.38}$$

where T denotes a PWIT with $d = 1$. We will also see later in Theorem 6 of Section 5 that just local weak convergence (4.38) and the fact that the PWIT is an infinite graph with all distinct edge weights is enough to give us

$$\big(K_n,\, MST(K_n)\big) \xrightarrow{d} \big(T,\, MSF(T)\big). \tag{4.39}$$

This limit is typical of the way the theory of local weak convergence informs one about a problem in combinatorial optimization. At a conceptual level, the limit (4.39) answers almost any question one may have about the qualitative behavior of $MST(K_n)$ for large n, provided that the question meets one restriction; it be expressed in terms of a function of some neighborhood of the root of $MST(K_n)$.

A Randomized Root and a Total Sum

The expectation $E[C_n^{MST}(F_n)]$ can be written as a function of the neighborhood of the root, yet this representation seems to be a lucky break. Eventually one calls on this break routinely, but at least once it seems useful to examine the full logical chain; we begin by noting

$$C_n^{MST}(F_n) \stackrel{d}{=} \sum_e \xi_e \mathbf{1}(e \in MST(K_n))$$

so now if we let R denote a root chosen at random from the vertex set of K_n we also have

$$\frac{1}{n} E\big[\sum_e \xi_e \mathbf{1}(e \in MST(K_n))\big] = \frac{1}{2} E\big[\sum_{e:R\in e} \xi_e \mathbf{1}(e \in MST(K_n))\big]$$

and from the local weak convergence $MSF(K_n) \to MSF(T)$ we have

$$\sum_{e:R\in e} \xi_e \mathbf{1}(e \in MST(K_n)) \stackrel{d}{\longrightarrow} \sum_{e:R\in e} \xi_e \mathbf{1}(e \in MST(T)). \tag{4.40}$$

Now, once one checks uniform integrability (which is a not particularly easy, but which is a doable technical exercise), one finds from the basic distributional limit (4.40) that

$$E\big[\sum_{e:R\in e} \xi_e \mathbf{1}(e \in MST(K_n))\big] \to E\big[\sum_{e:R\in e} \xi_e \mathbf{1}(e \in MST(T))\big] = 2\zeta(3),$$

where the last equality comes from Lemma 4. From the links of the chain we find

$$\frac{1}{n} E\big[C_n^{MST}(F_n)\big] \to \zeta(3),$$

just as we hoped to show.

5 Minimal Cost Perfect Matchings

A perfect matching of a finite graph is a collection of disjoint edges with the property that at most one vertex of the graph fails to be contained in an edge of the collection. Perfect matchings are among the most intensively studied structures of graph theory, and they suggest many natural random variables. Here we focus on the prince among these, the minimal cost C_n of a perfect matching of the complete graph K_n with independent edge weights ξ_e with distribution F.

The assignment problem discussed in the introduction is precisely the perfect matching problem for the bi-partite graph $K_{n,n}$, and the analysis of C_n is intimately connected to that of A_n. Thus, the following theorem is a direct analog of the $\zeta(2)$ limit theorem for the expected assignment cost $E[A_n]$, but here we rescale the F-weights to be consistent with our discussion of the local weak limit theory of K_n.

Theorem 5 *If C_n denotes the minimal cost of a perfect matching of the complete graph K_n on n vertices with independent edge costs ξ_e having the exponential distribution with mean n, then*

$$\lim_{n\to\infty} \frac{1}{n} E[C_n] = \frac{1}{2}\sum_{k=1}^{\infty} k^{-2} = \frac{1}{2}\zeta(2). \tag{5.41}$$

Given the earlier experience with the maximum partial matching on the weighted uniform tree T_n, one has every reason to expect that the local weak convergence of K_n should provide the key to the proof of this theorem. This view is well founded, but the earlier analyses only provide one with a proper orientation. There are many devils in the details, and a complete proof of Theorem 5 would take more space than we have here.

Nevertheless, we can provide the better part of one-half of the proof; that is, we can provide most of the argument that proves the limit infimum of $E[C_n]/n$ is not less than $\zeta(2)/2$. While doing so, we can also underscore the main structural features of the full proof. In particular, we develop a basic distributional identity and solve the identity to find an unexpected connection to the logistic distribution. We can then use the distributional identity to construct a special stochastic process which can be used in turn to define an inclusion rule for the edges of a PWIT to be in a minimal cost matching. Finally, given just a few properties of the freshly constructed process, we can complete the analysis of the limit infimum.

In the course of this development, we can also complete a promise made in Section 2; specifically, we can formalize the *involution invariance* property that one finds for all limit objects that are given by the standard construction discussed in Section 2. Even though involution invariance seems perfectly innocuous, it has an unavoidable part in the proof of Theorem 5. In the earliest stage, it helps weed out defective candidates for the limit objects, and at a later stage it guides us to a proper understanding of the matchings on the PWIT that *can* be attained as limits of matching on K_n.

5.1 A Natural Heuristic — Which Fails for a Good Reason

The local weak convergence of K_n almost compels one to think of using good matchings on the PWIT to help in the hunt for good matchings on K_n. The natural hope is that a perfect matching in the PWIT with minimal average cost per vertex should be close to a perfect matching of K_n for large n; thus, weak convergence theory might set us on the trail of a proof of Theorem 5. This hope can be justified, but only after one restricts attention to an important subset of the matchings on the PWIT that satisfy a natural symmetry property. If one ignores this restriction, there are matchings on the PWIT that are simply too good to be tied to matchings on K_n.

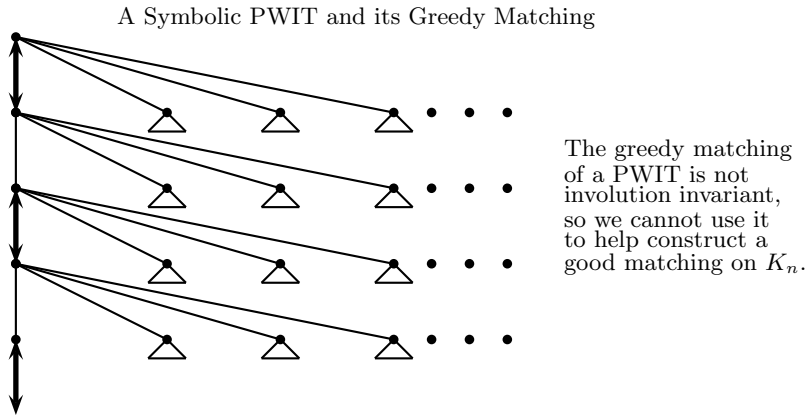

Figure 7: To construct the greedy matching of a PWIT one takes the shortest edge e out of the root, removes all of the vertices incident to e, and proceeds recursively. The expected cost per vertex of the greedy matching is equal to one-half, which beats the unbeatable limiting average cost of $\zeta(2)/2$ per vertex for K_n.

An Example: Greedy Matching on the PWIT

In a finite weighted graph one reasonable way to hunt for a good matching is the *global greedy algorithm* where one picks the edge e with minimal weight, removes all of the edges incident to either endpoint of e, and proceeds recursively. Unfortunately, in a PWIT this method does not have a chance to start — a PWIT has no edge of minimal cost.

In the PWIT, a reasonable alternative to the global greedy algorithm is the *local greedy algorithm* where one starts by choosing the edge of minimal cost that is incident to the root. In a PWIT such an edge e exists, and we even know that it has expected value equal to 1. One can then delete e, remove the edges incident to e, and proceed recursively along the infinite forest produced by this first step.

The local greedy algorithm produces a matching in the PWIT with an expected cost per vertex equal to one-half, and this is intriguing since it is even smaller than the value $\zeta(2)/2$ that we expected to show to be asymptotically optimal for the K_n.

Unfortunately, the greedy matching for the PWIT simply cannot be the limit of a sequence of matchings for K_n. Moreover, one can even see this fact from first principles. The problem is that the greedy matching lacks a symmetry property that one must find in every limit object like the limit of the optimal matchings of K_n.

5.2 Involution Invariance and the Standard Construction

As the discussion of the greedy matching suggests, the measures on $\mathcal{G}_\star$ that one obtains by the standard construction of Section 2 are not completely general. The random rooting process gives these measures a basic symmetry that we call *involution invariance*.

This property turns out to be fundamental in the deeper applications of the objective method. With some experience, one finds that involution invariance serves as a graph theoretical surrogate for the translation invariance that is central to the theory of stationary point processes. The main goal of this section is to provide a patient development of the simplest properties of involution invariance.

The Defining Elements

If $\widetilde{\mathcal{G}}_\star$ denotes the set of pairs (G, v), where $G \in \mathcal{G}_\star$ and where $v \in V$ is a neighbor of the root of G, then one can view an element $\widetilde{G} = (G, v)$ of $\widetilde{\mathcal{G}}_\star = \mathcal{G}_\star \times V$ as a graph with a *distinguished directed edge*; specifically, the distinguished edge $(\mathrm{root}(G), v)$ may be viewed as a directed edge from the root of G to the neighbor vertex v. One can therefore define an involution

$$\iota : \widetilde{\mathcal{G}}_\star \to \widetilde{\mathcal{G}}_\star$$

by taking $\iota(\widetilde{G})$ to be the element of $\widetilde{\mathcal{G}}_\star$ that one obtains simply by *reversing* the direction of the distinguished edge of $\widetilde{G}$; that is, one makes v the new root and takes the old root to be the new distinguished neighbor.

Now, given any probability measure μ on $\mathcal{G}_\star$, one can define a new positive measure $\widetilde{\mu}$ on $\widetilde{\mathcal{G}}_\star$ by (1) taking the marginal measure of $\widetilde{\mu}$ on $\mathcal{G}$ to be μ and by (2) taking the conditional measure $\widetilde{\mu}_G$ on the neighbors of the root of G to be the counting measure. The measure $\widetilde{\mu}$ one obtains in this way is said to be *induced* on $\widetilde{\mathcal{G}}_\star$ by μ.

One should note that $\widetilde{\mu}$ is *not* a probability measure, except in some special cases. Moreover, since the definition of $\mathcal{G}_\star$ permits infinite vertex sets and permits vertices with infinite degree, $\widetilde{\mu}$ is often an infinite measure. Nevertheless, $\widetilde{\mu}$ is always a nonnegative σ-finite measure on the complete metric space $\widetilde{\mathcal{G}}_\star$. Thus, with proper care, all of the usual results of weak convergence theory may be applied to sequences of induced measures.

Definition 4 (Involution Invariance) *A probability measure μ on $\mathcal{G}_\star$ is said to be* **involution invariant** *provided that the σ-finite measure $\widetilde{\mu}$ induced on $\widetilde{\mathcal{G}}_\star$ by μ is invariant under the involution map ι; that is to say,*

$$\widetilde{\mu}(A) = \widetilde{\mu}(\iota^{-1}(A)) \quad \textit{for all Borel } A \subset \widetilde{\mathcal{G}}_\star.$$

One of the consequences of involution invariance is that it essentially characterizes the measures on $\mathcal{G}_\star$ that one can obtain via the standard construction. The next lemma formalizes this assertion for finite graphs.

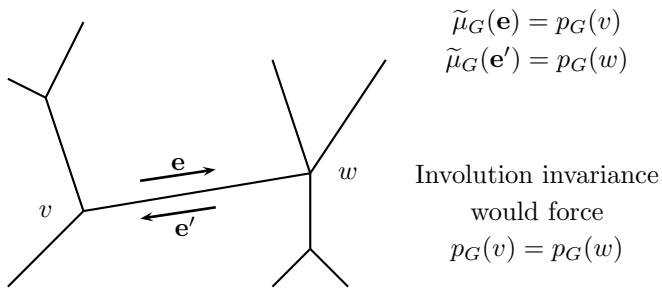

Figure 8: Given $G = (V, E)$ we let $p_G(v)$ denote the probability that v is the root of G, and if $\widetilde{\mu}$ is the induced measure, we write $\widetilde{\mu}_G$ for the marginal measure given G. For any directed edge (v, w) of G, the measure $\widetilde{\mu}_G$ puts mass $p_G(v)$ on the edge (v, w) and puts mass $p_G(w)$ on the edge (w, v). If one has involution invariance, then $\widetilde{\mu}_G(v, w) = \widetilde{\mu}_G(w, v)$, and we see that $p_G(v) = p_G(w)$. Thus, $p_G(\cdot)$ is constant on each connected component of G whenever one has involution invariance.

Lemma 5 *If μ is the probability distribution of $\mathbf{G}[X]$ where $\mathbf{G}[X]$ is given by the standard construction and $\mathbf{G}[X]$ is finite with probability one, then μ is involution invariant. Conversely, if the probability measure μ is involution invariant and concentrated on the finite connected graphs in $\mathcal{G}_\star$, then μ is the distribution of a $\mathcal{G}_\star$-valued random variable $\mathbf{G}[X]$ that is given by the standard construction.*

To prove the direct implication, we first condition on the event that the underlying graph of $\mathbf{G}[X]$ is the fixed finite graph G. If G has n vertices and m undirected edges, then the marginal measure $\widetilde{\mu}_G$ puts mass $1/n$ on each of the $2m$ directed edges of G, so we trivially get involution invariance. To prove the converse, one notes by the captioned discussion of Figure 8 that $p_G(\cdot)$ is constant on each connected component of G. Finally, if we assume that G is connected, we see $p_G(\cdot)$ is constant on G. This is precisely what it means to say that the root X of G is chosen uniformly from its vertex set, so the proof of the lemma is complete.

Connection to the Objects of the Objective Method

One of the basic features of involution invariance is that it is preserved under local weak convergence in the metric space $\mathcal{G}_\star$. To prove this fact, one first shows that the topology on $\mathcal{G}_\star$ extends in a natural way to a topology on $\widetilde{\mathcal{G}}_\star$ that makes $\widetilde{\mathcal{G}}_\star$ a complete separable metric space. Next one shows that the transformation $\mu \to \widetilde{\mu}$ is a continuous mapping with respect to weak convergence; specifically

$$\mu_n \xrightarrow{d} \mu_\infty \quad \Rightarrow \quad \widetilde{\mu}_n \xrightarrow{d} \widetilde{\mu}_\infty.$$

Finally, one shows that the involution map $\iota : \widetilde{\mathcal{G}}_\star \to \widetilde{\mathcal{G}}_\star$ is continuous.

5.3 Involution Invariance and the Convergence of MSTs

The formalities of involution invariance give one no hint that the notion has much force, but the next theorem shows that involution invariance can be genuinely powerful. In fact, there is something fundamental about involution invariance, since, in a way, it replaces the spatial invariance that one uses constantly in the theory of homogeneous point processes. The pleasantly general Theorem 6 contains as special cases Frieze's $\zeta(3)$ theorem which was discussed earlier and the convergence theorem for MSTs of random samples in the cube which we will discuss in Section 6.

Theorem 6 (MST Convergence Theorem) *Let G_∞ denote a $\mathcal{G}_\star$-valued random variable such that with probability one G_∞ has infinitely many vertices and no two of the edges of G have the same length. Further, let $\{G_n : n = 1, 2, ...\}$ denote a sequence of $\mathcal{G}_\star$-valued random variables such that for each n the distribution of G_n is given by the standard construction and such that for each n the vertex set of G_n has cardinality n with probability one. If*

$$G_n \xrightarrow{d} G_\infty \quad \text{as } n \to \infty, \tag{5.42}$$

then one has the joint weak convergence in $\mathcal{G}_\star \times \mathcal{G}_\star$,

$$\big(G_n,\, MST(G_n)\big) \xrightarrow{d} \big(G_\infty,\, MSF(G_\infty)\big). \tag{5.43}$$

Further, if N_n denotes the degree of the root of $MST(G_n)$ and N denotes the degree of the root of $MSF(G_\infty)$

$$N_n \xrightarrow{d} N \quad \text{and} \quad E[N_n] \to E[N] = 2, \tag{5.44}$$

and, if L_n denotes the sum of lengths of the edges incident to the root of $MST(G_n)$ and L denotes the corresponding quantities for $MSF(G_\infty)$, then

$$L_n \xrightarrow{d} L. \tag{5.45}$$

The applications of the limit (5.43) are easy to understand without working through the convergence proof. Nevertheless, a survey on local weak convergence should give at least one such proof in reasonable detail, and the limit (5.43) is a attractive candidate. The proof demonstrates several basic facets of involution invariance, and the net result is genuinely useful.

To begin the proof, we first note that the hypothesis (5.42) tells us the sequence $\{G_n\}$ is tight, and if we consider the sequence of pairs $(G_n,\, MST(G_n))$ one can check that they determine a sequence $\{\mu_n\}$ of measures on $\mathcal{G}_\star \times \mathcal{G}_\star$ that is also tight. We then let μ denote a measure on $\mathcal{G}_\star \times \mathcal{G}_\star$ that one obtains as the weak limit of a subsequence $\{\mu_{n_j}\}$.

By Skorohod's embedding theorem, we may suppose the $\mathcal{G}_\star \times \mathcal{G}_\star$-valued random variables $(G_{n_j},\, MST(G_{n_j}))$ are all defined on a common probability space and that there is a $\mathcal{G}_\star$-valued random variable S such that

$$\big(G_{n_j},\, MST(G_{n_j})\big) \to (G_\infty, S) \quad \text{with probability one.} \tag{5.46}$$

From the definition of μ as the weak limit of the sequence $\{\mu_{n_j}\}$, we see that the proof of the limit (4.39) boils down to showing

$$S(\omega) = MSF(G_\infty)(\omega) \quad \text{with probability one.} \tag{5.47}$$

This identity will be proved in two steps — one for each direction of the set inclusions.

The First Inclusion is Easy

To show that $MSF(G_\infty)(\omega) \subset S(\omega)$ for almost all ω, we begin by considering an edge $e = (u, v) \in MSF(G_\infty)$ that has length s. By the definition of the minimal spanning forest, we may assume without loss of generality that the component $c(G_\infty, u; s)$ is finite. For convenience of notation, we will also assume that the subsequence n_j is in fact the full sequence; when the argument is done, one will see that no generality was lost.

Now take a neighborhood H of u that is large enough to contain $c(G_\infty, u; s)$ and note by the almost sure convergence of G_n to T that there is a finite $N(\omega)$ such that for all $n \geq N(\omega)$ there is an isomorphism between H and a neighborhood of the root of G_n. Also, since $c(G_\infty, u; s)$ is finite, there is an $\varepsilon > 0$ such that each edge of $c(G_\infty, u; s)$ has length less than $s - \varepsilon$.

If $e^n = (u^n, v^n)$ denotes the edge associated with e by the neighborhood isomorphism, then for large n the length of e^n is at least $s - \varepsilon/2$ and all edges of $c(G_n, u^n; s)$ have length strictly less than $s - \varepsilon/2$. Therefore, there can be no path in G_n from u^n to v^n that has all of its edges shorter than the length of e^n. Thus, we find that there a finite $N'(\omega)$ such that e^n is in $MST(G_n)$ for all $n \geq N'(\omega)$. Since $MST(G_n)$ converges to S in $\mathcal{G}_\star$, we get our first inclusion

$$MSF(G_\infty)(\omega) \subset S(\omega) \text{ with probability one.} \tag{5.48}$$

A More Subtle Inclusion

To complete the proof we just need to show that the set difference

$$D(\omega) = S(\omega) \setminus MSF(G_\infty)(\omega) \tag{5.49}$$

is empty with probability one. By the general lemma below, we also see that it suffices to show that the number of edges in D that meet the root of G has expectation zero.

Lemma 6 *Let G be an involution invariant element of $\mathcal{G}_\star$ with root r and let W be an involution invariant subset of the edges of G. If*

$$E\big[\sum_{e:r\in e} \mathbf{1}(e \in W)\big] = 0,$$

then W is empty with probability one.

This lemma is proved in Aldous and Steele [7], and, since the way that it uses involution invariance is more simply illustrated by the proof of Lemma 7, we will not repeat the proof Lemma 6 here. Instead, we will go directly to its application.

Naturally, we want to apply Lemma 6 where W is taken to be the difference set D defined by equation (5.49), so our task is to estimate the corresponding expectation. If we let N_n denote the degree of root of G_n in $MST(G_n)$ and let N denote the degree of the root of G_∞ in $MSF(G_\infty)$, then we have the usual Skorohod embedding and the inclusion relation (5.48) give us the bound

$$N \leq \liminf_{n\to\infty} N_n, \tag{5.50}$$

and the definitions of D, N_n and N also give us

$$\sum_{e:r\in e} \mathbf{1}(e \in D) \leq (\liminf_{n\to\infty} N_n) - N. \tag{5.51}$$

Since an n-vertex tree has exactly $n-1$ edges, we always have the trivial limit

$$E[N_n] = 2(n-1)/n \nearrow 2,$$

so, by the bound (5.50) and Fatou's Lemma, we have

$$E[N] \leq E[\liminf_{n\to\infty} N_n] \leq \liminf_{n\to\infty} E[N_n] = 2.$$

This observation almost completes the proof of Theorem 6, and the next lemma does the rest. It is all we need to confirm that the expectation of the sum (5.51) is zero.

Lemma 7

$$E[N] = 2. \tag{5.52}$$

In fact, only half of the lemma remains to be proved; we just need to show $E[N] \geq 2$. To begin, we consider an edge $e = (r, v)$ that is incident to the root r of $MSF(G)$. If we write f for "finite" and ∞ for "infinite," then each such edge must be one of four possible types

$$(f, f), \quad (f, \infty), \quad (\infty, f), \quad \text{and } (\infty, \infty)$$

according to the cardinality of the respective components of $MSF(G_\infty) \setminus \{e\}$ that contain r and v. We will write N_{ff}, $N_{f\infty}$, $N_{\infty f}$, and $N_{\infty\infty}$ for the numbers of edges of the corresponding types.

From the definition of the minimal spanning forest, we know that every vertex is in an infinite component of the MSF, so N_{ff} is always zero. A more important observation is that unless r has infinite degree, the argument given in the caption of Figure 9 gives us

$$2N_{f\infty} + N_{\infty\infty} \geq 2. \tag{5.53}$$

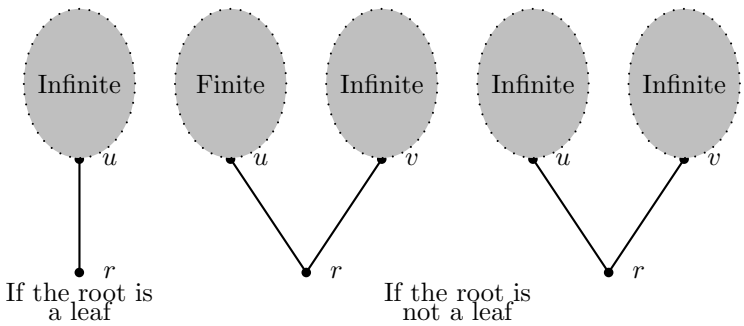

Figure 9: If the root of G_∞ has finite degree, we are sure to have $2N_{f\infty}+N_{\infty\infty} \geq 2$ because no edge of the MSF of G_∞ can connect two finite components of the MSF and because in the three remaining cases one always has either $N_{f\infty} \geq 1$ or $N_{\infty\infty} \geq 2$.

The key step turns out to be a very simple involution invariance argument that shows that $N_{f\infty}$ and $N_{\infty f}$ have the same expectation. Once this is confirmed, one can complete the proof of the lemma by noting that the bound (5.53) gives us

$$\begin{aligned} E[N] &= E[N_{f\infty}] + E[N_{\infty f}] + E[N_{\infty\infty}] \\ &= 2E[N_{f\infty} + E[N_{\infty\infty}] \geq 2. \end{aligned}$$

Finally, to establish the equality of $E[N_{f\infty}]$ and $E[N_{\infty f}]$, we first recall the measure $\tilde{\mu}$ that one uses to define involution invariance. For any directed edge $e^{\rightarrow}$ we have

$$E[N_{f\infty}] = \tilde{\mu}\{e^{\rightarrow} \text{ is of type } (f, \infty)\},$$

and we also have

$$\begin{aligned} E[N_{\infty f}] &= \tilde{\mu}\{e^{\rightarrow} \text{ is of type } (\infty, f)\} \\ &= \tilde{\mu}\{e^{\leftarrow} \text{ is of type } (f, \infty)\}, \end{aligned}$$

where $e^{\leftarrow}$ is the reversal of $e^{\rightarrow}$. Now, since G_∞ is involution invariant, we have equality of reversals, so

$$\tilde{\mu}\{e^{\leftarrow} \text{ is of type } (f, \infty)\} = \tilde{\mu}\{e^{\rightarrow} \text{ is if type } (f, \infty)\},$$

and we find at last that $E[N_{\infty f}] = E[N_{f\infty}]$. This completes the proof of Lemma 7, and hence completes the proof of Theorem 6 as well.

A Thought on Involution Invariance

The way that involution invariance does its work is by telling us that "whatever is true from the perspective of the root r" is also true from the perspective of the root's neighbor v. In the case of a connected graph, one can chain this argument from vertex to vertex and conclude that "whatever is true from the perspective of one vertex is true from the perspective of any other vertex." This statement is far from precise, but the derivation of the identity $E[N_{\infty f}] = E[N_{f\infty}]$ suggests that it provides honest guidance.

Results Related to the MST Convergence Theorem

The MST Convergence Theorem given here is new, and it generality serves to underscores the benefit of working with the local weak convergence of sequences of $\mathcal{G}_\star$-valued random variables. The simplicity of Theorem 6 should not conceal the fact that it contains as special cases both the basic convergence theory for MSTs for weighted complete graphs developed by Aldous [1] and the corresponding theory for random Euclidean points developed by Aldous and Steele [6]. Avram and Bertsimas [12] independently studied both of these cases, and one further finds that their calculations are also much more transparent when viewed in the light of local weak convergence.

5.4 A Heuristic That Works by Focusing on the Unknown

The greedy algorithm gave us an excellent matching on the PWIT, but it lacked involution invariance and did not help a bit in our hunt for a good matching on K_n. Moreover, the symmetry of the PWIT makes it hard to see how one can do better; almost any matching one might find by good guessing is likely to face the same failings that were met by the greedy matching. Nevertheless, a good involution invariant matching can be found, provided that one takes a hint from the analysis of the maximal partial matching problem.

The overarching plan is to consider the minimal cost per vertex of an involution invariant perfect matching on the PWIT and to treat this cost as an "unknown." We then begin the hunt for an equation that contains this unknown, and, in due course, heuristic considerations will lead us to a distributional identity. This identity can then be solved rigorously, and its solution can be used in turn to build a joint process with one marginal being a PWIT and another marginal being random variables that satisfy the distributional identity. These auxiliary variables can then used to construct an "inclusion rule" that provides a test for an edge's membership in the matching on the PWIT.

One small wrinkle this time is that we will focus on a simple difference, rather than the more subtle bonus that we used before. Specifically, we focus on the difference

$$D(T) = C(T) - C\big(T \setminus \{r\}\big),$$

where $C(T)$ is the minimal cost of a perfect matching of T and $C(T \setminus \{r\})$ is the minimal cost of a perfect matching of the graph $T \setminus \{r\}$. Naturally, both $C(T)$ and $C(T \setminus \{r\})$ are infinite, so $D(T)$ is not really well defined. Nevertheless, in our analysis of the maximal partial matching problem we met a similar difficulty,

and we still managed to prevail. We have every reason to hope for a similar outcome this time.

If j is a child of the root, then we let T^j denote the descendants of j and we view T^j as a rooted graph with root j. We then introduce the analogous heuristic difference $D(T^j)$, and note that even though $D(T)$ and $D(T^j)$ are not honestly defined, one can continue in the brave tradition of physical scientists and argue that nevertheless they should satisfy the distributional identity

$$D(T) = \min_{1 \le j < \infty} \big\{ \xi_j - D(T^j) \big\}. \tag{5.54}$$

To see this, we first note that the minimal cost of a matching that contains the edge (r, j) from the root r to the vertex j has cost

$$\xi_j + C\big(T^j \setminus \{j\}\big) + \sum_{i: i \ne j} C(T^i) \tag{5.55}$$

and that $C(T)$ is just the minimum of all of these values. Also, we have the direct identity

$$C\big(T \setminus \{r\}\big) = \sum_{1 \le i < \infty} C(T^i). \tag{5.56}$$

The difference of equations (5.55) and (5.56) is just $\xi_j - C(T^j) + C(T^j \setminus \{j\})$, so in the end we see that

$$C(T) - C\big(T \setminus \{r\}\big) = \min_{1 \le j < \infty} \big\{ \xi_j - C(T^j) + C\big(T^j \setminus \{j\}\big) \big\}. \tag{5.57}$$

5.5 A Distributional Identity with a Logistic Solution

Equation (5.54), or its expanded equivalent (5.57), motivates us to begin our rigorous analysis by first considering the distributional identity

$$X \stackrel{d}{=} \min_{1 \le i < \infty} \big\{ \xi_i - X_i \big\} \quad -\infty < X < \infty, \tag{5.58}$$

which we interpret to be an identity for a distribution F that is fully specified by supposing (1) $X \sim F$, (2) the X_i are independent, (3) $X_i \sim F$ for all $i = 1, 2, ...$, and (4) the sequence $\{X_i\}$ is independent of the process $0 < \xi_1 < \xi_2 < \cdots$ which is a Poisson process with constant intensity equal to one.

Naturally, one wants to know about the existence and uniqueness of a distribution F that satisfies this distributional identity. The next lemma provides an answer of striking simplicity.

Lemma 8 (Logistic Solution of a Distributional Identity) *The unique solution of the identity (5.58) is the logistic distribution*

$$F(y) = 1/(1 - e^{-y}), \quad -\infty < y < \infty \tag{5.59}$$

corresponding to the density function

$$f(y) = (e^{y/2} + e^{-y/2})^{-2}, \quad -\infty < y < \infty.$$

Better Than a Proof — A Derivation

Rather than just give a direct verification of this lemma, one does better to see how the probability theory of the distributional identity leads one to discover the connection to the logistic distribution. Thus, for the moment, we assume that the distribution F satisfies the identity (5.58), and we will try to calculate F.

We first note that the distributional identity (5.58) asserts that

$$1 - F(y) = P\left(\min_{1 \le i < \infty} \{\xi_i - X_i\} \ge y\right)$$

where the $\{X_i : i = 1, 2, ...\}$ on the righthand side are i.i.d. with distribution F and $\{\xi_i : i = 1, 2, ...\}$ is an independent Poisson process. Therefore, if we consider the collection of points $\mathcal{P}$ in $\mathbb{R}^2$ given by $\{\,(\xi_i, X_i) \,:\, i = 1, 2, ...\}$ we see that $\mathcal{P}$ is a Poisson process on $(0, \infty) \times (-\infty, \infty)$ with intensity $dzdF$. As a consequence, we have

$$\begin{aligned} 1 - F(y) =& P\big(\text{no point of } \mathcal{P} \text{ is in } \{(z, x) : \; z - x \le y\}\,\big) \\ &= \exp\left(-\iint_{z-x<y} dz\, dF(x)\right) \\ &= \exp\left(-\int_0^\infty \{1 - F(z - y)\}\; dz\right) \\ &= \exp\left(-\int_{-y}^\infty \{1 - F(u)\} du\right). \end{aligned}$$

When we differentiate this identity we find

$$F'(y) = \big(1 - F(-y)\big)\big(1 - F(y)\big), \tag{5.60}$$

and the symmetry of this equation implies that the density $F'(\cdot)$ is symmetric. As a consequence, we see that equation (5.60) is equivalent to

$$F'(y) = F(y)\big(1 - F(y)\big),$$

which is easily seen to be separable. We can rewrite it as

$$dy = \frac{dF}{F(1 - F)} = \frac{dF}{F} + \frac{dF}{1 - F}$$

and solve by inspection to find

$$y = \log \frac{F}{1 - F} \quad \text{or} \quad F(y) = 1/(1 - e^{-y}).$$

Thus, F is simply is the logistic distribution. It is famous for popping up in odd places, but its appearance here still a bit surprising. The remainder of our analysis does not call on any theory of the logistic distribution beyond the easily checked fact that it has variance $\pi^2/3$.

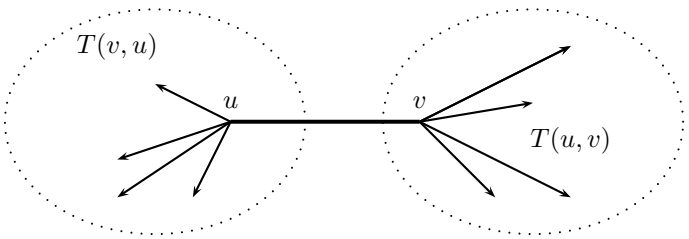

Figure 10: When one cuts a PWIT at an edge, one obtains two trees that one can label by using the directed edges associated with the cut edge. The tree $T(u,v)$ labelled by the directed edge (u,v) has the root v and contains all of the edges that follow v in the direction of (u,v).

5.6 A Stochastic Process that Constructs a Matching

With the explicit solution of the distributional identity (5.58) in our hands, the next step is to construct a joint process where the distributional identity (5.58) suggests an inclusion criterion that can be used to define a good matching on the PWIT. Specifically, we will construct simultaneously (1) a PWIT and (2) a special collection of random variables that are associated with the edges of the PWIT. The variables in this special collection are intended to mimic the differences $D(T^j)$ that led us to the logistic recursion. To put this plan into action, we will need to introduce some notation that makes more explicit the consequences of cutting a PWIT at one of its edges.

Notation for a Split PWIT

When we take a PWIT and delete the edge (u,v) we get two disjoint trees that we can view as being rooted at u and v respectively. To help keep track of these trees (and the difference functionals that we will associate with them), we assign to each edge of the PWIT a pair of *directed* edges.; formally, we just distinguish between (u,v) and (v,u). Given any graph G we let $\mathbf{E}_G$ denote the set of edges obtained by directing the edges of G, and, if T is a PWIT with edge weights ξ_e, then we give the same weight to each of the directed edges associated with e. Finally, if $\vec{e} = (u,v)$ is a directed edge and $X(\mathbf{e})$ is a random variable associated with that edge, we will often simply write $X(u,v)$. In the case of the PWIT, the weight of the directed edge $\xi_{\vec{e}}$ will also be written as $\xi(u,v)$. By definition we have $\xi(u,v) = \xi(v,u)$, but, typically, $X(u,v)$ and $X(v,u)$ will be different.

Lemma 9 (Triple Tree Process) *There exists a process $(\mathcal{T}, \xi_{\vec{e}}, X(\vec{e}) : \vec{e} \in \mathbf{E}_{\mathcal{T}})$ where $\mathcal{T}$ is a PWIT with associated edge lengths $\{\, \xi_{\vec{e}} : \vec{e} \in \mathbf{E}_{\mathcal{T}} \}$ and where $\{\, X(\vec{e}) : \vec{e} \in \mathbf{E}_{\mathcal{T}} \}$ is stochastic process indexed by the directed edges of $\mathcal{T}$ with the three following properties:*

(i) for each directed edge $\vec{e} = (u,v) \in \mathbf{E}_{\mathcal{T}}$ we have

$$X(u,v) = \min\{\xi(v,w) - X(v,w) : (v,w) \in \mathbf{E}_{\mathcal{T}},\ w \neq u\}, \tag{5.61}$$

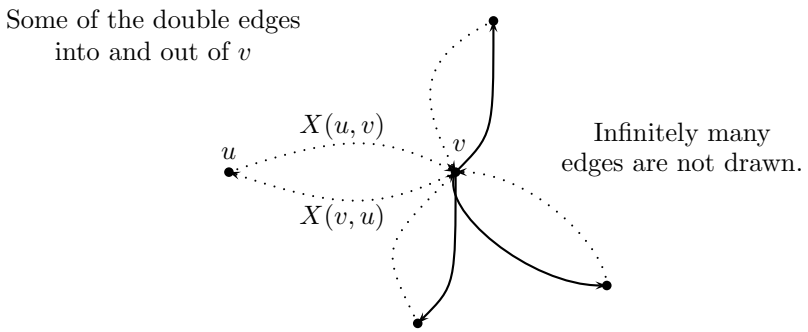

Figure 11: The edges that feature in the identity for $X(u,v)$ are those directed edges out of v indicated by the solid arrows; note that the edge (v,u) is not one of these. Note also that the weight ξ_e of an edge does not depend on its direction, but the value of X does.

(ii) for each $\vec{e} \in \mathbf{E}_{\mathcal{T}}$ that is directed away from the root of $\mathcal{T}$ the random variable $X(\vec{e})$ has the logistic distribution, and

(iii) for each pair of oppositely directed edges (u,v) and (v,u) in $\mathbf{E}_{\mathcal{T}}$, the random variables $X(u,v)$ and $X(v,u)$ are independent.

The proof of Lemma 9 is easy and instructive; it shows how one can exploit a distributional identity to construct useful processes on trees and other graphs. To begin, we fix an integer $g \geq 1$ and then for each directed edge (u,v) with u in generation g and v in generation $g+1$ (relative to the root), we create by fiat an independent logistic random variable $X(u,v)$. We next use the identity (5.61) to define $X(\mathbf{e})$ for each edge within g generations of the root. This construction gives us a collection $\mathcal{C}_g$ of random variables whose joint distribution satisfies properties (i), (ii) and (iii) for all vertices in the first g generations. Moreover, the distribution of the collection $\mathcal{C}_g$ is equal to the marginal distribution of the larger collection $\mathcal{C}_{g+1}$ restricted to the first g generations, so by the Kolmogorov consistency theorem there is a collection $\mathcal{C}_\infty$ such that for each g the marginal distribution of $\mathcal{C}_\infty$ restricted to first g generations is equal to the distribution of $\mathcal{C}_g$.

A Useful Addendum to the Triple Tree Lemma

Once the triple process is in hand, one can go further; in fact, one can show that the process has some intrinsic independence properties that are not an explicit part of the construction. Moreover, we will find that this additional independence has a key role in our later computations.

First we take a fixed $0 < x < \infty$, and we condition on the event that there exist an edge a the root that has length x. We call this edge (r,v) and note

that as in Figure 10, the edge determines two subtrees of the PWIT that one could label $T(r,v)$ and $T(v,r)$. As we have noted before, a Poisson process conditioned to have a point at x is again a Poisson process when this point is deleted, so by the definition of the PWIT we see that $T(u,v)$ and $T(v,u)$ are conditionally independent copies of the original PWIT. We therefore find an addendum to Lemma 9:

> (iv) conditional on the existence of an edge (r,v) from the root having length x, the random variables $X(r,v)$ and $X(v,r)$ are independent random variables with the logistic distribution.

Construction of an Excellent Matching on $\mathcal{T}$

For the triple process $(\mathcal{T},\xi,X)$ given by Lemma 9, one can use the additional information carried by the X process to specify a good matching on $\mathcal{T}$. As before, the idea is to use an *inclusion criterion* that is motivated by the heuristic derivation of the distributional identity (5.58).

To define this matching, we first consider the map $\varphi : V \to V$ on the set V of vertices of the PWIT that is given by $\varphi(v) = v^*$ where v^* is the vertex for which $\xi_e - X(\mathbf{e})$ is minimized when $\mathbf{e} = (v,v^*)$. The next lemma confirms that the set of edges (v,v^*) is a matching; we will subsequently show that it is a good one. More precisely, one can show that this matching is optimal in the sense that no other involution invariant matching of a PWIT can have a lower average expected length per edge.

Lemma 10 (Matching Lemma) *The set of edges (v,v^*) defined by the rule $\varphi(v) = v^*$ is a matching on the special PWIT $\mathcal{T}$.*

To prove Lemma 10, we first note that it is exactly equivalent to showing that φ is idempotent; that is, we must show that $(v^*)^* = v$ for all $v \in V$. Now, by the definition of φ and the recursion (5.61) for X, we have

$$\begin{aligned}\xi(v,v^*) - X(v,v^*) &= \min\{\xi(v,y) - X(v,y) : (v,y) \in \mathcal{T}\}\\ &< \min\{\xi(v,y) - X(v,y) : (v,y) \in \mathcal{T},\, y \neq v^*\}\\ &= X(v^*,v),\end{aligned}$$

or, a bit more simply

$$\xi(v,v^*) < X(v,v^*) + X(v^*,v). \tag{5.62}$$

On the other hand, if z is a neighbor of v other than v^*, then we have

$$\begin{aligned}\xi(v,z) - X(v,z) &> \min\{\xi(v,y) - X(v,y) : (v,y) \in \mathcal{T}\}\\ &= \min\{\xi(v,y) - X(v,y) : (v,y) \in \mathcal{T},\, y \neq z\}\\ &= X(z,v).\end{aligned}$$

This inequality tells us $\xi(v,z) > X(v,z) + X(z,v)$ whenever $z \neq v$, so we see that v^* is the *unique* neighbor of v that satisfies the inclusion criterion (5.62).

We can therefore see that we could have just as well *defined* $\varphi(v)$ as the unique v^* satisfying the criterion (5.62). Finally, if one notes that both sides of the criterion (5.62) are symmetric we see that $\varphi(\varphi(v)) = v$, just as we hoped to show.

Wisdom of a Matching Rule

Given the maneuvering that led us to the matching rule $r^* = \varphi(r)$, one has a right to expect that there is no invariant matching that has a lower expected cost per vertex. The next lemma asserts that this expectation is justified.

Lemma 11 (Optimality on the PWIT) *Let M denote any matching of $\mathcal{T}$ that is involution invariant. If (r, s) is the edge of M that contains the root r of $\mathcal{T}$, then one has*

$$E[\xi(r, s)] \geq E[\xi(r, r^*)]. \tag{5.63}$$

The proof of Lemma 11 is unfortunately too lengthy to be included here, and even a convincing sketch would run beyond our boundaries, so for most details we must refer the reader to Proposition 18 of Aldous [4], which additionally proves that inequality (5.63) is strict except when the matching M is almost surely identical to the matching defined by φ.

5.7 Calculation of a Limiting Constant: $\pi^2/6$

If M_{opt} is the matching given by Lemma 10 and if $r^* = \varphi(r)$ is the vertex that is matched to the root r by M_{opt}, then the expected value per edge of the weight of the M_{opt} is simply equal to $E[\xi(r, r^*)]$. This expected value turns out to be surprisingly easy to compute.

Lemma 12

$$E[\xi(r, r^*)] = \frac{\pi^2}{6}$$

To prove the lemma we first fix $0 < x < \infty$ and then we condition on the event A_x that there exists an edge (r, v) incident to the root of $\mathcal{T}$ that has length x. The inclusion criterion (5.62) then tells us r is matched to v if and only if we have

$$x < X(r, v) + X(v, r).$$

By the addendum to Lemma 9 we know that conditional on the event A_x the pair $\big(X(r, v), X(v, r)\big)$ same distribution as a pair (X, Y) of independent logistics, so by the uniform intensity of the Poisson process on $[0, \infty)$ we have

$$E[\xi(r, r^*)] = \int_0^\infty x\, P(x < X + Y)\, dx = \frac{1}{2} E[(X + Y)^2 \mathbf{1}(X + Y > 0)].$$

Finally, the symmetry of the logistic distribution lets us rewrite the last expectation as

$$\frac{1}{4} E[(X + Y)^2] = \frac{1}{2} E[X^2] = \pi^2/6,$$

where in the last step we used the fact noted earlier that the logistic distribution has variance $\pi^2/3$.

5.8 Passage from a PWIT Matching to a K_n Matching

We are now in position to sketch the connections that complete the proof of Theorem 5, and we begin by noting that the observations in hand already give us one half of the theorem. For each n we let M_n denote the optimal matching on K_n and we view M_n as a $\mathcal{G}_\star$-valued random variable by setting the root of M_n to be the root of K_n. By the tightness of the joint process (M_n, K_n) we know that for every subsequence of $n = 1, 2, ...$ there is a further subsequence n_j such that (M_{n_j}, K_{n_j}) converges weakly in $\mathcal{G}_\star \times \mathcal{G}_\star$. By Skorohod's theorem and our earlier analysis of the weak limit of K_n, we may even assume that

$$(M_{n_j}, K_{n_j}) \to (\mathcal{M}, \mathcal{T}) \quad \text{with probability one,} \tag{5.64}$$

where $\mathcal{T}$ is a PWIT and $\mathcal{M}$ is a matching on $\mathcal{T}$. Moreover, from the fact that M_n is randomly rooted, we see that $\mathcal{M}$ is involution invariant.

Now, if $c(M_n)$ denotes the cost of matching the root in M_n and $c(\mathcal{M})$ denotes the cost of the edges that meets the root of $\mathcal{T}$, then from the limit (5.64) and the definition of the topology on $\mathcal{G}_\star$, we have

$$c(M_{n_j}) \to c(\mathcal{M}) \quad \text{with probability one.}$$

From this limit, Fatou's Lemma, and Lemma 12 we then find

$$\liminf_{n\to\infty} E[c(M_{n_j})] \geq E[c(\mathcal{M})] \geq \pi^2/6.$$

Finally, by the generality of the subsequence $\{n_j\}$, we therefore find in the notation of Theorem 5 that

$$\liminf_{n\to\infty} \frac{2}{n} E[C_n] \geq \pi^2/6. \tag{5.65}$$

This proves half of the equality (5.41) of Theorem 5; the bad news is that this is the easy half.

The Harder Half — and Its Two Parts

The harder half of Theorem 5 breaks into two natural steps. First one needs to show that there is a low-cost ε-feasible matching for K_n for large values of n, and then one needs to show that such an ε-feasible matching can be modified to provide an honest matching without substantially increasing the total cost.

To make this plan precise, we first recall that an ε-feasible matching of a graph G with n vertices is a subset M^* of edges of G with the property that each vertex of G is in some edge of M^* and at most εn elements of M^* fail to be isolated from the other elements of M^*. The fact that one can cheaply convert a low-cost ε-feasible matching of K_n to an honest low-cost complete matching of K_n is (almost) purely combinatorial. We do not deal here with that part of the argument, but one can refer to Proposition 2 of Aldous [3] for a proof of an analogous result in the context of bipartite matching.

How to Use a Matching on a PWIT to Find One on K_n

We continue to work with the special PWIT $\mathcal{T}$ that was constructed as part of Triple Tree Process of Lemma 9, and we let $\mathcal{M}$ denote the matching defined on $\mathcal{T}$ that was constructed in Lemma 10. Next, we fix a $\varrho > 0$, and we let $N_\varrho(r, \mathcal{T})$ denote the neighborhood of radius ϱ about the root r of $\mathcal{T}$. We view $N_\varrho(r, \mathcal{T})$ as an element of $\mathcal{G}_\star$ with root r, and we let $N_\varrho(r, K_n)$ denote the analogously defined ϱ-neighborhood of the root of K_n. Here one should note that $N_\varrho(r, \mathcal{T})$ is always a tree since it is a subtree of $\mathcal{T}$, but the corresponding neighborhood of K_n given by $N_\varrho(r, K_n)$ will not necessarily be a tree. Nevertheless, since ϱ is fixed, the probability that $N_\varrho(r, K_n)$ is a tree will go to one as n becomes large.

Now, for any $G \in \mathcal{G}_\star$ that is a feasible realization of $N_\varrho(r, \mathcal{T})$ and for any $w \in \mathcal{T}$, we let $q(w|G)$ denote the conditional probability that r is matched to w in $\mathcal{M}$ given the event that the neighborhood $N_\varrho(r, \mathcal{T})$ is isomorphic to G as an element of $\mathcal{G}_\star$. We will use $q(w|G)$ to guide a sampling process on K_n that we can use to help build a matching.

Specifically, for any $1 \leq n < \infty$ we define a subset $\mathcal{M}(\varrho, n)$ of the *directed edges* of K_n by the following process: for each $u \in K_n$ we choose (u, v) with probability $q(v|\, N_\varrho(u, K_n)\,)$. Also, in just the same way, we define a subset $\mathcal{M}(\varrho, \infty)$ of the *directed edges* of $\mathcal{T}$ — that is, for each $u \in \mathcal{T}$ we choose the directed edge (u, v) with probability $q(v|\, N_\varrho(u, \mathcal{T}))$. The key observation is that the event $A(\varrho, \infty)$ defined by

$$A(\varrho, \infty) = \{\, r \text{ is in exactly two edges of } \mathcal{M}(\varrho, \infty)\,\}$$

satisfies the identity

$$P(A(\varrho, \infty)) = 1 - \eta(\varrho) \quad \text{where } \eta(\varrho) \to 0 \text{ as } \varrho \to \infty. \tag{5.66}$$

Thus, if we collapse each directed loop $\{(u, v), (v, u)\}$ of $\mathcal{M}(\varrho, \infty)$ to a single undirected edge, then for large ϱ we have a high probability of obtaining an honest matching on $\mathcal{T}$.

Now, we consider the analogous events for K_n. Specifically, we let $A(\varrho, n)$ denote the event that (1) the root r of K_n is in exactly two directed edges of $\mathcal{M}(\varrho, n)$ and (2) these edges have the form (r, v) and (v, r) for some v. The local weak convergence of K_n to $\mathcal{T}$ and the definition of $\eta(\varrho)$ then tells us that

$$\lim_{n\to\infty} P(\, A(\varrho, n)\,) = 1 - \eta(\varrho). \tag{5.67}$$

Next, we let $\mathcal{M}^u(\varrho, n)$ denote the set of *undirected* edges one obtains by collapsing each directed loop $\{(v, w), (w, v)\}$ of $\mathcal{M}(\varrho, n)$ into a single undirected edge. One can then show that the limit (5.67) implies that there is an $\varepsilon(\varrho)$ such that $\varepsilon(\varrho) \to 0$ as $\varrho \to \infty$ for which we have

$$\liminf_{n\to\infty} P(\mathcal{M}^u(\varrho, n) \text{ is } \varepsilon(\varrho)\text{-feasible}) \geq 1 - \varepsilon(\varrho). \tag{5.68}$$

Finally, to complete our sketch of the proof of Theorem 5, we just note that one can now use local weak convergence to relate the costs of the edges of $\mathcal{M}^u(\varrho, n)$ to the costs of those in the matching $\mathcal{M}$ on $\mathcal{T}$, and, in due course, one can use

this correspondence to show that the asymptotic mean cost of $\mathcal{M}^u(\varrho, n)$ is at most $n/2$ times $\pi^2/6$. This is precisely what one was brought to expect from the calculation in Lemma 12, and thus it completes the sketch.

When one looks back on this construction, one can confirm that the basic idea of sampling from a locally defined conditional distribution is pleasantly general. In fact, there are many problems where a similar sampling method can be applied, and this method of conditional sampling seems to express one of the more fundamental ways that one can exploit local weak convergence. Further elaboration of the sampling construction is expected to be part of Aldous and Steele [7].

5.9 Finally — Living Beyond One's Means

One of the original motivations for the study of the assignment problem was the discovery of Karp [37] that under the model of independent uniform costs c_{ij} the value of A_n provides an impressively tight lower bound on the minimal cost of a *tour* of the vertex set of the complete directed graph where the cost of using the directed edge (i, j) is c_{ij}. This discovery led in turn to sustained exploration of algorithms that use the assignment problem and branch and bound techniques for the travelling salesman problem (cf. Balas and Toth[13]), and it also motivated more detailed investigations of the precision of A_n as an approximation for the TSP cost (cf. Karp and Steele [39]).

While the objective method is particularly well focused on issues such as the convergence of the means $E(A_n)$ and the determination of this limit, it does not provide any information on the variance of A_n or other measures of concentration of A_n. For such information, other methods must be applied.

For example, Karp and Steele [39] used combinatorial and coupling methods to show that with high probability the greatest cost of an edge in an optimal assignment of size n is $O(\log^2 n/n)$, and Talagrand [62] used this observation to illustrate one of his isoperimetric inequalities to show that there is a constant $0 < c < \infty$ such that

$$\mathrm{Var}(A_n) \leq \frac{c}{n} (\log n)^4 \quad \text{for all } n \geq 1.$$

More recently, Frieze and Sorkin [23] improved the bounds of Karp and Steele [39] to show that with high probability the longest edge in $O(\log n/n)$, and the same technique was then further refined by Lee and Su [45] to show that there are positive constants c_1 and c_2 such that

$$c_1 n^{-5/2} \leq \mathrm{Var}(A_n) \leq \frac{c_2}{n} (\log n)^2. \tag{5.69}$$

Despite the wide gap between the bounds (5.69) they are the best that are currently known. Nevertheless, one certainly expects that in fact there is $\sigma^2 > 0$ such that

$$\mathrm{Var}(A_n) \sim \frac{\sigma^2}{n} \quad \text{as } n \to \infty. \tag{5.70}$$

Unfortunately, the objective method is powerless to help, and for the moment at least, the conjecture (5.70) seems to be far over the horizon.

6 Problems in Euclidean Space

The objective method is especially well suited for many problems of Euclidean combinatorial optimization. Typically, one begins with an optimization problem that is driven by a sample $\{X_i : 1 \le i \le n\}$ of independent random variables with the uniform distribution on the unit d-cube $[0,1]^d$, so after choosing a random "root" I that is uniformly distributed on $\{1, 2, ..., n\}$ and independent of the $\{X_i\}$, one can take immediate advantage of the classical fact that

$$n^{1/d}(X_i - X_I : 1 \le i \le n) \xrightarrow{d} \mathcal{P}_0 \quad \text{as } n \to \infty, \tag{6.71}$$

where $\mathcal{P}_0 = \mathcal{P} \cup \{0\}$ and $\mathcal{P}$ is a homogenous Poisson process on $\mathbb{R}^d$ with uniform intensity equal to 1.

Now, throughout this survey, we have interpreted the indicated convergence as local weak convergence as defined in Section 2, so it may seem unfair to call this result classical. Nevertheless, once one agrees to interpret the two sides of the limit (6.71) as rooted geometric graphs where the root is the vertex at the origin of $\mathbb{R}^d$ and where the edge lengths are given by Euclidean distances, then (up to an irrelevant rotation) local weak convergence (6.71) is in fact equivalent to the traditional weak convergence of point process theory.

The bottom line is that for problems of combinatorial optimization for points in $\mathbb{R}^d$, the Poisson limit (6.71) assumes the role that Grimmett's Lemma 1 played for the uniform trees and that Aldous's Theorem 3 played for the exponentially weighted complete graphs. Thus, for Euclidean problems the first step of the objective method comes *gratis*; the game only begins with the introduction of an optimization problem.

6.1 A Motivating Problem

One of the early motivations for the objective method was a problem sparked by an empirical observation of R. Bland, who noted from simulations of an algorithm for the MST that for n independent uniformly distributed points in the unit square that one consistently finds that

$$\sum_{e \in MST} |e|^2 \to c > 0 \quad \text{as } n \to \infty,$$

where $|e|$ denotes the usual Euclidean distance from u to v when $e = (u, v)$. Stimulated by Bland's observation, Steele [58] proved that if the random variables $\{X_i\}$ are independent and have a distribution μ with compact support, then for all $0 < \alpha < d$ one actually has

$$n^{(\alpha-d)/d} \sum_{e \in MST} |e|^\alpha \to c(\alpha, d) \int_{\mathbb{R}^d} f(x)^{(d-\alpha)/d}\, dx \quad \text{a.s.}, \tag{6.72}$$

where f is the density of the absolutely continuous part of μ and $c(\alpha, d)$ is a constant that depends only on α and d. This result goes beyond Bland's

conjecture in several respects, but, ironically, the subadditive method used in Steele [58] falls short of covering the motivating case $\alpha = d$. This is where the objective method entered the picture.

Benefits of Hindsight — and A Proper Point of View

Fifteen years ago it was not clear how one could make further progress on Bland's conjecture, but with the benefit of hindsight a good plan is obvious. One simply needs to extend the Poisson convergence (6.71) to the joint convergence the point sets and the corresponding spanning trees.

When this was done in Aldous and Steele [6], there was some mystery to emergence of a *forest* as the limit of a sequence of minimal spanning trees. Now the reasons for the appearance of the minimal spanning forest are understood more deeply, and, if one simply applies the MST Convergence Theorem of Section 5.3 to the Poisson limit (6.71), one obtains a theorem that explains most of the qualitative behavior of the MST of a random sample from the d-cube.

Theorem 7 *For the randomly rooted normalized sample*

$$S_n = n^{1/d}(X_i - X_I : 1 \le i \le n)$$

of independent random variables with the uniform distribution on $[0,1]^d$*, one has the joint local weak convergence*

$$\big(\, S_n,\, MST(S_n)\,\big) \;\xrightarrow{d}\; \big(\, \mathcal{P}_0,\, MSF(\mathcal{P}_0)\,\big). \tag{6.73}$$

A Question of Means

On a conceptual level, the limit (6.73) tells one almost everything there is to say about $MST(S_n)$ for large n. Nevertheless, there are inevitably technical hurdles to be cleared when one applies Theorem 7 to a concrete problem. For example, to obtain the convergence of the mean of $MST(S_n)$, one needs some information on the uniform integrability of the set of variables $\{MST(S_n) : n = 1, 2, ..\}$.

Uniform integrability of this sequence was proved in Aldous and Steele [6] by direct estimates, and later the "lens geometry" of the MST and spacefilling curves were used in Chapter 5 of Steele [60] to give more inclusive moment bounds. In the end, one finds for all $p \ge 1$ and all d there is a constant $\beta_d > 0$ such that for random uniform samples of size n from the d-cube one has

$$\sum_{e \in MST} |e|^d \to \beta_d \quad \text{in } L^p \text{ as } n \to \infty.$$

Moreover, one can identify the limit constant β_d in terms of the limit object by the formula

$$\beta_d = \frac{1}{2} E\Big[\sum_{e \in R} |e|^d\Big] < \infty \tag{6.74}$$

where R is the set of edges of the MSF of $\mathcal{P}_0$ that are incident to the root at the origin of $\mathbb{R}^d$.

The objective method almost always gives one a characterization of the limit constant in terms of the limit object, and we have seen examples in Sections 3 through 5 where this characterization could be pressed to provide an exact determination of the limit constant. In this respect, Euclidean problems are different. Despite tantalizing characterizations such as that provided by formula (6.74), none of the limit constants of finite dimensional Euclidean optimization have been determined analytically. The reason for this failing seems to be that in the Euclidean case, one does not have the self-similarity that is crucial to the recursions that lead us to the limit constants in results like Theorem 2 and Theorem 5.

6.2 Far Away Places and Their Influence

Throughout the theory of Euclidean combinatorial optimization, one finds hints that tractable problems must be *local* in some sense; or, looked at the other way around, one finds that in tractable problems the points that are far away cannot have much influence on the problem's solution in the neighborhood of a near-by point. This intuition can be formalized in several ways, but the recently introduced notion of *stability* is perhaps most in tune with the philosophy of the objective method.

Sums of Local Contributions

To begin, we consider nonnegative functions of the form $\xi(x;\mathcal{X})$ where $x \in \mathbb{R}^d$ and $\mathcal{X}$ is a subset of $\mathbb{R}^d$ that is locally finite in the sense that each bounded subset of $\mathbb{R}^d$ contains only finitely many elements of $\mathcal{X}$. Also, if a is a positive scalar and $y \in \mathbb{R}^d$ we denote the usual dilations and translations of $\mathcal{X}$ by

$$a\mathcal{X} := \{ax : x \in \mathcal{X}\} \quad \text{and} \quad y + \mathcal{X} := \{y + x : x \in \mathcal{X}\}.$$

If we assume ξ is translation invariant in the sense that

$$\xi(y + x; y + \mathcal{X}) = \xi(x;\mathcal{X}) \quad \text{for all } y \text{ and } \mathcal{X},$$

then the sum $H_\xi(\mathcal{X})$ defined by setting

$$H_\xi(\mathcal{X}) := \sum_{x\in\mathcal{X}} \xi(x;\mathcal{X}) \tag{6.75}$$

is also translation invariant in the sense that $H_\xi(y+\mathcal{X}) = H_\xi(\mathcal{X})$ for all $y \in \mathbb{R}^d$. Moreover, sums of this form (6.75) may be used to represent almost all of the functionals of interest in Euclidean combinatorial optimization.

For example, the length of the MST of $\mathcal{X}$ may be written as $H_\xi(\mathcal{X})$ if we take $\xi(x;\mathcal{X})$ to be half of the sum of the edges of the MST that are incident to x, so

$$\xi(x;\mathcal{X}) = \frac{1}{2}\sum_{y\in\mathcal{X}} |x-y|\mathbf{1}(\,(x,y) \in MST(\mathcal{X})\,).$$

For a less obvious example, consider the representation $H_\xi(\mathcal{X})$ of the number of connected components of the nearest neighbor graph of $\mathcal{X}$. In this case we can take

$$\xi(x;\mathcal{X}) = 1/\mathrm{card}(C_x)$$

where C_x is the component of the nearest neighbor graph of $\mathcal{X}$ that contains the vertex x. Essentially all of the Euclidean functionals discussed in Steele [60] or Yukich [65] may be represented as $H_\xi(\mathcal{X})$ for an appropriate choice of $\xi(x;\mathcal{X})$.

Stable Functionals and the Influence Function

One benefit of the representation $H_\xi(\mathcal{X})$ is that the summands $\xi(x;\mathcal{X})$ suggest a natural way to formalize the idea of local dependence. To explain this idea, we first take $r > 0$ and let $B(x;r)$ denote the Euclidean ball of radius r about x, then, for any locally finite point set $\mathcal{S} \subset \mathbb{R}^d$ and any integer $m \geq 1$ we set

$$\overline{\xi}(\mathcal{S};m) := \sup_{n\in\mathbb{N}} \big(\text{ess sup}_{m,n}\{\xi(0;(\mathcal{S}\cap B(0;m))\cup\mathcal{A})\}\big)$$

and where ess $\sup_{m,n}$ denotes the essential supremum with respect to Lebesgue measure on $\mathbb{R}^{dn}$ and where $\mathcal{A} \subset \mathbb{R}^d \setminus B(0;m)$ is a set of cardinality n. Analogously, we define

$$\underline{\xi}(\mathcal{S};m) := \inf_{n\in\mathbb{N}} \big(\text{ess inf}_{m,n}\{\xi(0;(\mathcal{S}\cap B(0;m))\cup\mathcal{A})\}\big),$$

and we think of and $\underline{\xi}(\mathcal{S};m)$ and $\overline{\xi}(\mathcal{S};m)$ as measures of the changes that one may make by a finite additions to $\mathcal{S}$ outside a ball of radius m. The interesting case occurs when these two possibilities are comparable for large m.

Definition 5 (Stable Functionals and the Influence Function)
The functional ξ is said to be stable *on the locally finite set $\mathcal{S}$ provided that one has*

$$\lim_{m\to\infty} \overline{\xi}(\mathcal{S};m) = \lim_{m\to\infty} \underline{\xi}(\mathcal{S};m). \tag{6.76}$$

Moreover, when ξ is stable on $\mathcal{S}$, we let $\xi_\infty(\mathcal{S})$ denote the value of the common limit and we call $\xi_\infty(\mathcal{S})$ the influence function *of ξ on $\mathcal{S}$.*

The locally finite sets that are of most interest to us are naturally the realizations of homogeneous Poisson processes, and one should note that if $\mathcal{P}_\tau$ is the homogeneous Poisson process with constant intensity τ on $\mathbb{R}^d$ then the sequence of random variables $\{\overline{\xi}(\mathcal{P}_\tau;m) : m = 1,2,...\}$ is nonincreasing and the sequence $\{\underline{\xi}(\mathcal{P}_\tau;m) : m = 1,2,...\}$ is nondecreasing. Thus, both sequences converge almost surely, and we see that ξ is almost surely stable on $\mathcal{P}_\tau$ if the two limits are equal with probability one. Many of the natural functionals $\xi(x;\mathcal{X})$ of Euclidean combinatorial optimization have this property, and there are some useful consequences of this fact.

A Weak Law of Large Numbers

Now consider a sequence $X_1, X_2, \ldots$ of independent d-dimensional random variables that have a common density f. Next, set $\mathcal{X}_n = \{X_1, ..., X_n\}$ and consider the scaled summand

$$\xi_n(x; \mathcal{X}) := \xi(n^{1/d}x; n^{1/d}\mathcal{X}_n) \tag{6.77}$$

together with the corresponding scaled sums

$$\widetilde{H}_\xi(\mathcal{X}_n) := \sum_{x \in \mathcal{X}_n} \xi_n(x; \mathcal{X}_n) = \sum_{x \in \mathcal{X}_n} \xi(n^{1/d}x; n^{1/d}\mathcal{X}_n).$$

When ξ is stable for almost every realization of any homogeneous Poisson process, these sums typically satisfy a weak of large numbers; in fact, this notion of stability was introduced in Penrose and Yukich [53] for the purpose of framing such weak laws. The next theorem is typical of the results obtained there.

Theorem 8 *For each constant $0 < \tau < \infty$ we suppose that ξ is almost surely stable for the homogenous Poisson process $\mathcal{P}_\tau$ with intensity τ, and we let $\xi_\infty(\mathcal{P}_\tau)$ denote the corresponding influence function. If the sequence*

$$\left\{ \xi(n^{1/d}X_1; n^{1/d}\mathcal{X}_n) : 1 \le n < \infty \right\} \quad \textit{is uniformly integrable,} \tag{6.78}$$

where observations in the sample $\mathcal{X}_n = \{X_1, X_2, ..., X_n\}$ are i.i.d. with density f on $\mathbb{R}^d$, then the influence function satisfies

$$I[\tau] = E[\xi_\infty(\mathcal{P}_\tau)] < \infty \quad \textit{for each } 0 < \tau < \infty \textit{ such that } f(\tau) > 0, \tag{6.79}$$

and the normalized sums satisfy

$$n^{-1}\widetilde{H}_\xi(\mathcal{X}_n) \to \int_{\mathbb{R}^d} I[f(x)]f(x)dx \quad \text{in } L^1. \tag{6.80}$$

Consequences of Homogeneity

Many of the functionals addressed by Theorem 8 have the property that there is a constant $\gamma > 0$ such that

$$\xi(ax; a\mathcal{X}) = a^\gamma \xi(x; \mathcal{X}) \quad \text{for all } a \in \mathbb{R}^+,$$

and, in this case the righthand side of the limit (6.80) is more simply written as

$$E[\xi_\infty(\mathcal{P}_1)] \int_{\mathbb{R}^d} f(x)^{(d-\gamma)/d} dx. \tag{6.81}$$

Moreover, when ξ is *scale invariant*, or homogeneous of order 0, the limit (6.81) simply boils down to $E[\xi_\infty(\mathcal{P}_1)]$, and the limit given by Theorem 8 does not depend on the density of underlying point set. Formula (6.81) also makes it clear that although ξ_∞ may at first seem subtle to define, it has an inevitable place in the limit theory of the normalized sums $\tilde{H}_\xi$.

Confirmation of Stability

Theorem 8 applies to many problems of computational and it provides limit laws for the minimal spanning tree, the k-nearest neighbors graph ([34], [20]), the Voronoi graph [32], the Delaunay graph, and the sphere of influence graphs ([27], [33]). Nevertheless, before Theorem 8 can be applied, one needs to prove the associated summands $\xi(x, \mathcal{X})$ are indeed stable on $\mathcal{P}_\tau$ for all $0 \leq \tau < \infty$.

In some cases, stability may be proved easily, but, in other cases, the proof may be quite challenging. For example, even the stability of the MST is not easy without the help of some powerful tools. For example, Penrose and Yukich [53] prove stability MST with help from the idea of a *blocking set*, which is a set of points that effectively isolates the MST in a neighborhood of the origin from any influence by points outside of a certain large disk. This basic structure was introduce by Kesten and Lee [40] (cf. Lemmas 3 and 5), and it provides one of the keys to the of the central limit theory for the MST as further developed by Lee ([43], [44]).

A Cox Process Coupling

Finally, we should note that the proof of Theorem 8 given in Penrose and Yukich [53] calls on an interesting coupling argument that is likely to have further applications. The key observation is that one may simulate samples that closely approximate the samples from a general density f with help from a Cox process that has a random *but conditionally uniform* intensity. This approximation provides one with a new way to convert results for uniform samples to results for nonuniform samples, and, in the right circumstances, it seems to have advantages over older methods where one first approximates f by a locally flat density and then creates a coupling. Nevertheless, each method ultimately relies on some sort of smoothness of the sums $\widetilde{H}_\xi$.

6.3 Euclidean Methods and Some Observations in Passing

There are so many connections between percolation theory, minimum spanning trees, and the objective method that we cannot survey them in detail, but we do want to review a few developments that seem particularly informative. The first of these is a recent observation of Bezuidenhout, Grimmett, and Löffler [16] that draws a simple connection between MSTs and the notion of free energy that originates in statistical mechanics.

MSTs, Free Energy, and Percolation Connections

To begin, if V is a subset of $\mathbb{R}^d$ and $\alpha > 0$, we let $\Gamma(\alpha, V)$ denote the graph with vertex set V and edge set consisting of pairs (x, y) of elements of V such that $|x - y| \leq \alpha$. Next, we let $\mathcal{P}$ denotes the homogeneous Poisson process on $\mathbb{R}^d$ with unit intensity, and take $S(\alpha)$ to be the size of the connected component of $\Gamma(\alpha, \mathcal{P} \cup \{0\})$ that contains 0, the origin of $\mathbb{R}^d$, and, finally, we define $\kappa(\alpha)$, the *free energy of the Poisson process*, by setting

$$\kappa(\alpha) = E\big[1/S(\alpha)\big].$$

The free energy turns out to have a remarkably simple relationship to the empirical distribution of the edge lengths of the minimal spanning tree T_n of the set $V_n = \mathcal{P} \cap [-n, n]^d$. Specifically, if we consider the empirical distribution of the set of edge lengths of T_n,

$$F_n(\alpha) = (|V_n| - 1)^{-1} \sum_{e \in T_n} \mathbf{1}(|e| \leq \alpha),$$

then we have the succinct limit

$$\lim_{n \to \infty} F_n(\alpha) = 1 - \kappa(\alpha) \quad \text{a.s. and in } L^1. \tag{6.82}$$

This engaging identity may be explained in just a few lines. The key observation is that if $\nu(\alpha, n)$ denotes the number of connected components of the graph $\Gamma(\alpha, \mathcal{P} \cap [-n, n]^d)$, then by counting the missing edges we have

$$F_n(\alpha) = \frac{|V_n| - \nu(\alpha, n)}{|V_n| - 1}. \tag{6.83}$$

To finish off, one now just needs to note that the Ergodic Theorem for the Poisson process can be used to show

$$\lim_{n \to \infty} \nu(\alpha, n)/|V_n| = E\big[1/S(\alpha)\big] := \kappa(\alpha) \quad \text{a.s. and in } L^1,$$

so the limit (6.82) follows immediately from the identity (6.83).

Naturally, if one is to make good use of this observation more work is needed, and the real effort in Bezuidenhout, Grimmett, and Löffler [16] goes into the development of the analytical properties of the free energy and its analogs for Bernoulli site percolation. In that setting, they show that the free energy defect $H(\alpha) = 1 - \kappa(\alpha)$ is indeed a distribution function, and they also show that one has the convergence of the corresponding mth moments for all $0 \leq m < \infty$,

$$\int_0^\infty \alpha^m dF_n(\alpha) \to \int_0^\infty \alpha^m dH(\alpha) \quad \text{a.s. and in } L^1.$$

A Second Percolation Connection — Forest vs. Tree

When the minimal spanning forest for the Poisson process was introduce in Aldous and Steele [6], there was a strong suspicion that the minimal spanning forest might actually be a tree — at least in low dimensions. This conjecture was confirmed in dimension $d = 2$ by Alexander [8], Example 2.8, but, the question remains open for dimension $d \geq 3$.

The suspicion now is that the MSF of the Poisson process is not a tree if the dimension is sufficiently high, and Theorem 2.5, part (ii), of Alexander [8] provides some indirect evidence for this conjecture. Because of results of Hara and Slade ([28], [29]) that show one does not have percolation at the critical point for dimension $d > 19$ for Bernoulli bond percolation, Alexander's theorem implies that in high dimensions a related MSF does not have any doubly infinite paths. If one relies on the analogies between Bernoulli bond percolation, Bernoulli site percolation, and continuum percolation, Alexander's result suggests in turn that if the MSF of the Poisson process is a tree, then in high dimensions that tree must consist of a sequence of finite trees that are connected to one singly infinite path. Finally, for such a structure to cover all of the point of a Poisson process, this path would then be forced to wander around space in a way that does not seem feasible.

Inclusion Criteria and Computational Complexity

Almost all of the probability problems of Euclidean combinatorial optimization may be framed in terms of a graph G_n that depends on a random sample $\{X_1, X_2, ..., X_n\}$ in $\mathbb{R}^d$, and for many of these problems the subadditive method introduced by Beardwood, Halton and Hammersley [14] provides a competing technology to the objective method. Although the objective method has intuitive and analytical advantages in almost every case where it applies, there are still instances where the only concrete results have been found with the subadditive method.

For example, subadditive methods work quite naturally for traveling salesman problem (TSP), but when one tries to apply the objective method to the TSP one some deep and interesting questions emerge. Specifically, one needs to define an analog $\mathcal{T}$ of the traveling salesman tour for the Poisson process in $\mathbb{R}^d$, and in practical terms this means one needs to specify an inclusion criterion that tells one when an edge (x, y) is in $\mathcal{T}$. We were able to provide such an inclusion criterion for the maximal matching problem, and classical results for the MST suggested the defining criterion for the minimal spanning forest. On the other hand, the theory of computational complexity suggest that one is unlikely to find an inclusion criterion that will provide a suitable analog for the TSP of the Poisson process. The basic intuition is that no problem that is NP-complete for finite sets of points can be expected to have an inclusion criterion that provides a suitable limit object on the Poisson process.

Further Comparison to Subadditive Methods

In addition to the examples suggested by computational complexity, there are other cases where subadditive methods seem to have an advantage over the objective method. One such situation is provided by the asymptotic analysis of heuristic methods, and the greedy matching problem [11] is perhaps the most natural example.

Nevertheless, there are surely cases where subadditive methods have been used in the past, but where the objective now provides an easier or more informative approach. Confirmed cases on this list include the limit theory for MST with power weighted edges [58] and MST vertex degrees [61]; and, with some systematic effort, the list can probably be extended to include the theory of Euclidean semi-matchings [59], optimal cost triangulations [56] and the K-median problem [35]. Moreover, there are many cases where the objective method quickly gives one the essential limit theory, yet subadditive methods appear to be awkward to apply. Here the list can be made as long as one likes, but one should certainly include the limit theory for Voronoi regions [32] and the sphere of influence graphs ([27], [33]).

6.4 Recurrence of Random Walks in Limits of Planar Graphs

Most of the local weak limits that we have surveyed here have their origins in combinatorial optimization, but the objective method and the theory of local weak convergence can also be applied to problems that are more purely probabilistic. One striking example is given by the recurrence theory for random walks on planar graphs recently given by Benjamini and Schramm [15].

Theorem 9 *Suppose that G is a $\mathcal{G}_\star$-valued random variable that is a weak limit in $\mathcal{G}_\star$ of a sequence $\{G_n[X_n] : n = 1, 2, ...\}$ of uniformly rooted random graphs where*

(1) for each $n \geq 1$, the graph G_n is planar with probability one, and

(2) $\max\{\deg(v) : v \in G_n\} \leq M < \infty$ *for all* $n \geq 1$,

then with probability one, the random walk on the graph $G(\omega)$ is recurrent.

This theorem is clearly part of the theory of local weak convergence, but the theorem and its proof differ greatly the other results in this survey. Even a sketch of the proof of Theorem 9 would take us far off our path, but we should at least note that one key to the proof is the remarkable Circle Packing Theorem of Koebe [41]. This theorem says that if a finite graph $G = (V, E)$ is planar, then there exist a collection C of circles in the plane with disjoint interiors and a one-to-one correspondence $\phi : V \to C$ such that $(x, y) \in E$ if and only if the circles $\phi(x)$ and $\phi(y)$ make contact.

From the perspective of this survey, there are two attractive features of Theorem 9. First, its very difference serves as a reminder that there must be other, as yet unimagined, areas where the theory of local weak convergence can play a natural role. Second, Theorem 9 suggest a useful generic question: What are the properties of a sequence of random rooted finite graphs that are

preserved under local weak convergence? In many cases this question is bound to yield only pedestrian answers, but the Benjamini-Schramm Theorem shows that there are indeed cases when the generic question can lead one to remarkable results.

7 Limitations, Challenges, and Perspectives

This treatment of the objective method is part survey and part tutorial. The basic intention has been to make the objective method more accessible, while also illustrating its effectiveness on problems with some subtlety and substance. Any survey has an element of advocacy, and here we certainly hoped to show the objective method in a positive light. Nevertheless, one should be aware of the method's limitations.

Some Intrinsic Limitations

First, the objective method deals with reasonably complex entities such as the Skeleton Tree $PGW^{\infty}(1)$, the PWIT, or — de minimus — the Poisson process in $\mathbb{R}^d$. In many cases, one can make immediate and effective use of these off-the-shelf structures, but in more original problems, one is forced to invent (or discover) new objects. Any such object begins life as the weak limit in $\mathcal{G}_\star$ of a sequence of random finite graphs with uniformly distributed random roots, but, if it is to be used with effect, one almost always needs an independent characterization. This highly constructive feature of the objective method means that it seldom provides short proofs. Ironically, this remains so even when the result that one wants to prove is "conceptually clear" from the point of view offered by the limiting object.

Second, when one focuses directly on a limit object, one tends to lose almost all trace of how one gets to that limit. As a consequence, the objective method is typically ill suited for obtaining information about rates of convergence. To be sure, there are special circumstances where rates may be found; for example, the recursion used in our analysis of the maximum partial matching problem provides at least some rate information. Nevertheless, this is an exception that reflects the hybrid nature of one particular problem.

Third, the theory of local weak convergence has a substantial limitation on the type of information that it can provide. Basically, the method only addresses those problems that may be reduced to a calculation in a neighborhood of a randomly chosen root. For example, we are able to address total-sum-of-length problems only because of a lucky piece of arithmetic; the sum of the lengths of the edges in a graph with n vertices is just n times the expected value of half of the sum of the lengths of the edges incident to a randomly chosen root.

Many similar sounding problems are not so lucky. For example, consider the problem of calculating the expected length of the longest path in the MST of a uniform random sample of size n from the unit square. This problem sounds quite similar to the total length question, but, on reflection, one finds that it cannot be reduced to a calculation that just depends on a neighborhood of a

random root; thus, it falls outside of the scope of the objective method. For the same reason, the objective method cannot help us show that the variance of A_n is asymptotic to σ^2/n, nor can it help with more general measures of concentration. Local weak convergence is always a prisoner that cannot escape the neighborhoods of a random root.

A Current Challenge: The Largest Independent Set

The limitations on the objective method are substantial, but one still finds a wide range of challenges for which the objective method offers the most likely path for progress. Here we will just discuss one problem that we find especially attractive.

An *independent set* S in a graph G is collection of vertices such that no two elements of S are joined by an edge of G, and the *independence number* $\alpha(G)$ is the maximum cardinality of an independent set in G. The independence number is one of the most studied quantities of graph theory, and it also leads one to an interesting test of potential of the objective method.

For any integer $k \geq 3$ there is a standard and model for a *random k-regular graph* with n vertices, and the properties of these graphs a studied in detail in the monographs of Bollobás [17] and Janson, et.al. [36]. Now, if one starts to think about the independence number of a random k-regular graph $G_{k,n}$, one almost immediately comes to conjecture that there exists a constant $\alpha_k > 0$ such that

$$\lim_{n\to\infty} n^{-1} E[\alpha(G_{k,n})] = \alpha_k. \tag{7.84}$$

The origins of this conjecture are lost in the mists of time, and no doubt it has occurred independently to many people. From the perspective of this survey, one of its interesting features is that it offers several parallels to the random assignment problem as that problem was understood circa 1986.

For example, no one has yet proved that indicated limit (7.84) actually exists — yet no one doubts that it does. Moreover, some concrete limits have already been place on the possible size of α_k. Specifically, Frieze and Suen [24] studied the greedy algorithm for constructing an independent set for $G_{3,n}$, and they have proved

$$\liminf_{n\to\infty} n^{-1} E[\alpha(G_{3,n})] \geq 6\log(3/2) - 2 = 0.432\cdots \tag{7.85}$$

Finally, the methods of statistical mechanics have already found a role. In particular, Hartmann and Weigt [31] have used the replica method to study the independence number in a model that is closely related to $G_{k,n}$.

When we view conjecture (7.84) from the perspective of the objective method, we do not need long to hit on a promising plan. In fact, we can take one step almost automatically, since it is reasonably obvious that $G_{k,n}$ converges in $\mathcal{G}_\star$ to the randomly rooted infinite k-regular tree $\mathcal{T}$. Thus, the intrigue only begins once one starts to look for a large involution invariant independent subset S of $\mathcal{T}$, and, remarkably enough, an attractive candidate comes to mind almost immediately.

To define our candidate S, we first flip a coin. If it comes up heads, we take the root r to be an element of S, and then we add to S every other vertex on the paths that descend from r, but if the coin comes up tails, we take all of the children of r to be in S and then add to S every other vertex on the paths that descend from these children. The set S is clearly an independent set in $\mathcal{T}$, and one can check that S is involution invariant. No checking is needed to see that S has density one-half, and, since no independent set can have density larger than one-half, the principles developed in Section 3 might seem to suggest that we are well on our way to proving that $\alpha_r = 1/2$ for all r.

Unfortunately, we meet a bump in the road. Our candidate S cannot be realized as the weak limit of a sequence of independent subsets of $G_{n,k}$, even though S does satisfy the necessary condition of involution invariance. The problem with S is that it exhibits a extreme form of long-range dependence, and, in this particular case, one can exploit that dependence to show that S is not the weak limit of sequence independent subsets of $G_{n,k}$ as $n \to \infty$.

Thus, we come back to the challenge of defining an independent subset $\mathcal{T}$ that can be realized as the limit of independent subsets of the sequence $G_{n,k}$, $n = 1, 2, \ldots$. One natural thought is that it may be possible to characterize α_k as the maximum density of an involution invariant random independent set in $\mathcal{T}$ that does not exhibit some particular type of long-range dependence, but, so far, there are only the faintest hints how one might specify the precise form of dependence that one should rule out.

Some Final Perspectives

One can never know how a mathematical theory will evolve, and even well-informed guesses are still just guesses. Nevertheless, the objective method and the theory of local weak convergence have a certain inevitability to them that invites speculation.

At the most concrete level, the objective method helps one to decouple two basic tasks: the proof of the existence of a limit and the determination of a limit constant. This contrasts with analytic methods where techniques based on recursions and generating functions typically address these two tasks simultaneously. It also contrasts with subadditive methods where one may prove the existence of a limit and then be left without any serious indication of its numerical value.

At a more abstract level, the objective method helps one focus on the most essential elements of a complex limit theorem. Just as one finds in the traditional theory of weak convergence that Brownian motion appears as the limit of many different processes, one finds in the theory of local weak convergence that basic objects such as the skeleton tree and the PWIT appear as the limit of many different processes. These new structures do not have the amazing universality of Brownian motion, yet they do have *some* measure of universality. Moreover, even though these semi-discrete structures are simpler than Brownian motion, they still have substantial internal structure that can support concrete computation; the calculation in Lemma 4 showing that $\zeta(3)$ is the average length per vertex of the PWIT provides one example. Finally, the list of objects is surely

far from exhausted, and perhaps, it has only been begun.

One surely expects to see new local weak limits with all the charm of the skeleton tree or the PWIT. Recent support for this view is offered by Angel and Schramm (2002) which proves local weak convergence of uniform random triangulations of the sphere with n vertices to a limit random triangulation of the plane and by Gamarnik (2002) where arguments motivated by the PWIT-based analysis of the random assignment are applied to a linear programming relaxation of random K-SAT.

Finally, we should note that the objective method is tightly bound with the theory of recursive distributional equations. For example, the fixed point equations (3.12) and (5.58) were at the heart of our analysis of the minimum cost perfect matching problems, and this situation seems to be typical. In fact, equations that characterize an unknown distribution by equality in distributional with a function of a collection of independent copies of itself arise in a wide variety of probabilistic settings. Aldous and Bandyopadhyay (2002) provide a survey of such equations, both with and without direct connections to the objective method.

Acknowledgements: The authors are pleased to thank Kenneth Alexander, Harry Kesten, Sungchul Lee, Mathew Penrose, Gordon Slade, Z. Su, and Joseph Yukich for kindly providing preprints, references, and comments. The research of David Aldous is supported by the NSF under Grant DMS-9970901.

References

[1] Aldous, D.J. (1990): A random tree model associated with random graphs. Random Structures Algorithms, **1**, 383–402.

[2] Aldous, D.J. (1991): Asymptotic fringe distributions for general families of random trees. Ann. Appl. Probab. **1**, 228–266.

[3] Aldous, D.J. (1992): Asymptotics in the random assignment problem. Probab. Th. Rel. Fields, **93**, 507–534.

[4] Aldous, D.J. (2001): The $\zeta(2)$ limit in the random assignment problem. Random Structures Algorithms, **18**, 381–418.

[5] Aldous, D.J. and Bandyopadhyay, A. (2002): A Survey of Max-type Recursive Distributional Equations. Technical Report, U.C. Berkeley.

[6] Aldous, D.J. and Steele, J.M. (1992): Asymptotics of Euclidean Minimal Spanning Trees on Random Samples. Probab. Th. Rel. Fields, **92** 247–258.

[7] Aldous, D.J. and Steele, J.M. (2002): The asymptotic essential uniqueness property for minimal spanning trees on random points, manuscipt in preparation.

[8] Alexander, K.S. (1995) Percolation and minimal spanning forest in infinite graphs. Ann. Probab. **23**, 87–104.

[9] Alm, S.E. and Sorkin, G.B. (2002): Exact expectations and distributions for the random assigmnent problem. Combinatorics, Probability, and Computing (to appear).

[10] Angel, O. and Schramm, O. (2002): Uniform Infinite Planar Triangulations, arXiv:math.PR/0207153.

[11] Avis, D., Davis, B., and Steele, J.M. (1988): Probabilistic analysis of a greedy heuristic for Euclidean matching. Probability in the Engineering and Information Sciences, **2**, 143–156.

[12] Avram, F. and Bertsimas, D. (1992): The minimum spanning tree constant in geometric probability and under the independent model: a unified approach. Annals of Applied Probability, **2**, 113–130.

[13] Balas, E. and Toth, P. (1985): Branch and bound methods. In The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimzation. Lawler, E.L, Lenstra, J.K, Rinnooy Kan, A.H.G., and Smoys, D.B. (eds), Wiley, NY.

[14] Beardwood, J., Halton, J.H., and Hammersley, J.M. (1959): The shortest path through many points. Proceedings of the Cambridge Philosphical Society, **55**, 299–327.

[15] Benjamini, I. and Schramm, O. (2001): Recurrence of distributional limits of finite planar graphs. Electronic Joural of Probability, **6**, Paper No. 23, 1–13.

[16] Bezuidenhout, C., Grimmett, G., and Löffler, A. (1998): Percolation and minimal spanning trees. J. Statist. Phys., **92**, 1-34.

[17] Bollobás, B. (1985): Random Graphs. Academic Press, London.

[18] Devroye, L. (1998): Branching processes and their application in the analysis of tree structures and tree algorithms. In M. Habib, editor, Probabilistic Methods for Algorithmic Discrete Mathematics, Springer-Verlag.

[19] Dyer, M. E., Frieze, A. M., and McDiarmid, C. (1986): On linear programs with random costs. Math. Programming, **35**, 3–16.

[20] Eppstein, D., Paterson, M.S. and Yao, F.F. (1997): On nearest-neighbor graphs. Discrete Comput. Geom., **17**, 263–282.

[21] Frieze, A.M. (1985): On the value of a random minimum spanning tree problem. Discrete Appl. Math., **10**, 47–56.

[22] Frieze, A.M. and McDiarmid, C.J.H. (1989): On random minimum lenght spanning trees. Combinatorica, **9**, 363–374.

[23] Frieze, A. and Sorkin, G.B. (2001): The probabilistic relationship between the assignment problem and asymmetric traveling salesman problems. Proceedings of SODA, ACM Publishers, 652–660.

[24] Frieze, A. and Suen, S. (1994): On the independence number of random cubic graphs. Random Structures and Algorithms, **5**, 640–664.

[25] Gamarnik, D. (2202): Linear Phase Transition in Random Linear Constraint Satisfaction Problem. Technical Report: IBM T.J. Watson Research Center.

[26] Grimmett, G.R. (1980): Random labelled trees and their branching networks. J. Austral. Math. Soc. (Ser. A), **30**, 229–237.

[27] Füredi, Z. (1995): The expected size of a random sphere of influence graph, *Intuitive Geometry*, Bolyai Mathematical Society, **6**, 319-326.

[28] Hara, T. and Slade, G. (1990): Mean-field critical behaviour for percolation in high dimensions. Commun. Math. Phys., **128**, 333–391.

[29] Hara, T. and Slade, G. (1994): Mean-field behaviour and the lace expansion. Probability and Phase Transition (G. Grimmett, ed.), Kluwer, Dordrecht.

[30] Harris, T.H. (1989): The Theory of Branching Processes. Dover Publications, New York.

[31] Hartmann, A.K. and Weigt, M. (2001): Statistical mechanics perspective on the phase transition in vertex covering of finite-connectivity random graphs. Theoretical Computer Science, **265**, 199–225.

[32] Hayen, A. and Quine, M.P. (2000): The proportion of triangles in a Poisson-Voronoi tessellation of the plane. Adv. in Appl. Probab., **32**, 67-74.

[33] Hitczenko, P., Janson, S., and Yukich, J.E. (1999): On the variance of the random sphere of influence graph. Random Structures Algorithms, **14**, 139–152.

[34] Henze, N. (1987): On the fraction of random points with specified nearest neighbor interrelations and degree of attraction. Adv. Appl. Prob., **19**, 873–895.

[35] Hochbaum, D. and Steele, J.M. (1982): Steinhaus' geometric location problem for random samples in the plane. Advances in Applied Probability, **14**, 55–67.

[36] Janson, S., Luczak, T., and Ruciński (2000): Random Graphs. Wiley Interscience Publishers, New York.

[37] Karp, R.M. (1979): A patching algorithm for the non-symmetric traveling salesman problem. SIAM Journal on Computing, **8**, 561–573.

[38] Karp, R.M. (1987): An upper bound on the expected cost of an optimal assignment. In: Discrete Algorithms and Complexity: Proceedings of the Japan-U. S. Joint Seminar, Academic Press, New York.

[39] Karp, R.M. and Steele, J.M. (1985): Probabilistic analysis of heuristics. In The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimzation. Lawler, E.L, Lenstra, J.K, Rinnooy Kan, A.H.G., and Smoys, D.B. (eds), Wiley, NY, 181–205.

[40] Kesten, H. and Lee, S. (1996): The central limit theorem for weighted minimal spanning trees on random points. Annals of Applied Probability, **6**, 495–527.

[41] Koebe, P. (1936): Kontaktprobleme der Konformen Abbildung. Ber. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Kl. **88**, 141–164.

[42] Kurtzberg, J. M. (1962): On approximation methods for the assignment problem. J. Assoc. Comput. Mach., **9**, 419–439.

[43] Lee, S. (1997): The central limit theorem for Euclidean minimal spanning trees I. Annals of Applied Probability, **7**, 996–1020.

[44] Lee, S. (1999): The central limit theorem for Euclidean minimal spanning trees II., Advances in Applied Probability, **31**, 969–984.

[45] Lee, S. and Su, Z. (2001): On the fluctutation in the random assigmnent problem. Technical Report, Yonsei University Department of Mathematics, April 2001.

[46] Lovász, L. and Plummer, M.D. (1986): Matching Theory. Annals of Discrete Mathematics, vol. 29. North-Holland Publishers. Amsterdam.

[47] Mahmoud, H.M. (1992): Evolution of Random Search Trees, Wiley, New York, 1992.

[48] Matoušek, J. and Neštřil, Y. (1998): Discrete Mathematics, Oxford Universtity Press, Oxford.

[49] Meyn, S.P. and Tweedie, R.L. (1993): Markov Chains and Stochastic Stability, Springer-Verlag, New York.

[50] Mézard, M. and Parisi, G. (1987): On the solution of the random link matching problem. J. Physique, **48**, 1451–1459.

[51] Parisi, G. (1998): A conjecture on radom bipartite matching. ArXiv Condmat 9801176.

[52] Penrose, M.D. (1996): The random minimal spanning tree in high dimensions. Ann of Probability, **24** 1903–1925.

[53] Penrose, M.D. and Yukich, J.E. (2002): Weak laws of large numbers geometric probability, Annals of Applied Probability, to appear.

[54] Propp, J. and Wilson D. (1998): Coupling from the past: a user's guide, In D. Aldous and J. Propp, editors, Microsurveys in Discrete Probability, number 41 in DIMACS Ser. Discrete Math. Theoret. Comp. Sci., pages 181–192.

[55] Steele, J.M. (1981): Subadditive Euclidean functionals and non-linear growth in geometric probability. Annals of Probability, **9**, 365–376.

[56] Steele, J.M. (1982): Optimal triangulation of random samples in the plane. *Annals of Probability*, **10**, 548–553.

[57] Steele, J.M. (1987): On Frieze's $\zeta(3)$ limit for the lenths of minimal spanning trees, Discrete Applied Mathematics, **18**, 99–103.

[58] Steele, J.M. (1988): Growth rates of Euclidean minimal spanning trees with power weighted edges. *Annals of Probability*, **16**, 1767–1787, 1988.

[59] Steele, J.M. (1992): Euclidean semi-matchings of random samples. *Mathematical Programming*, **53**, 127–146, 1992.

[60] Steele, J.M. (1997): Probability Theory and Combinatorial Optimization, NSF-CBMS Volume 69. Society for Industrial and Applied Mathematics, Philadelphia.

[61] Steele, J.M., Shepp, L.A. J.M. Eddy, W. (1987): On the number of leaves of a Euclidean minimal spanning tree. J. Appl. Probab., **24**, 809–826.

[62] Talagrand, M. (1995): Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. IHES, **81**, 73–205.

[63] Vitter, J.S. and Flajolet, P. (1990): Analysis of algorithms and data structures. In *Handbook of Theoretical Computer Science*, volume A: Algorithms and Complexity (Chapter 9), North-Holland, 431–524.

[64] Walkup, D. W. (1979): On the expected value of a random assigmnent problem. SIAM J. Comput., **8**, 440–442.

[65] Yukich, J.E. (1998): Probability Theory of Classical Euclidean Optimization Problems, Lecture Notes in Mathematics, **1675**, Springer-Verlag, New York.