The Annals of Applied Probability
2014, Vol. 24, No. 6, 2414–2454
DOI: 10.1214/13-AAP981

BELIEF PROPAGATION FOR OPTIMAL EDGE COVER IN THE RANDOM COMPLETE GRAPH

BY MUSTAFA KHANDWAWALA[1] AND RAJESH SUNDARESAN[1,2]

Indian Institute of Science

We apply the objective method of Aldous to the problem of finding the minimum-cost edge cover of the complete graph with random independent and identically distributed edge costs. The limit, as the number of vertices goes to infinity, of the expected minimum cost for this problem is known via a combinatorial approach of Hessler and Wästlund. We provide a proof of this result using the machinery of the objective method and local weak convergence, which was used to prove the $\zeta(2)$ limit of the random assignment problem. A proof via the objective method is useful because it provides us with more information on the nature of the edge's incident on a typical root in the minimum-cost edge cover. We further show that a belief propagation algorithm converges asymptotically to the optimal solution. This can be applied in a computational linguistics problem of semantic projection. The belief propagation algorithm yields a near optimal solution with lesser complexity than the known best algorithms designed for optimality in worst-case settings.

1. Introduction. Suppose that we are given a graph G with vertex set V and edge set E, denoted $G = (V, E)$. Each edge $e \in E$ has a weight $\xi_e \in \mathbf{R}_+$. Alternatively, we are given a bipartite graph with a vertex set $V = V_1 \cup V_2$, a union of two disjoint vertex subsets, and an edge set $E \subset V_1 \times V_2$. An *edge cover* for the graph is a subset of edges that hits (covers) every vertex. The cost of an edge cover is the sum of the weights of edges in the cover. Our interest in this paper is on minimum-cost edge covers on the complete graph (denoted K_n when $|V| = n$) and on the complete bipartite graph (denoted $K_{n,n}$ when $|V_1| = |V_2| = n$), when the edge weights are independent random variables, each with the exponential distribution of mean 1.

The following example on a bipartite graph illustrates how minimum-cost edge covers arise in practice.

Received December 2012; revised October 2013.

[1]Supported by the Department of Science and Technology, Government of India, by a TCS fellowship grant.

[2]Supported by an Indo-US Science and Technology Forum research fellowship grant.

MSC2010 subject classifications. Primary 60C05; secondary 68Q87, 82B44.

Key words and phrases. Belief propagation, edge cover, local weak convergence, objective method, semantic projection.

An example of semantic projection. Computational linguists have recently been interested in machine-based natural language processing. These include part-of-speech tagging, parsing, and at a higher level, semantic role parsing [12] which, for example, would enable an automatic recognition that the sentences "Mary sold the book to John" and "The book was sold by Mary to John" have the same semantic roles. (This example is taken from Wikipedia [19].) Currently, English is blessed with the availability of a large amount of annotated texts as training data while most others languages lack this advantage. Semantic projection exploits the availability of (1) parallel corpora of translated texts and (2) higher quality parsing tools in one language in order to transfer annotations from the resource-rich language to the other.

Padó and Lapata [12] provide one method to do this where a minimum-cost edge cover naturally arises. The source and target sentences in the two languages are first broken into *linguistic units* to yield sets V_1 and V_2 of the respective linguistic units. These linguistic units are then viewed as vertices of a complete bipartite graph. Let R be some finite set of *semantic roles*, which can be viewed for our purposes as abstract annotations. The parsing tool on the source side is used to find a semantic role assignment $\mathsf{role}_1 : R \to 2^{V_1}$, where the subscript refers to the source language. A *dissimilarity* measure based on linguistic considerations is then assigned to every pair of linguistic units across the languages and is denoted $\xi : V_1 \times V_2 \to \mathbf{R}_+$. A decision procedure uses these dissimilarity scores to find a subset $\mathcal{C} \subset V_1 \times V_2$ of semantically aligned units. Padó and Lapata [12] argue that a minimum-cost edge cover is a good choice for this semantic alignment. It allows a linguistic unit in one language (an element of say V_2) to map to several units in the other language (a subset of V_1), and vice-versa. For example, the linguistic units "to be on time" and "punctual" (English) could both be mapped with small, but possibly different, dissimilarity scores to "pünktlich" (German), and both edges may be picked by a good candidate edge cover. The covering property of the edge cover enables all source and target vertices to participate and thus has the potential to capture important connections between linguistic units, which may otherwise be missed. The minimum cost property attempts to provide an economical semantic alignment and further captures global alignments as compared to previously proposed local decision procedures. Once the minimum-cost edge cover is found by the decision procedure, semantic roles are then assigned on the target side as

$$\mathsf{role}_2(r) = \{j \mid \text{there is an } i \in \mathsf{role}_1(r) \text{ such that } (i, j) \in \mathcal{C}\}.$$

Padó and Lapata [12] compare the goodness of their decision procedures based on minimum-cost edge cover (and perfect matching) with some other prior approaches on a data set of about 1000 sentences. Real data sets are of course much larger. The resulting graph, when restricted to edges of small weight (i.e., edges signifying low dissimilarity and therefore good correspondence), can be modeled as a large, but sparse, random graph. If $|V_1| = O(|V_2|) = n$, algorithms used by

Padó and Lapata [12] to find the minimum-cost edge cover take $O(n^3)$ operations, in the worst case.

The actual results of the Padó and Lapata [12] experiments need not concern us here. For a list of challenges that arise in the implementation of the above approach and methods to address them, we refer the linguistically inclined reader to [13] and references therein. What we shall take with us as we move forward are the observations that (1) edge covers arise in practice on large graphs that can be modeled by sparse random graphs and (2) algorithmic simplifications that reduce complexity are of practical value.

We shall for simplicity focus on minimum-cost edge covers on the complete graph K_n on n vertices. All our results carry over to $K_{n,n}$ with only scaling factor modifications. Recall that the edge capacities are independent, each edge having the exponential distribution with mean 1. This is a typical *mean-field* model which captures sparsity of the graph depicting linguistic units and associated edges in the above example, but ignores correlations among edge weights. See Section 11 for another geometric setting where the same mean field models arise. Let C_n be the cost of the minimum-cost edge cover of K_n. We prove that the expected value of C_n converges to the constant $W(1) + W(1)^2/2$, which is approximately 0.728. (The function $W(\cdot)$ is Lambert's W-function, which is the inverse of $f : [0, \infty) \to [0, \infty)$, $f(x) = xe^x$; $W(1) \approx 0.567$.) Further, and more importantly from an application perspective, we show that a belief propagation algorithm can be used to find asymptotically optimal edge covers in $O(n^2)$ steps. The results, with only scaling factor changes, hold for the complete bipartite graphs $K_{n,n}$.

The result regarding the limit on $K_{n,n}$ has been proved before by Hessler and Wästlund in [10] using a combinatorial approach. A proof based on a game formulation is contained in [16]. We discuss these works at the end of this Introduction. Our focus in this article is on using the *objective method* for this problem and on devising a belief propagation algorithm.

The roots of the objective method lie in Aldous's 1992 paper [1] on the assignment problem. The problem of finding the minimum cost matching on the complete bipartite graph with independent and identically distributed edge costs, termed as the random assignment problem in literature, inspired a series of works in combinatorial probability. Mézard and Parisi [11], using the *cavity method* of statistical physics, conjectured in 1987 that the expected minimum cost for the random assignment problem on the bipartite graph $K_{n,n}$ converges to $\zeta(2) = \sum_{k=1}^{\infty} k^{-2}$ as n goes to infinity. This was proved rigorously by Aldous [4] in 2001 by extending the proof of existence of the limit contained in [1]. Several other proofs have been provided for the limit in subsequent works.

In [4], Aldous related the problem on $K_{n,n}$ to one on a suitable limit object. Several calculations become easier on the limit object. In this case, the limit is a tree, the so-called Poisson weighted infinite tree or PWIT, with many useful symmetries. Aldous used these symmetries to construct a *distributional identity*, that then served as a guide for solving the random assignment problem rigorously. With this approach, Aldous showed that the following quantities converge to the corresponding quantities on the limit object:

- the expected cost of optimal matching on $K_{n,n}$;
- the distribution of the cost of the matching edge incident on a typical node of $K_{n,n}$;
- the probability that the matching edge incident on a typical node of $K_{n,n}$ is the kth smallest of all the edges incident on it.

It turns out that the limit object, and hence the answers, remain the same for problems on the complete bipartite graph $K_{n,n}$ and on the complete graph K_n. One dividend of a proof via the objective method is that we have answers to several ancillary questions such as the second and third bullets above. The ability of the objective method to provide these auxiliary results motivates us to solve the problem of optimal edge cover via the objective method.

From an algorithms perspective, the cavity equations suggest a natural iterative decentralized message passing algorithm, some versions of which are commonly called belief propagation (BP) in the computer science literature. For many combinatorial optimization problems, a BP algorithm can be set up to converge to the correct solution on graphs without cycles. Bayati, Shah and Sharma [7] proved that the BP algorithm for maximum weight matching on bipartite graphs converges to the correct value as long as the maximum weight matching is unique. Salez and Shah [14] studied the random assignment problem and proved a tighter connection with the limit object. They showed that that a BP algorithm on $K_{n,n}$ converges to an update rule on the limit PWIT of [4]. The iterates on the limit graph converge in distribution to the minimum cost assignment. The iterates are near the optimal solution in $O(n^2)$ steps, whereas the worst case optimal algorithm on bipartite graphs is $O(n^3)$ [expected time $O(n^2 \log n)$ for i.i.d. edge capacities]; see Salez and Shah [14] and references therein. We show a similar complexity improvement for the edge-cover problem.

The objective method is quite powerful to be applicable to several combinatorial probability problems. See Aldous and Steele [3] for a survey. Aldous and Bandopadhyay [5], Section 7.5, outline the steps of Aldous's program to establish the validity of the cavity method, which we quote in Section 11. However, each problem requires specific proofs, and we are still far from a complete theory applicable to a wide class of problems. The edge-cover problem itself poses some modest problem-specific challenges which we overcome in this paper. These include (1) a proof of existence and uniqueness of a solution to the distributional identity associated with the edge-cover problem, (2) a proof of a property called *endogeny* of a process on the tree associated with the distributional identity, (3) a proof of optimality of the edge-cover selection on the PWIT as suggested by the distributional identity and eventually (4) a proof that a BP algorithm converges to an asymptotically optimal edge cover on the random complete graph. See Section 11 for a more detailed summary.

Before we end this introduction, we would like to mention two other approaches that have been used to solve related combinatorial optimization problems, in particular, matching, edge cover and travelling salesman problems. One approach used by Wästlund in [16, 18] calls for a "boundary conditioning" parameter to study "diluted" versions of the optimization problems, eventually driving the parameter to infinity, and thereby relating the resulting limiting problem with the undiluted versions. For example, in the matching case, diluted matching is a partial matching with each unmatched vertex paying a cost equal to the parameter. Wästlund then formulates the optimization problem in terms of a game played on the graph. A second and more combinatorial approach is used by Wästlund in [17] for matching and TSP and in [10] for the edge-cover problem. These works study the respective optimization problems as certain flow problems on bipartite graphs. The feasible solutions to these flow problems have a fixed number of edges k. A recursive relation on k is obtained for the cost of the optimal solution. As our focus is on the objective method, we do not dwell any more on these approaches.

2. Main results.

Our first result establishes the limit of the expected minimum cost of the random edge-cover problem.

THEOREM 1. *On K_n, we have*

$$\lim_{n\to\infty} \mathrm{E}C_n = W(1) + \frac{W(1)^2}{2}. \tag{1}$$

Our second result shows that a belief propagation algorithm gives an edge cover that is asymptotically optimal as $n \to \infty$. We will use the result that the update rule of BP converges to an update rule on a limit infinite tree. For this we define the BP algorithm on an arbitrary graph $G = (V, E)$ with edge costs. For an edge $e = \{v, w\} \in E$, we write its cost as $\xi_G(e)$ or $\xi_G(v, w)$. For each vertex $v \in V$, we associate a nonempty subset of its neighbors $\pi_G(v)$. By taking a union of all edges of the form $\{v, w\}$, $w \in \pi_G(v)$, we get an edge cover of G which we will denote by $\mathcal{C}(\pi_G)$.

The BP algorithm is an iterative message passing algorithm. In each iteration $k \geq 0$, every vertex $v \in V$ sends a message $X_G^k(w, v)$ to each neighbor $w \sim v$ according to the following rules:

Initialization:

$$X_G^0(w, v) = 0. \tag{2}$$

Update rule:

$$X_G^{k+1}(w, v) = \min_{u\sim v, u\neq w} \{(\xi_G(v, u) - X_G^k(v, u))^+\}. \tag{3}$$

Decision rule:

$$\pi_G^k(v) = \arg\min_{u\sim v}\{(\xi_G(v,u) - X_G^k(v,u))^+\}, \tag{4}$$

$$\text{Edge cover} = \mathcal{C}(\pi_G^k(v)). \tag{5}$$

We analyze the belief propagation algorithm for $G = K_n$ and i.i.d. exponential random edge costs, and prove that after a sufficiently large number of iterates, the expected cost of the assignment given by the BP algorithm is close to the limit value in Theorem 1.

THEOREM 2. *On K_n, we have*

$$\lim_{k\to\infty}\lim_{n\to\infty} \mathrm{E}\Bigg[\sum_{e\in\mathcal{C}(\pi_{K_n}^k)} \xi_{K_n}(e)\Bigg] = W(1) + \frac{W(1)^2}{2}. \tag{6}$$

The formal statements on the bipartite complete graph $K_{n,n}$ with i.i.d. expontial distribution of mean 1 are the following and are stated without proof.

THEOREM 3. *On $K_{n,n}$, we have*

$$\lim_{n\to\infty} \mathrm{E}C_n = 2W(1) + W(1)^2. \tag{7}$$

THEOREM 4. *On $K_{n,n}$, we have*

$$\lim_{k\to\infty}\lim_{n\to\infty} \mathrm{E}\Bigg[\sum_{e\in\mathcal{C}(\pi_{K_{n,n}}^k)} \xi_{K_{n,n}}(e)\Bigg] = 2W(1) + W(1)^2. \tag{8}$$

3. Local weak convergence. In this section, we recollect the terminology for defining convergence of graphs.

3.1. *Rooted geometric networks.* A graph $G = (V, E)$ along with a length function $l : E \to (0, \infty]$ is called a *network*. The *distance* between two vertices in the network is the infimum of the sum of lengths of the edges of a path connecting the two vertices, the infimum being taken over all such paths. We call the network a *geometric network* if for each vertex $v \in V$ and positive real ρ, the number of vertices within a distance ρ of v is finite. We denote the space of geometric networks by $\mathcal{G}$.

A geometric network with a distinguished vertex v is called a *rooted geometric network* with root v. We denote the space of all connected rooted geometric networks by $\mathcal{G}_*$. In $\mathcal{G}_*$ we do not distinguish between rooted isomorphisms of the same network. We will use the notation (G, o) to denote an element of $\mathcal{G}_*$ which is the isomorphism class of rooted networks with underlying network G and root o.

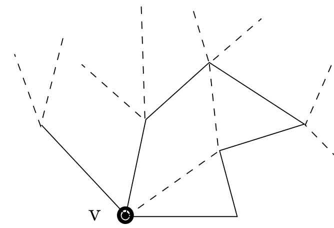

FIG. 1. *Neighborhood $\mathcal{N}_\rho(G)$ of graph G. The solid edges form the neighborhood, and form paths of length at most ρ from the root v. Dashed edges are the other edges of G.*

3.2. *Local weak convergence.* We call a positive real number ρ a *continuity point* of G if no vertex of G is exactly at a distance ρ from the root of G. Let $\mathcal{N}_\rho(G)$ denote the neighborhood of the root of G up to distance ρ. $\mathcal{N}_\rho(G)$ contains all vertices of G which are within a distance ρ from the root of G (Figure 1). We take $\mathcal{N}_\rho(G)$ to be an element of $\mathcal{G}_*$ by inheriting the same length function l as G, and the same root as that of G.

We say that a sequence of rooted geometric networks $G_n, n \geq 1$, *converges locally* to an element G_∞ in $\mathcal{G}_*$ if for each continuity point ρ of G_∞, there is an n_ρ such that for all $n \geq n_\rho$, there exists a graph isomorphism $\gamma_{n,\rho}$ from $\mathcal{N}_\rho(G_\infty)$ to $\mathcal{N}_\rho(G_n)$ that maps the root of the former to the root of the latter, and for each edge e of $\mathcal{N}_\rho(G_\infty)$, the length of $\gamma_{n,\rho}(e)$ converges to the length of e as $n \to \infty$.

The space $\mathcal{G}_*$ can be suitably metrized to make it a separable and complete metric space. One can then consider probability measures on this space and endow that space with the topology of weak convergence of measures. This notion of convergence is called *local weak convergence*.

In our setting of complete graphs $K_n = (V_n, E_n)$ with random i.i.d. edge costs $\{\xi_e, e \in E_n\}$, we regard the edge costs to be the lengths of the edges, and declare a vertex of K_n chosen uniformly at random as the root of K_n. This makes K_n along with its root a random element of $\mathcal{G}_*$. We rescale the edge costs such that for each n, $\{\xi_e, e \in E_n\}$ are i.i.d. random variables with mean n exponential distribution. We will denote this random, rooted, rescaled version of the n-vertex complete graph by $\overline{K}_n$ to distinguish it from the K_n defined earlier. Theorem 5 stated below (from [1]) says that the sequence of random geometric networks $\overline{K}_n$ converges in the local weak sense to an element of $\mathcal{G}_*$ called the *Poisson weighted infinite tree* (*PWIT*).

3.3. *Poisson weighted infinite tree.* We use the notation from [14] to define the PWIT.

Denote by $\mathcal{V}$ the set of all finite words over the alphabet $\mathbf{N} = \{1, 2, 3, \ldots\}$. Let ϕ denote the empty string and "." the concatenation operator. For any $v \in \mathcal{V}$ write $|v|$ for the length of string v, and if $v \neq \phi$ write $\dot{v}$ for the string obtained by removing the last letter of v.

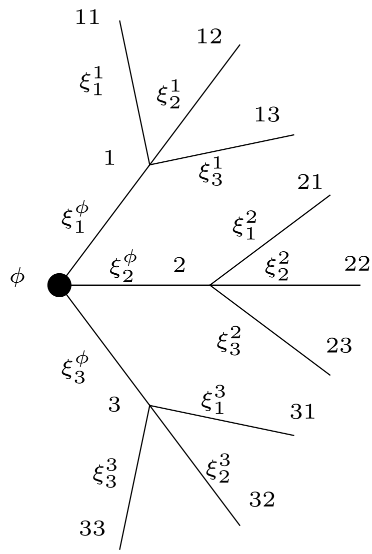

FIG. 2. *PWIT $\mathcal{T}$ up to depth* 2, *with only the first three children of each vertex shown.*

Construct an undirected graph $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ on $\mathcal{V}$ with the edge set

$$\mathcal{E} = \{\{v, v.i\}, v \in \mathcal{V}, i \in \mathbf{N}\}.$$

Set ϕ to be the root of $\mathcal{T}$. Then $\mathcal{T}$ is an infinite rooted tree with each vertex having a countably infinite number of children. Construct a family of independent Poisson processes of intensity 1 on $\mathbf{R}_+ : \{\boldsymbol{\xi}^v = (\xi_1^v, \xi_2^v, \ldots), v \in \mathcal{V}\}$. Assign to each edge $\{v, v.i\}$ in $\mathcal{E}$ the length ξ_i^v. $\mathcal{T}$ is then a random element of $\mathcal{G}_*$, and we call it the Poisson weighted infinite tree (PWIT) (Figure 2).

THEOREM 5 ([1]). *The sequence of uniformly rooted random networks $\overline{K}_n$ converges to the PWIT $\mathcal{T}$ as $n \to \infty$ in the sense of local weak convergence.*

A similar result was earlier established by Hajek [9], Section IV, for a class of sparse Erdős–Rényi random graphs. The above theorem says that if we look at an arbitrary, large but fixed neighborhood of the root of $\overline{K}_n$, then for large n it looks like the corresponding neighborhood of the root of $\mathcal{T}$. This suggests that if boundary conditions can be ignored, we may be able to relate optimal edge covers on $\overline{K}_n$ with an appropriate edge cover on $\mathcal{T}$ [to be precise, an optimal *involution invariant* edge cover (Section 5) on the PWIT]. Furthermore, the local neighborhood of the root of $\overline{K}_n$ is a tree for large enough n (with high probability). So we may expect belief propagation on $\overline{K}_n$ to converge. Both the above observations are true in the matching case; the former was established in [1, 4], and the latter was shown in [14]. We now extend these ideas to prove similar results for the edge-cover problem.

4. Recursive distributional equation.

4.1. *A heuristic recursion.* The PWIT $\mathcal{T}$ is an infinite graph, and it is clear that any edge cover on it must have infinite cost. So it does not make sense to talk about a minimum-cost edge cover on $\mathcal{T}$. However, for a moment let us pretend to perform operations on the minimum cost as if it were a finite quantity. Write $C(\mathcal{T})$ for this minimum cost, and define

$$D(\mathcal{T}) = \big(C(\mathcal{T}) - C\big(\mathcal{T} \setminus \{\phi\}\big)\big)^+, \tag{9}$$

where $C(\mathcal{T} \setminus \{\phi\})$ is the minimum cost of edge cover on the subgraph of $\mathcal{T}$ obtained by removing the root. Note that $D(\mathcal{T})$ denotes the difference between the minimum cost of edge cover of $\mathcal{T}$ and the minimum cost of partial edge cover of $\mathcal{T}$ where the root ϕ can be left uncovered.

If j is a child of the root, let $\mathcal{T}^j$ denote the induced subgraph of $\mathcal{T}$ containing j and all its descendants, and view it as a rooted network with root j (Figure 3). Define $D(\mathcal{T}^j)$ accordingly, and observe from the symmetry of $\mathcal{T}$ that $\{D(\mathcal{T}^j), j \geq 1\}$ are i.i.d., and have the same distribution as $D(\mathcal{T})$. We give a heuristic argument that $D(\mathcal{T})$ satisfies the following relation:

$$D(\mathcal{T}) = \min_{j \geq 1} \big(\xi_j^\phi - D\big(\mathcal{T}^j\big)\big)^+. \tag{10}$$

We can write $C(\mathcal{T} \setminus \{\phi\})$ in terms of edge covers on the subtrees $\mathcal{T}^j$, $j \geq 1$, as

$$C\big(\mathcal{T} \setminus \{\phi\}\big) = \sum_{j \in \mathbf{N}} C\big(\mathcal{T}^j\big). \tag{11}$$

Let us consider edge covers in which the edges covering the root are incident on the vertices in a fixed subset A of the children of the root. The minimum cost among such edge covers can be written as

$$\sum_{j \in A} \big(\xi_j^\phi + \min\big\{C\big(\mathcal{T}^j\big), C\big(\mathcal{T}^j \setminus \{j\}\big)\big\}\big) + \sum_{i \in \mathbf{N} \setminus A} C\big(\mathcal{T}^i\big).$$

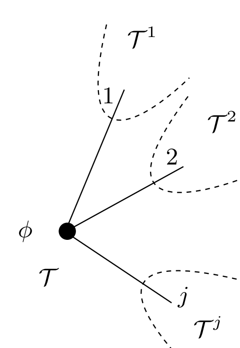

FIG. 3. *PWIT $\mathcal{T}$ with the subtrees $\mathcal{T}^j$ at node j.*

$C(\mathcal{T})$ is the minimum of the above value taken over all nonempty A, that is,

$$C(\mathcal{T}) = \min_{A \text{ nonempty}} \left\{ \sum_{j \in A} \left(\xi_j^\phi + \min\{C(\mathcal{T}^j), C(\mathcal{T}^j \setminus \{j\})\}\right) + \sum_{i \in \mathbf{N} \setminus A} C(\mathcal{T}^i) \right\}. \tag{12}$$

Thus we can write

$$\begin{aligned} D(\mathcal{T}) &= \left(\min_{|A| \geq 1} \sum_{j \in A} \left(\xi_j^\phi - \left(C(\mathcal{T}^j) - C(\mathcal{T}^j \setminus \{j\})\right)^+\right) \right)^+ \\ &= \left(\min_{|A| \geq 1} \sum_{j \in A} \left(\xi_j^\phi - D(\mathcal{T}^j)\right) \right)^+. \end{aligned}$$

To minimize the term within parentheses, we must include all those indices j for which the summand $(\xi_j^\phi - D(\mathcal{T}^j))$ is negative. If the terms are positive for all indices j, A must be the singleton where the minimum is attained among all indices. By then taking the positive part, equation (10) follows.

Although $D(\mathcal{T})$ and $D(\mathcal{T}^j)$ are not well-defined quantities, we shall prove that there is a nonnegative random variable X and i.i.d. random variables X_j, $j \geq 1$, having the same distribution as X, such that

$$X = \min_{j \geq 1} (\xi_j - X_j)^+, \tag{13}$$

where $\{\xi_j, j \geq 1\}$ are points of a Poisson process of rate 1 on $\mathbf{R}_+$, independent of $\{X_j, j \geq 1\}$.

4.2. *Recursive distributional equations and recursive tree processes*. Equations of the form of (13) are termed as recursive distributional equations in [5]. Specifically, if $\mathcal{P}(S)$ denotes the space of probability measures on a space S, a recursive distributional equation (RDE) is a fixed-point equation on $\mathcal{P}(S)$ of the form

$$X \stackrel{\mathrm{D}}{=} g\left(\xi; (X_j, 1 \leq j < N)\right), \tag{14}$$

where X_j, $j \geq 1$ are i.i.d. S-valued random variables having the same distribution as X, and are independent of the pair (ξ, N), ξ is a random variable on some space and N is a random variable on $\mathbf{N} \cup \{+\infty\}$. g is a given S-valued function. A solution to the RDE is a common distribution of X, X_j, $j \geq 1$, satisfying (14).

We can use relation (14) to construct a tree indexed stochastic process, say $X_i, i \in \mathcal{V}$, which is called a recursive tree process (RTP) [5]. Associate to each vertex $i \in \mathcal{V}$, an independent copy (ξ_i, N_i) of the pair (ξ, N), and require X_i to satisfy

$$X_i \stackrel{\mathrm{D}}{=} g\left(\xi_i; (X_{i.j}, 1 \leq j < N_i)\right)$$

with X_i independent of $\{(\xi_{i'}, N_{i'}) \mid |i'| < |i|\}$. If $\mu \in \mathcal{P}(S)$ is a solution to the RDE (14), there exists a stationary RTP; that is, each X_i is distributed as μ. Such a process is called an invariant RTP with marginal distribution μ.

4.3. *Solution to the edge cover RDE.*

THEOREM 6. *The unique solution to the RDE* (14) *is the c.d.f. F_* whose complementary c.d.f. $\overline{F}_*$ is given by*

$$\overline{F}_*(y) = \begin{cases} W(1)e^{-y}, & \text{if } y \geq 0, \\ 1, & \text{if } y < 0. \end{cases} \tag{15}$$

The function W above is Lambert's W-function, the inverse of $f : \mathbf{R}_+ \to \mathbf{R}_+$, $f(x) = xe^x$. In particular, $W(1)e^{W(1)} = 1$.

PROOF. Let μ be a solution to the RDE (13), and let F be its c.d.f. Take X_j, $j \geq 1$ i.i.d. with distribution μ. Then $\{(\xi_j, X_j), j \geq 1\}$ is a Poisson process on $\mathbf{R}_+ \times \mathbf{R}_+$ with intensity $\mathrm{d}z\,\mathrm{d}F(x)$. For $y \in \mathbf{R}_+$,

$$\begin{aligned} \mathrm{P}(X > y) &= \mathrm{P}\Big(\min_{j\geq 1}(\xi_j - X_j)^+ > y\Big) \\ &= \mathrm{P}\big(\text{No point of } \{(\xi_j, X_j)\} \text{ in } \{(z, x) : z - x \leq y\}\big) \\ &= \exp\left(-\int_{z=0}^{y}\int_{x=0}^{\infty} \mathrm{d}F(x)\,\mathrm{d}z - \int_{z=y}^{\infty}\int_{x=z-y}^{\infty} \mathrm{d}F(x)\,\mathrm{d}z\right) \\ &= e^{-y}\exp\left(-\int_{t=0}^{\infty}\int_{x=t}^{\infty} \mathrm{d}F(x)\,\mathrm{d}t\right) \\ &= e^{-y}\exp\left(-\int_0^{\infty}(1 - F(t))\,\mathrm{d}t\right). \end{aligned}$$

Writing $\overline{F}(t) = 1 - F(t)$, we have

$$\overline{F}(y) = e^{-y}\exp\left(-\int_0^{\infty}\overline{F}(t)\,\mathrm{d}t\right) \qquad \text{for all } y \geq 0.$$

Let $c = \exp(-\int_0^\infty \overline{F}(t)\,\mathrm{d}t)$. Then, using $\overline{F}(t) = ce^{-t}$ in the expression for c gives

$$c = \exp\left(-\int_0^{\infty} ce^{-t}\mathrm{d}t\right) = e^{-c}.$$

The unique c satisfying the above equation is $c = W(1)$. This proves that F must be the c.d.f. F_*. □

5. Unimodularity and involution invariance. In Section 3 we defined the space $\mathcal{G}_*$ as the set of connected rooted geometric networks. Now define $\mathcal{G}_{**}$ as the space of connected geometric networks with an ordered pair of distinguished vertices. Again, we do not distinguish between isomorphisms in $\mathcal{G}_{**}$, and denote by (G, o, x) the isomorphism class of elements with underlying network G and distinguished vertex pair (v, o). We endow this space with the topology of local convergence in the same way as $\mathcal{G}_*$, except that for the isomorphism between the local neighborhoods of two graphs, we require that the distinguished ordered vertex pair of one graph maps to the distinguished pair of the other graph. There is a suitable metric for this convergence that makes $\mathcal{G}_{**}$ a complete separable metric space.

A probability measure μ on $\mathcal{G}_*$ is called *unimodular* if it satisfies the following for all Borel $f : \mathcal{G}_{**} \to [0, \infty]$:

$$\int \sum_{x \in V(G)} f(G, o, x)\, \mathrm{d}\mu(G, o) = \int \sum_{x \in V(G)} f(G, x, o)\, \mathrm{d}\mu(G, o).$$

A measure μ on $\mathcal{G}_*$ that satisfies the above for all Borel f supported on $\{(G, x, y) | x \sim y\}$ is said to be *involution invariant*. It is clear that the set of unimodular measures is a subset of the set of involution invariant measures. Proposition 2.2 of [2] shows that involution invariance is equivalent to unimodularity.

Involution invariance is characterized alternatively in [3] as follows. Given a measure μ on $\mathcal{G}_*$, define a measure μ^* on $\mathcal{G}_{**}$ by letting its marginal measure on $\mathcal{G}_*$ to be μ and the conditional measure on the second vertex given a rooted geometric network G to be the counting measure on the neighbors of the root of G. Specifically,

$$\mu^*(\cdot) = \int_{\mathcal{G}_*} \sum_{v \sim o} \mathbf{1}_{\{(G,o,v) \in \cdot\}}\, \mathrm{d}\mu(G, o).$$

Then μ is involution invariant if μ^* is invariant under the *involution* transformation

$$\iota : \mathcal{G}_{**} \to \mathcal{G}_{**}, \iota(G, o, v) = (G, v, o).$$

Involution ι swaps the order of the distinguished pair of vertices, leaving all else unchanged.

The definitions carry forward when the graphs in $\mathcal{G}_*$ are appended with maps from their edge sets to a complete separable metric space. An edge cover C on a graph G can be represented as the graph G with a map on the edge set of $G : e \mapsto \mathbf{1}_{\{e \in C\}}$. We say that a random edge cover C on a random graph G is involution invariant if the distribution of G with the above map on its edges is involution invariant.

In our model, the complete graphs $\overline{K}_n$ are randomly rooted. Write C_n^* for the minimum-cost edge over on $\overline{K}_n$ having the same root as $\overline{K}_n$. By symmetry it is easy to see that its distribution is involution invariant. From Section 5.2 of [3], we see that involution invariance is preserved under weak limits in the metric space $\mathcal{G}_*$ appended with the $\{0, 1\}$-map on the edge set. Consequently, if the sequence $C_n^*, n \geq 1$, converges to an element C^*, then the distribution of C^* will be involution invariant. This motivates us to study involution invariant edge covers on the limit PWIT.

6. Optimal involution invariant edge cover on the PWIT.

6.1. *A tree process based on the RDE.* In the PWIT we split each undirected edge into two directed edges. For a general graph G, we use the notation $\overrightarrow{E}(G)$ to denote the set of directed edges so obtained. If ξ_e is the cost of the undirected edge

$e = \{v, w\}$, we assign the same cost to both of the corresponding directed edges and write the costs as $\xi(u, v) = \xi(v, u) = \xi_e$. To each directed edge $\overrightarrow{e} = (u, v)$, we will assign a random variable denoted by $X(\overrightarrow{e})$ or $X(u, v)$. Typically, $X(u, v)$ will be different from $X(v, u)$. The X process is constructed in the following lemma, which is an analogue of Lemma 5.8 of [3] and is proved similarly. We include the proof here for completeness.

LEMMA 1. *There exists a process*

$$\big(\mathcal{T}, (\xi_e, e \in E(\mathcal{T})), (X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}(\mathcal{T}))\big),$$

where $\mathcal{T}$ *is a PWIT with edge lengths* $\{\xi_e, e \in E(\mathcal{T})\}$, *and* $\{X(\overrightarrow{e}), \overrightarrow{e} \in \overrightarrow{E}(\mathcal{T})\}$ *is a stochastic process satisfying the following properties*:

(a) *For each directed edge* $(u, v) \in \overrightarrow{E}(\mathcal{T})$,

$$X(u, v) = \min\{(\xi(v, w) - X(v, w))^+ : (v, w) \in \overrightarrow{E}(\mathcal{T}), w \neq u\}. \tag{16}$$

(b) *If* $(u, v) \in \overrightarrow{E}(\mathcal{T})$ *is directed away from the root of* $\mathcal{T}$, *then* $X(u, v)$ *has the distribution* F_* *as in* (15).

(c) *If* $(u, v) \in \overrightarrow{E}(\mathcal{T})$, *the random variables* $X(u, v)$ *and* $X(v, u)$ *are independent*.

(d) *For a fixed* $z > 0$, *conditional on the event that there exists an edge of length* z *at the root, say* $\{\phi, v_z\}$, *the random variables* $X(\phi, v_z)$ *and* $X(v_z, \phi)$ *are independent random variables, each having the distribution* F_*.

PROOF. Fix an integer $d \geq 1$. We create independent random variables from the distribution F_*, and assign one to each directed edge (v, w) of $\mathcal{T}$ where v is at depth $d - 1$, and w is at depth d from the root. Then if $d > 1$, use relation (16) to recursively define random variables $X(t, u)$, where $t \sim u$ are vertices of $\mathcal{T}$ within depth d from the root. This generates a collection of random variables $\mathscr{C}_d$ whose joint distribution satisfies properties (a), (b) and (c) in the statement of the lemma for all vertices of $\mathcal{T}$ up to a depth d from the root. It is easy to see that the sequence of collections $\{\mathscr{C}_d, d \geq 1\}$ satisfies the conditions of Kolmogorov consistency theorem. So there exists a collection $\mathscr{C}_\infty$ such that the restriction to random variables corresponding to vertices up to depth d is equal in distribution to the collection $\mathscr{C}_d$ for each $d \geq 1$. This implies that random variables in $\mathscr{C}_\infty$ satisfy the properties (a), (b) and (c).

To prove property (d), observe that a Poisson process conditioned to have a point at z is also a Poisson process of the same intensity when that point is removed. Now conditional on the existence of the edge $\{\phi, v_z\}$ of length z, if we remove this edge the PWIT splits into two subtrees. Letting ϕ and v_z to be the roots of these two subtrees, we find that the two subtrees are independent copies of the original PWIT $\mathcal{T}$. From the construction in the previous paragraph, it is clear that conditionally the random variables $X(\phi, v_z)$ and $X(v_z, \phi)$ are independent, and have the same distribution F_*. $\square$

6.2. *An involution invariant edge cover on the PWIT.* We use the process $\{X(\overrightarrow{e})\}$ to construct an edge cover $\mathcal{C}_{\text{opt}}$ on $\mathcal{T}$.

For each vertex v of the PWIT, define a set

$$\mathcal{C}_{\text{opt}}(v) = \arg\min_{y \sim v}\{(\xi(v, y) - X(v, y))^+\}. \tag{17}$$

In words, include in $\mathcal{C}_{\text{opt}}(v)$ all $y \sim v$ such that $\xi(v, y) - X(v, y) < 0$, and if there is no such y, then $\mathcal{C}_{\text{opt}}(v) = \{w\}$ where w is the unique (with probability 1) neighbor of v that minimizes $\xi(v, \cdot) - X(v, \cdot)$. Alternatively,

$$\mathcal{C}_{\text{opt}}(v) = \arg\min_{A}\left\{\sum_{y \in A}(\xi(v, y) - X(v, y)) : A \subset N_v, A \text{ nonempty}\right\}. \tag{18}$$

Define the edge cover to be

$$\mathcal{C}_{\text{opt}} = \bigcup_{v}\{\{v, w\} : w \in \mathcal{C}_{\text{opt}}(v)\}.$$

The following lemma reassures us that the chosen edge cover does not include wasteful edges.

LEMMA 2. *For any two vertices v, w of $\mathcal{T}$, we have*

$$v \in \mathcal{C}_{\text{opt}}(w) \quad \Longleftrightarrow \quad \xi(v, w) < X(v, w) + X(w, v).$$

As a consequence,

$$v \in \mathcal{C}_{\text{opt}}(w) \quad \Longleftrightarrow \quad w \in \mathcal{C}_{\text{opt}}(v).$$

PROOF. Suppose $w \in \mathcal{C}_{\text{opt}}(v)$. If $\xi(v, w) < X(v, w)$ then, since $X(w, v) \geq 0$, we have $\xi(v, w) < X(v, w) + X(w, v)$.

If $\xi(v, w) \geq X(v, w)$, then definition (17) of $\mathcal{C}_{\text{opt}}(v)$ and w's membership to this set implies that w is the only element of

$$\arg\min_{y \sim v}\{(\xi(v, y) - X(v, y))^+\},$$

that is,

$$\xi(v, w) - X(v, w) < (\xi(v, y) - X(v, y))^+ \qquad \text{for all } y \sim v, y \neq w.$$

Hence,

$$\begin{aligned}\xi(v, w) - X(v, w) &< \min\{(\xi(v, y) - X(v, y))^+ : y \sim v, y \neq w\} \\ &= X(w, v),\end{aligned}$$

where the last equality follows from (16). We have thus established one direction of the first statement, that is,

$$w \in \mathcal{C}_{\text{opt}}(v) \quad \Longrightarrow \quad \xi(v, w) < X(v, w) + X(w, v).$$

Conversely, suppose that $\xi(v, w) < X(v, w) + X(w, v)$. Then $X(w, v) > \xi(v, w) - X(v, w)$. Also $X(w, v) \geq 0$. Therefore,

$$X(w, v) \geq \big(\xi(v, w) - X(v, w)\big)^+,$$

that is,

$$\min_{y \sim v, y \neq w} \big(\xi(v, y) - X(v, y)\big)^+ \geq \big(\xi(v, w) - X(v, w)\big)^+.$$

It follows that

$$w \in \arg\min_{y \sim v} \big(\xi(v, y) - X(v, y)\big)^+$$

and hence $w \in \mathcal{C}_{\text{opt}}(v)$. Thus we have established the first statement of the lemma, which is

$$w \in \mathcal{C}_{\text{opt}}(v) \quad \iff \quad \xi(v, w) < X(v, w) + X(w, v).$$

The condition on the right-hand side above is symmetric in v, w, and hence the second statement of the lemma is proved. □

The following lemma asserts that the edge cover $\mathcal{C}_{\text{opt}}$ satisfies involution invariance. See Section 5 for definition. The proof is similar to the proof of Lemma 24 of [4].

LEMMA 3. *$\mathcal{C}_{\text{opt}}$ is involution invariant.*

PROOF. Given $\xi_e, X(\vec{e}), \vec{e} \in \vec{E}(\mathcal{T})$, the edge cover $\mathcal{C}_{\text{opt}}$ does not depend on the vertex labels (which are strings from $\mathcal{V}$). Relation (16) for the X process is also independent of the labels of the vertices. The proof of the lemma is then complete by showing that the measure of the X process constructed in Lemma 1 is involution invariant.

From the proof of Lemma 1 it is clear that the joint distribution of X process is determined by the property that for any $d > 1$,

$$\{X(v, w) | v \text{ at depth } d-1 \text{ from the root}, w \text{ at depth } d \text{ from the root}\}$$

are independent random variables with distribution F_*. We need to show that this property is invariant under the involution map.

If ϕ is the root (first distinguished vertex) of $\mathcal{T}$, and $u \sim \phi$ is the second distinguished vertex, then under the involution map, u becomes the root and ϕ the second distinguished vertex. Write $\mathcal{T}_u$ for the subtree containing u obtained by removing the edge $\{\phi, u\}$. For an arbitrary Borel set B, define the event

$$A := \{(X(v, w), v \text{ at depth } d-1 \text{ from } u, w \text{ at depth } d \text{ from } u) \in B\}.$$

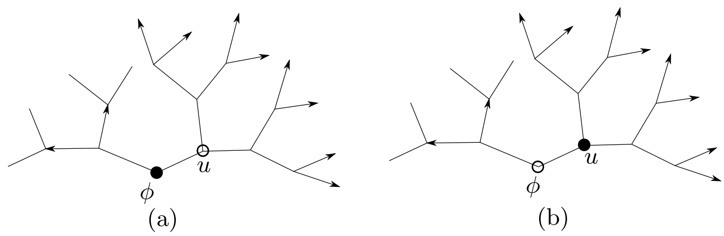

FIG. 4. *The edges involved in events* A (a) *and* $\iota^{-1}(A)$ (b) *are shown with arrow heads. Here* $d = 3$. *The vertex with a filled circle is the root, and the vertex with an unfilled circle is the second distinguished vertex.*

The inverse image of A in the involution map is

$$\begin{aligned}\iota^{-1}(A) = \big\{\big(X(v_1, w_1), v_1 \in \mathcal{T}_u, v_1 \text{ at depth } d \text{ from } \phi,\\ w_1 \text{ at depth } d+1 \text{ from } \phi;\\ X(v_2, w_2), v_2 \in \mathcal{T} \setminus \mathcal{T}_u, v_2 \text{ at depth } d-2 \text{ from } \phi,\\ w_2 \text{ at depth } d-1 \text{ from } \phi\big) \in B\big\}.\end{aligned}$$

Figure 4 shows the edges involved. It is clear that the random variables considered above are independent with distribution F_*. Consequently the measure of the set $\iota^{-1}(A)$ equals the measure of A. This completes the proof. Note that we have used here the simpler notion of involution invariance described in Section 5 rather than *spatial invariance* as used in [4]. □

6.3. *Evaluating the cost*. In the following theorem we evaluate the cost of the edge cover $\mathcal{C}_{\text{opt}}$ on the $\mathcal{T}$. For obvious reasons, the expectation is twice the right-hand side of (6).

THEOREM 7.

$$\mathrm{E}\Big[\sum_{v \in \mathcal{C}_{\text{opt}}(\phi)} \xi(\phi, v)\Big] = 2W(1) + W(1)^2.$$

PROOF. Denote by D the event that $\xi(\phi, v) > X(\phi, v)$ for all $v \sim \phi$. Under the event D, there is only one vertex in $\mathcal{C}_{\text{opt}}(\phi)$, say y. By Lemma 2, y is the only neighbor of ϕ satisfying $\xi(\phi, y) < X(\phi, y) + X(y, \phi)$. Also, from (16), $X(y, \phi) > 0$. Conversely, if there is a neighbor y of ϕ that satisfies (i) $X(y, \phi) > 0$, (ii) $\xi(\phi, y) > X(\phi, y)$ and (iii) $\xi(\phi, y) < X(\phi, y) + X(y, \phi)$, then from (16), we have

$$0 < X(y, \phi) = \min\{(\xi(\phi, v) - X(\phi, v))^+, v \sim \phi, v \neq y\},$$

which implies $\xi(\phi, v) > X(\phi, v)$ for every $v \sim \phi, v \neq y$. This and (ii) together imply that the event D holds, and $\mathcal{C}_{\text{opt}}(\phi) = \{y\}$.

Now fix a $z > 0$, and condition on the event that there is a neighbor v_z of ϕ with $\xi(\phi, v_z) = z$. Call this event E_z. If we condition a Poisson process to have a point at some location, then the conditional process on removing this point is again a Poisson process with the same intensity. This shows that under E_z, $X(\phi, v_z)$ and $X(v_z, \phi)$ both have the same distribution F_*. Also they are independent. Using these facts and the characterization of the event D in the previous paragraph, the expected cost under D can be written as

$$
\begin{aligned}
&\mathrm{E}\Big[\Big(\sum_{v\in\mathcal{C}_{\text{opt}}(\phi)} \xi_{\mathcal{T}}(\phi, v)\Big)\mathbf{1}_D\Big] \\
&\quad = \int_{z=0}^{\infty} zP\{X(v_z, \phi) > 0, z > X(\phi, v_z), z < X(\phi, v_z) + X(v_z, \phi)\}\,\mathrm{d}z \\
&\quad = \int_{z=0}^{\infty}\Big(zP\{X(\phi, v_z) = 0\}\mathrm{P}\{X(v_z, \phi) > z\} \\
&\qquad\qquad + \int_{x=0}^{z} zP\{X(v_z, \phi) > z - x\}\,\mathrm{d}F_*(x)\Big)\,\mathrm{d}z \qquad (19)\\
&\quad = \int_{z=0}^{\infty}\Big(z\big(1 - W(1)\big)W(1)e^{-z} + \int_{x=0}^{z} zW(1)e^{-(z-x)}W(1)e^{-x}\,\mathrm{d}x\Big)\,\mathrm{d}z \\
&\quad = W(1)\big(1 - W(1)\big) + 2W(1)^2 \\
&\quad = W(1) + W(1)^2.
\end{aligned}
$$

In the second equality above, we condition on $X(\phi, v_z) = 0$ and $X(\phi, v_z) = x \in (0, z)$, respectively, in the two terms of the integrand.

Under the event D^c, $\mathcal{C}_{\text{opt}}(\phi)$ contains all v for which $\xi(\phi, v) < X(\phi, v)$. The expected cost over this event is given by

$$
\begin{aligned}
&\mathrm{E}\Big[\Big(\sum_{v\in\mathcal{C}_{\text{opt}}(\phi)} \xi(\phi, v)\Big)\mathbf{1}_{D^c}\Big] \\
&\quad = \mathrm{E}\Big[\sum_{v} \xi(\phi, v)\mathbf{1}_{\{\xi(\phi,v)<X(\phi,v)\}}\Big] \\
&\quad = \sum_{v} \mathrm{E}\big[\xi(\phi, v)\mathbf{1}_{\{\xi(\phi,v)<X(\phi,v)\}}\big] \\
&\quad = \sum_{v} \int_{y=0}^{\infty} \mathrm{P}\{\xi(\phi, v) > y, \xi(\phi, v) < X(\phi, v)\}\,\mathrm{d}y \\
&\quad = \int_{y=0}^{\infty} \sum_{v} P\{y < \xi(\phi, v) < X\}\,\mathrm{d}y \\
&\qquad (X \text{ is a } F_*\text{-distributed r.v. independent of the Poisson process}) \qquad (20)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{y=0}^{\infty} \mathrm{E}\big[\text{Number of Poisson points in } [y, X]\big] \, \mathrm{d}y \\
&= \int_{y=0}^{\infty} \mathrm{E}\big[(X - y)^{+}\big] \, \mathrm{d}y \\
&= \int_{y=0}^{\infty} \int_{x=y}^{\infty} \overline{F_*}(x) \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{y=0}^{\infty} \int_{x=y}^{\infty} W(1) e^{-x} \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{y=0}^{\infty} W(1) e^{-y} \, \mathrm{d}y \\
&= W(1).
\end{aligned}
$$

Combining (19) and (20) completes the proof. □

In passing, we remark that $\mathcal{C}_{\mathrm{opt}}(\phi)$ is finite almost surely.

6.4. *Optimality in the class of involution invariant edge covers.* We now show that our candidate edge cover $\mathcal{C}_{\mathrm{opt}}$ has the minimum expected cost among involution invariant edge covers on the PWIT.

THEOREM 8. *Let $\mathcal{C}$ be an involution invariant edge cover of the PWIT $\mathcal{T}$. Write $\mathcal{C}(\phi)$ for the set of vertices of $\mathcal{T}$ adjacent to the root ϕ in $\mathcal{C}$. Then*

$$
\mathrm{E}\Big[\sum_{v \in \mathcal{C}(\phi)} \xi(\phi, v)\Big] \geq \mathrm{E}\Big[\sum_{v \in \mathcal{C}_{\mathrm{opt}}(\phi)} \xi(\phi, v)\Big].
$$

Let us first set up some notation that will simplify the proof steps. For each directed edge (v, w) of $\mathcal{T}$, define a random variable

$$
Y(v, w) = \min\left\{ \sum_{y \in A} \big(\xi(w, y) - X(w, y)\big) \,\middle|\, \begin{array}{c} A \subset N_w \setminus \{v\}, \\ A \text{ nonempty} \end{array} \right\}, \tag{21}
$$

where N_w is the set of neighbors of w. It is easy to see that the random variable can be written as

$$
Y(v, w) = \begin{cases} \displaystyle\min_{y \sim w, y \neq v} \{\xi(w, y) - X(w, y)\} c, \\ \qquad \text{if } \xi(w, y) - X(w, y) \geq 0 \qquad \text{for all } y \sim w, y \neq v, \\ \displaystyle\sum_{y \sim w, y \neq v} \big(\xi(w, y) - X(w, y)\big) \mathbf{1}_{\{\xi(w,y) - X(w,y) < 0\}}, \\ \qquad \text{otherwise.} \end{cases}
$$

Note that $(Y(v, w))^{+} = X(v, w)$.

Suppose that $\mathrm{E}[\sum_{v \in \mathcal{C}(\phi)} \xi(\phi, v)] < \infty$. Then $\mathcal{C}(\phi)$ is a finite set with probability 1 because $\{\xi(\phi, v), v \sim \phi\}$ are points of a Poisson process of rate 1. For such an edge cover $\mathcal{C}$, define

$$A(\mathcal{C}) = \sum_{v \in \mathcal{C}(\phi)} X(\phi, v) + \max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v, \phi). \tag{22}$$

The max operation in the above equation is over an infinite number of vertices; however, in the remark after the proof of Lemma 4, we will show that effectively $Y(v, \phi)$ assumes only finitely many values as we vary v, and hence the max operation as well as $A(\mathcal{C})$ are almost surely well defined.

The following two lemmas will be used to prove Theorem 8.

LEMMA 4. *Let $\mathcal{C}$ be an edge cover rule on the PWIT such that*

$$\mathrm{E}\left[\sum_{v \in \mathcal{C}(\phi)} \xi(\phi, v)\right] < \infty.$$

Then almost surely,

$$\sum_{v \in \mathcal{C}(\phi)} \xi(\phi, v) \geq A(\mathcal{C}).$$

Furthermore,

$$\sum_{v \in \mathcal{C}_{\mathrm{opt}}(\phi)} \xi(\phi, v) = A(\mathcal{C}_{\mathrm{opt}}).$$

LEMMA 5. *Let $\mathcal{C}$ be an edge cover rule on the PWIT such that*

$$\mathrm{E}\left[\sum_{v \in \mathcal{C}(\phi)} \xi(\phi, v)\right] < \infty.$$

If $\mathcal{C}$ is involution invariant, *we have* $\mathrm{E}[A(\mathcal{C})] \geq \mathrm{E}[A(\mathcal{C}_{\mathrm{opt}})]$.

PROOF OF THEOREM 8. If $\mathrm{E}[\sum_{v \in \mathcal{C}(\phi)} \xi(\phi, v)] = \infty$, the statement of the theorem is trivially true. Assume that it is finite. We are now in a position to apply Lemmas 4 and 5 as follows to get the result

$$\begin{aligned}
\mathrm{E}\left[\sum_{v \in \mathcal{C}(\phi)} \xi(\phi, v)\right] &\geq \mathrm{E}[A(\mathcal{C})] \qquad \text{(Lemma 4)} \\
&\geq \mathrm{E}[A(\mathcal{C}_{\mathrm{opt}})] \qquad \text{(Lemma 5)} \\
&= \mathrm{E}\left[\sum_{v \in \mathcal{C}_{\mathrm{opt}}(\phi)} \xi(\phi, v)\right] \qquad \text{(Lemma 4)}.
\end{aligned}$$

□

Let us now complete the proofs of Lemmas 4 and 5.

PROOF OF LEMMA 4. From (21), we have

$$Y(v,\phi) \le \sum_{y\in A} \bigl(\xi(\phi,y) - X(\phi,y)\bigr)$$

for all $A \subset N_\phi \setminus \{v\}$, A nonempty.

For any $v \notin \mathcal{C}(\phi)$, we can choose $A = \mathcal{C}(\phi)$ to obtain

$$Y(v,\phi) \le \sum_{y\in\mathcal{C}(\phi)} \bigl(\xi(\phi,y) - X(\phi,y)\bigr).$$

This implies

$$\max_{v\notin\mathcal{C}(\phi), v\sim\phi} Y(v,\phi) \le \sum_{y\in\mathcal{C}(\phi)} \bigl(\xi(\phi,y) - X(\phi,y)\bigr). \tag{23}$$

Thanks to the finite expectation assumption in the lemma, $\mathcal{C}(\phi)$ is a finite set almost surely, and so $\sum_{y\in\mathcal{C}(\phi)} X(\phi,y)$ is finite. Rearrangement of (23) then yields

$$\sum_{v\in\mathcal{C}(\phi)} \xi(\phi,v) \ge A(\mathcal{C}).$$

Now recall the alternate characterization of $\mathcal{C}_{\text{opt}}$ via

$$\mathcal{C}_{\text{opt}}(w) = \arg\min_A \biggl\{\sum_{y\in A} \bigl(\xi(w,y) - X(w,y)\bigr) : A \subset N_w, A \text{ nonempty}\biggr\}. \tag{24}$$

From (21) and (24), for any $v \notin \mathcal{C}_{\text{opt}}(\phi)$, we have

$$Y(v,\phi) = \sum_{y\in\mathcal{C}_{\text{opt}}(\phi)} \bigl(\xi(\phi,y) - X(\phi,y)\bigr) \tag{25}$$

and hence

$$\max_{v\notin\mathcal{C}_{\text{opt}}(\phi), v\sim\phi} Y(v,\phi) = \sum_{y\in\mathcal{C}_{\text{opt}}(\phi)} \bigl(\xi(\phi,y) - X(\phi,y)\bigr).$$

It follows by rearrangement that

$$\sum_{v\in\mathcal{C}_{\text{opt}}(\phi)} \xi(\phi,v) = A(\mathcal{C}_{\text{opt}}).$$

□

Let us quickly reassure the reader that the max operation in (22) is well defined. Notice that (25) implies that $Y(w,\phi)$ takes values in the finite set

$$\biggl\{\sum_{y\in\mathcal{C}_{\text{opt}}(\phi)} \bigl(\xi(\phi,y) - X(\phi,y)\bigr)\biggr\} \cup \bigl\{Y(v,\phi) | v \in \mathcal{C}_{\text{opt}}(\phi)\bigr\}.$$

That $\mathcal{C}_{\text{opt}}(\phi)$ is finite (almost surely) can be gleaned from Theorem 7. This validates the assertion that the max in the definition of $A(\mathcal{C})$ is well defined.

PROOF OF LEMMA 5. Define

$$\widetilde{A}(\mathcal{C}) = \sum_{v \in \mathcal{C}(\phi)} X(v,\phi) + \max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v,\phi). \tag{26}$$

We will prove Lemma 5 by showing the following two results:

(a) For an involution invariant edge cover $\mathcal{C}$,

$$\mathrm{E}[\widetilde{A}(\mathcal{C})] = \mathrm{E}[A(\mathcal{C})]. \tag{27}$$

(b) Almost surely,

$$\widetilde{A}(\mathcal{C}) \geq \widetilde{A}(\mathcal{C}_{\mathrm{opt}}). \tag{28}$$

We first prove (27). First, by involution invariance of $\mathcal{C}$, we have

$$\mathrm{E}\Bigg[\sum_{v \in \mathcal{C}(\phi)} X(\phi, v)\Bigg] = \mathrm{E}\Bigg[\sum_{v \in \mathcal{C}(\phi)} X(v,\phi)\Bigg]. \tag{29}$$

Indeed, the left-hand side equals

$$\int_{\mathcal{G}_*} \sum_{v \sim \phi} X(\phi, v) \mathbf{1}_{\{\{\phi, v\} \in \mathcal{C}\}} \,\mathrm{d}\mu_{\mathcal{C}}\big([G,\phi]\big),$$

where $\mu_{\mathcal{C}}$ is the probability measure on $\mathcal{G}_*$ corresponding to $(\mathcal{T}, \mathcal{C})$. By involution invariance, this equals

$$\int_{\mathcal{G}_*} \sum_{v \sim \phi} X(v, \phi) \mathbf{1}_{\{\{v, \phi\} \in \mathcal{C}\}} \,\mathrm{d}\mu_{\mathcal{C}}\big([G,\phi]\big),$$

which is equal to the right-hand side of (29). Thanks to the finite expectation assumption of the lemma, we saw in the proof of Lemma 4 that

$$\max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v,\phi)$$

is finite almost surely. Now observe that $A(\mathcal{C})$ [resp., $\widetilde{A}(\mathcal{C})$] is obtained by adding the almost surely finite random variable $\max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v,\phi)$ to the random variable which is the argument of the expectation on the left-hand side of (29) [resp., the right-hand side of (29)]. Taking expectation and using the equality in (29), we get (27).

Now we will prove (28). First condition on the event $L_1 = \{|\mathcal{C}_{\mathrm{opt}}(\phi)| > 1\}$. Observe that, under L_1, $\xi(\phi, y) - X(\phi, y) < 0$, $y \sim \phi$ if and only if $y \in \mathcal{C}_{\mathrm{opt}}(\phi)$, and there are at least two such y. Then, by (16),

$$X(v,\phi) = 0 \qquad \text{for all } v \sim \phi. \tag{30}$$

Also, from (21) and (24),

$$Y(v,\phi) \geq \sum_{y \in \mathcal{C}_{\mathrm{opt}}(\phi)} \big(\xi(\phi, y) - X(\phi, y)\big) = Y(w, \phi)$$

if $w \notin \mathcal{C}_{\mathrm{opt}}(\phi)$. This implies

$$Y(v,\phi) \geq \max_{w \notin \mathcal{C}_{\mathrm{opt}}(\phi), w \sim \phi} Y(w,\phi) \qquad \text{for all } v \sim \phi.$$

In particular,

$$\max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v,\phi) \geq \max_{w \notin \mathcal{C}_{\mathrm{opt}}(\phi), w \sim \phi} Y(w,\phi). \tag{31}$$

Combining (30) and (31) gives

$$\begin{aligned} &\sum_{v \in \mathcal{C}(\phi)} X(v,\phi) + \max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v,\phi) \\ &\quad \geq \sum_{v \in \mathcal{C}_{\mathrm{opt}}(\phi)} X(v,\phi) + \max_{v \notin \mathcal{C}_{\mathrm{opt}}(\phi), v \sim \phi} Y(v,\phi). \end{aligned} \tag{32}$$

Thus $\widetilde{A}(\mathcal{C}) \geq \widetilde{A}(\mathcal{C}_{\mathrm{opt}})$ under L_1.

Now consider the event $L_2 = \{|\mathcal{C}_{\mathrm{opt}}(\phi)| = 1\}$. Let

$$X_\phi^{(1)} = \min_{v \sim \phi}\big(\xi(\phi,v) - X(\phi,v)\big) \quad \text{and}$$

$$X_\phi^{(2)} = \min_{v \sim \phi}{}^{(2)}\big(\xi(\phi,v) - X(\phi,v)\big),$$

where $\min^{(2)}$ stands for the second minimum.

Let $\mathcal{C}_{\mathrm{opt}}(\phi) = \{u\}$. Then $X(u,\phi) = X_\phi^{(2)}$, and for $v \in \mathcal{C}(\phi) \setminus \mathcal{C}_{\mathrm{opt}}(\phi)$, $X(v,\phi) = (X_\phi^{(1)})^+$. So we get

$$\begin{aligned} &\sum_{v \in \mathcal{C}(\phi)} X(v,\phi) - \sum_{v \in \mathcal{C}_{\mathrm{opt}}(\phi)} X(v,\phi) \\ &\quad = \sum_{v \in \mathcal{C}(\phi) \setminus \mathcal{C}_{\mathrm{opt}}(\phi)} (X_\phi^{(1)})^+ - X_\phi^{(2)} \mathbf{1}_{\{u \notin \mathcal{C}(\phi)\}}. \end{aligned} \tag{33}$$

If $v \notin \mathcal{C}_{\mathrm{opt}}(\phi)$, then $Y(v,\phi) = X_\phi^{(1)}$. Also $Y(u,\phi) = X_\phi^{(2)}$. Since $X_\phi^{(2)} \geq X_\phi^{(1)}$, we get

$$\max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v,\phi) = X_\phi^{(2)} \mathbf{1}_{\{u \notin \mathcal{C}(\phi)\}} + X_\phi^{(1)} \mathbf{1}_{\{u \in \mathcal{C}(\phi)\}}$$

and

$$\max_{v \notin \mathcal{C}_{\mathrm{opt}}(\phi), v \sim \phi} Y(v,\phi) = X_\phi^{(1)}.$$

Therefore,

$$\max_{v \notin \mathcal{C}(\phi), v \sim \phi} Y(v,\phi) - \max_{v \notin \mathcal{C}_{\mathrm{opt}}(\phi), v \sim \phi} Y(v,\phi) = \big(X_\phi^{(2)} - X_\phi^{(1)}\big)\mathbf{1}_{\{u \notin \mathcal{C}(\phi)\}}. \tag{34}$$

Adding (33) and (34), and canceling $X_\phi^{(2)}\mathbf{1}_{\{u\notin\mathcal{C}(\phi)\}}$, we get

$$\sum_{v\in\mathcal{C}(\phi)} X(v,\phi) + \max_{v\notin\mathcal{C}(\phi), v\sim\phi} Y(v,\phi) - \sum_{v\in\mathcal{C}_{\text{opt}}(\phi)} X(v,\phi) - \max_{v\notin\mathcal{C}_{\text{opt}}(\phi), v\sim\phi} Y(v,\phi)$$
$$= \sum_{v\in\mathcal{C}(\phi)\setminus\mathcal{C}_{\text{opt}}(\phi)} (X_\phi^{(1)})^+ - X_\phi^{(1)}\mathbf{1}_{\{u\notin\mathcal{C}(\phi)\}} \ge 0,$$

where the last inequality follows because there exists a $v \in \mathcal{C}(\phi)\setminus\mathcal{C}_{\text{opt}}(\phi)$ by virtue of our assumption that $\mathcal{C}(\phi) \neq \mathcal{C}_{\text{opt}}(\phi)$. Thus $\widetilde{A}(\mathcal{C}) \ge \widetilde{A}(\mathcal{C}_{\text{opt}})$ under L_2 as well. □

7. Completing the lower bound. In the previous section we described an edge cover $\mathcal{C}_{\text{opt}}$ on the infinite tree $\mathcal{T}$. We showed that this edge cover satisfies the expected property of involution invariance, and it has the minimum expected cost among all edge covers having this property. We use this to show now that the expected cost of $\mathcal{C}_{\text{opt}}$ serves as an asymptotic lower bound on the expected cost of min-cost edge covers on $\overline{K}_n$.

THEOREM 9. *Let C_n^* be the optimal edge cover on $\overline{K}_n$. Then*

$$\liminf_{n\to\infty} \mathrm{E}\Bigg[\sum_{\{\phi,v\}\in C_n^*} \xi_{\overline{K}_n}(\phi,v)\Bigg] \ge 2W(1) + W(1)^2.$$

PROOF. Take a subsequence $\{n_k, k\ge 1\}$ for which the lim inf above is a limit. Now consider the joint sequence $(C_{n_k}^*, \overline{K}_{n_k})_{k\ge1}$ in $\mathcal{G}_* \times \mathcal{G}_*$. Because $\overline{K}_{n_k} \xrightarrow{\text{l.w.}} \mathcal{T}$, for every $\varepsilon > 0$ there is a compact subset $\mathcal{K}$ of $\mathcal{G}_*$, with $\mathrm{P}\{\overline{K}_{n_k} \in \mathcal{K}\} > 1-\varepsilon$ for all k. Also, we can take the graphs $\overline{K}_{n_k}$ to be on a common vertex set $\widetilde{\mathcal{V}}$, and assume that all graphs in $\mathcal{K}$ are defined on the same vertex set. Let $\widetilde{\mathcal{E}}$ denote the set of all possible edges. Let $\mathcal{K}_S$ denote the set $\{H \text{ is a subgraph of } G | G \in \mathcal{K}\}$. Since $C_{n_k}^*$ is a subgraph of $\overline{K}_{n_k}$, $\mathrm{P}\{C_{n_k}^* \in \mathcal{K}_S\} > 1-\varepsilon$ for all k. An element of $\mathcal{K}_S$ can be identified with an element of $\mathcal{K}\times\{0,1\}^{\widetilde{\mathcal{E}}}$, where 1 or 0 denotes the presence or absence of an edge, respectively. Since the latter is a compact set, so is $\mathcal{K}_S$. This shows that the sequence of random graphs $\{C_{n_k}^*\}_{k\ge1}$ is tight. By completeness of $\mathcal{G}_*$, we have that $\{(C_{n_k}^*, \overline{K}_{n_k}), k\ge 1\}$ is sequentially compact. Therefore, there exists a further subsequence $\{n_j, j\ge1\}$ of $\{n_k, k\ge1\}$ such that $(C_{n_j}^*, \overline{K}_{n_j})$ converges in the local weak sense to $(C^*, \mathcal{T})$. Since the C_n^* distribution is involution invariant, so is the distribution of C^*. By Skorohod's theorem we can assume the convergence occurs almost surely in some probability space. By the definition of local weak convergence

$$\sum_{\{\phi,v\}\in C_{n_j}^*} \xi_{\overline{K}_{n_j}}(\phi,v) \to \sum_{v\in C^*(\phi)} \xi_{\mathcal{T}}(\phi,v) \qquad \text{as } n\to\infty \text{ a.s.}$$

By Fatou's lemma

$$\liminf_{j\to\infty} \mathrm{E}\Bigg[\sum_{\{\phi,v\}\in C^*_{n_j}} \xi_{\overline{K}_{n_j}}(\phi, v)\Bigg] \ge \mathrm{E}\Bigg[\sum_{v\in C^*(\phi)} \xi_{\mathcal{T}}(\phi, v)\Bigg].$$

By Theorems 8 and 7,

$$\mathrm{E}\Bigg[\sum_{v\in C^*(\phi)} \xi_{\mathcal{T}}(\phi, v)\Bigg] \ge \mathrm{E}\Bigg[\sum_{v\in \mathcal{C}_{\mathrm{opt}}(\phi)} \xi_{\mathcal{T}}(\phi, v)\Bigg] = 2W(1) + W(1)^2.$$

This completes the proof. □

8. Belief propagation. To prove the upper bound on EC_n in order to complete the proof of Theorem 1, we will construct edge covers on $K_n, n \ge 1$, with costs $W(1) + W(1)^2/2 + o(1)$. This is achieved using belief propagation as described in Section 2.

We follow the approach of [14] to prove Theorem 2. In this section we will show the convergence of the BP algorithm on the PWIT $\mathcal{T}$, and relate the converged solution with the edge cover $\mathcal{C}_{\mathrm{opt}}$ of Section 6. In the next section we show that the belief propagation on $\overline{K}_n$ converges to belief propagation on $\mathcal{T}$ as $n \to \infty$.

8.1. *Convergence of BP on the PWIT.* In this section we will prove that the messages on $\mathcal{T}$ converge, and relate the resulting edge cover with the cover $\mathcal{C}_{\mathrm{opt}}$ of Section 6.

The message process can essentially be written as

$$X^{k+1}_{\mathcal{T}}(\dot{v}, v) = \min_{i\ge 1}\{(\xi_{\mathcal{T}}(v, v.i) - X^k_{\mathcal{T}}(v, v.i))^+\}, \tag{35}$$

where the initial messages $X^0_{\mathcal{T}}(\dot{v}, v)$ are i.i.d. random variables [zero in the case of our algorithm; see (2)].

By the structure of $\mathcal{T}$, it is clear that for a fixed $k \ge 0$, all the messages $X^k_{\mathcal{T}}(\dot{v}, v)$, $v \in \mathcal{V}$ share the same distribution. Also, it can be seen from the analysis of RDE (13) in Section 4 that if we denote the complementary c.d.f. of this distribution at some step k by $\overline{F}$, then after one update the complementary c.d.f. is given by the map

$$T\overline{F}(y) = \begin{cases} e^{-y}\exp\left(-\int_0^\infty \overline{F}(t)\,\mathrm{d}t\right), & \text{if } y \ge 0,\\ 1, & \text{if } y < 0.\end{cases}$$

The operator T thus defined on the space $\mathcal{D}$ of complementary c.d.f.'s of $\overline{\mathbf{R}}$-valued random variables has a unique fixed point $\overline{F}_*$ given by (15).

The following theorem shows that the fixed point $\overline{F}_*$ has the full space $\mathcal{D}$ as its domain of attraction. In other words, irrespective of the initial distribution, the common distribution of the messages $X^k_{\mathcal{T}}(\dot{v}, v)$, $v \in \mathcal{V}$ converges to the distribution F_* as $k \to \infty$.

THEOREM 10. *For any $\overline{F} \in \mathcal{D}$,*

$$\lim_{k\to\infty} T^k \overline{F} = \overline{F}_*.$$

PROOF. For any $y \geq 0$ and $k \geq 0$,

$$T^{k+1}\overline{F}(y) = e^{-y} \exp\left(-\int_0^\infty T^k \overline{F}(t)\,\mathrm{d}t\right).$$

Thus for $k \geq 1$, $T^k \overline{F}(y) = c_k e^{-y}$, where $c_k, k \geq 1$, are nonnegative real numbers satisfying

$$c_{k+1} = \exp\left(-\int_0^\infty c_k e^{-t}\,\mathrm{d}t\right) = e^{-c_k}.$$

It is easy to check that $c_k \to W(1)$. Consequently, $T^k \overline{F} \to \overline{F}_*$. □

8.2. *Endogeny and bivariate uniqueness.* We have established the convergence of the messages on $\mathcal{T}$ in distribution. We now ask for the joint convergence of the message process on the tree. In particular, the question is whether there is a limit process satisfying the requirements of Lemma 1.

An important property of the limiting process that allows us to come to this conclusion is *endogeny* introduced in [5]. Endogeny is a property of the recursive tree process (RTP) that it is measurable with respect to the i.i.d. process $(\xi_i, N_i), i \in \mathcal{V}$.

DEFINITION. An invariant RTP with marginal distribution μ is said to be endogenous if the root variable X_ϕ is almost surely measurable with respect to the σ-algebra

$$\sigma\big(\{(\xi_i, N_i) | i \in \mathcal{V}\}\big).$$

Endogeny is related to another property of the RTP termed as *bivariate uniqueness* again introduced in [5].

For a general RDE (14) write $T : \mathcal{P} \to \mathcal{P}(S)$ for the map induced by the function g. Let $\mathcal{P}^{(2)}$ denote the space of probability measures on $S \times S$ with marginals in $\mathcal{P}$. We now define a bivariate map $T^{(2)} : \mathcal{P}^{(2)} \to \mathcal{P}(S \times S)$, which maps a distribution $\mu^{(2)} \in \mathcal{P}^{(2)}$ to the joint distribution of

$$\begin{pmatrix} g\big(\xi; \big(X_j^{(1)}, 1 \leq j < N\big)\big) \\ g\big(\xi; \big(X_j^{(2)}, 1 \leq j < N\big)\big) \end{pmatrix},$$

where $(X_j^{(1)}, X_j^{(2)})_{j\geq 1}$ are independent with joint distribution $\mu^{(2)}$ on $S \times S$, and the family of random variables $(X_j^{(1)}, X_j^{(2)})_{j\geq 1}$ are independent of the pair (ξ, N).

It is easy to see that if μ is a fixed point of the RDE, then the associated *diagonal measure* $\mu^{\nearrow} := \mathrm{Law}(X, X)$ where $X \sim \mu$ is a fixed point of the operator $T^{(2)}$.

DEFINITION. An invariant RTP with marginal distribution μ is said to have the bivariate uniqueness property if $\mu^{\nearrow}$ is the unique fixed point of the operator $T^{(2)}$ with marginals μ.

Theorem 11 of [5] stated below shows that under certain assumptions, endogeny and bivariate uniqueness are equivalent.

THEOREM 11 (Theorem 11 of [5]). *Let S be a Polish space. Consider an invariant RTP with marginal distribution μ*:

(a) *If the endogenous property holds, then the bivariate uniqueness property holds.*

(b) *Conversely, suppose the bivariate uniqueness property holds. If also $T^{(2)}$ is continuous with respect to weak convergence on the set of bivariate distributions with marginals μ, then the endogenous property holds.*

(c) *The endogenous property holds if and only if $T^{(2)^n}(\mu \otimes \mu) \xrightarrow{\mathrm{D}} \mu^{\nearrow}$, where $\mu \otimes \mu$ is the product measure.*

The following theorem establishes the endogeny of the edge cover RDE.

THEOREM 12. *The invariant RTP with marginal μ_* (with c.d.f. F_*) associated with the edge cover RDE* (13) *is endogenous.*

PROOF. By Theorem 11(b) it is sufficient to prove bivariate uniqueness and continuity for the map $T^{(2)} : \mathcal{P}(\mathbf{R}_+ \times \mathbf{R}_+) \to \mathcal{P}(\mathbf{R}_+ \times \mathbf{R}_+)$, where $\mathbf{R}_+ = [0, \infty)$ and $T^{(2)}(\mu^{(2)})$ is the distribution of

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \min_{i \ge 1}(\xi_i - X_i)^+ \\ \min_{i \ge 1}(\xi_i - Y_i)^+ \end{pmatrix},$$

where $(X_i, Y_i)_{i \ge 1}$ are independent with joint distribution $\mu^{(2)}$ on $\mathbf{R}_+^2$, and are independent of $(\xi_i)_{i \ge 1}$ which are points of a Poisson process of rate 1 on $\mathbf{R}_+$.

To prove bivariate uniqueness, we have to show that if $\mu_*^{(2)}$ is a fixed point of the above map (with marginals μ_*), then $X = Y$ a.s. $(\mu_*^{(2)})$. By Lemma 1 of [6] this is equivalent to showing $X \stackrel{\mathrm{D}}{=} Y \stackrel{\mathrm{D}}{=} X \wedge Y$. Let $(X_i, Y_i)_{i \ge 1}$ be i.i.d. with distribution $\mu^{(2)}$. The set of points $\mathcal{P} := \{(\xi_i; (X_i, Y_i)) | i \ge 1\}$ forms a Poisson process on $(0, \infty) \times \mathbf{R}_+^2$ with intensity $\mathrm{d}t\, \mu_*^{(2)}(\mathrm{d}(x, y))$ at $(t; (x, y))$. Writing

$G(x, y) = \mathrm{P}\{X > x, Y > y\}$ for $x, y \in \mathbf{R}_+$, we get

$$
\begin{aligned}
G(x, y) &= \mathrm{P}\{\xi_i - X_i > x, \xi_i - Y_i > y, \text{ for all } i \geq 1\} \\
&= \mathrm{P}\{\text{No point of } \mathcal{P} \text{ in } \{(t; (u, v)) : t - u \leq x \text{ or } t - v \leq y\}\} \\
&= \exp\biggl(-\int_{t=0}^{x \vee y} \mathrm{d}t - \int_{t=x\vee y}^{\infty} \mathrm{P}\{t - X_1 \leq x \text{ or } t - Y_1 \leq y\}\,\mathrm{d}t\biggr) \\
&= e^{-x \vee y} \exp\biggl(-\int_{t=x\vee y}^{\infty} \mathrm{P}\{X_1 \geq t - x \text{ or } Y_1 \geq t - y\}\,\mathrm{d}t\biggr) \\
&= e^{-x \vee y} \exp\biggl(-\int_{t=x\vee y}^{\infty} \bigl(W(1)e^{-(t-x)} + W(1)e^{-(t-y)} \\
&\qquad\qquad - \mathrm{P}\{X_1 \geq t - x, Y_1 \geq t - y\}\bigr)\,\mathrm{d}t\biggr) \\
&= e^{-x \vee y} \exp\bigl(-W(1)e^{-x \vee y}(e^x + e^y)\bigr) \\
&\quad \times \exp\biggl(\int_{t=x\vee y}^{\infty} \mathrm{P}\{X_1 \geq t - x, Y_1 \geq t - y\}\,\mathrm{d}t\biggr).
\end{aligned} \tag{36}
$$

From this, setting $x = y$, it is clear that $G(x, x) = ce^{-x}$, $x \geq 0$, for some constant c. We now have to evaluate the constant.

Observe that the only place where $G(x, x)$ can be discontinuous (if at all) is at $x = 0$. As a consequence, with $x = y$ and the change of variable $z = t - x$, we see that the integral inside the exponent in (36) is $\int_0^\infty P(X_1 \geq z, Y_1 \geq z)\,\mathrm{d}z = \int_0^\infty P(X_1 > z, Y_1 > z)\,\mathrm{d}z = \int_0^\infty G(z, z)\,\mathrm{d}z$. With $x = y$ in (36), and integrating, we find that

$$c = e^{-2W(1)} e^c,$$

that is,

$$ce^{-c} = e^{-2W(1)}.$$

Since $W(1) = e^{-W(1)}$, it can be seen that $c = W(1)$ solves the above equation. Because $G(0, 0) \leq 1$, we have $c \leq 1$, and noting that the function $x \mapsto xe^{-x}$ is monotone increasing for $0 \leq x \leq 1$, we conclude that $c = W(1)$ is the only solution. Thus $G = \overline{F}_*$, that is, $X \wedge Y \stackrel{\mathrm{D}}{=} X \stackrel{\mathrm{D}}{=} Y$. This establishes bivariate uniqueness.

Now to establish endogeny it remains to prove the continuity hypothesis of Theorem 11(b). Note that we require continuity of the map $T^{(2)}$ only over the subset $\mathcal{P}_* \subset \mathcal{P}(\mathbf{R}_+^2)$ which contains probability distributions with both marginals equal to μ_*. We need to show that for any $\mu^{(2)} \in \mathcal{P}_*$ and a sequence $(\mu_n^{(2)})_{n\geq 1}$ in $\mathcal{P}_*$ such that $\mu_n^{(2)} \xrightarrow{\mathrm{D}} \mu^{(2)}$, we have $T^{(2)}(\mu_n^{(2)}) \xrightarrow{\mathrm{D}} T^{(2)}(\mu^{(2)})$.

Take a probability space $(\Omega, \mathcal{F}, P)$ in which there are random vectors $(X, Y) \sim \mu^{(2)}$ and a sequence of random vectors $\{(X_n, Y_n), n \geq 1\}$, with $(X_n, Y_n) \sim \mu_n^{(2)}$.

Then $(X_n, Y_n) \xrightarrow{\mathrm{D}} (X, Y)$. By following the steps of (36), for $x, y \in \mathbf{R}_+$, we can write

$$
\begin{aligned}
G_n(x, y) &= T^{(2)}(\mu_n^2)\big((x, \infty), (y, \infty)\big) \\
&= e^{-x\vee y} \exp\big(-W(1)e^{-x\vee y}(e^x + e^y)\big) \\
&\quad \times \exp\Big(\int_{t=x\vee y}^{\infty} \mathrm{P}\{X_n \ge t - x, Y_n \ge t - y\}\,\mathrm{d}t\Big) \\
&= e^{-x\vee y} \exp\big(-W(1)e^{-x\vee y}(e^x + e^y)\big) \\
&\quad \times \exp\Big(\int_{t=x\vee y}^{\infty} \mathrm{P}\big\{(X_n + x) \wedge (Y_n + y) \ge t\big\}\,\mathrm{d}t\Big) \\
&= e^{-x\vee y} \exp\big(-W(1)e^{-x\vee y}(e^x + e^y)\big) \\
&\quad \times \exp\big(\mathrm{E}\big[\big((X_n + x) \wedge (Y_n + y) - x \vee y\big)^+\big]\big).
\end{aligned} \tag{37}
$$

The same calculation also gives

$$
\begin{aligned}
G(x, y) &= T^{(2)}(\mu^{(2)})\big((x, \infty), (y, \infty)\big) \\
&= e^{-x\vee y} \exp\big(-W(1)e^{-x\vee y}(e^x + e^y)\big) \\
&\quad \times \exp\big(\mathrm{E}\big[\big((X + x) \wedge (Y + y) - x \vee y\big)^+\big]\big).
\end{aligned} \tag{38}
$$

Let

$$
\begin{aligned}
Z_n^{x,y} &:= \big((X_n + x) \wedge (Y_n + y) - x \vee y\big)^+ \quad \text{and} \\
Z^{x,y} &:= \big((X + x) \wedge (Y + y) - x \vee y\big)^+.
\end{aligned}
$$

Now $(X_n, Y_n) \xrightarrow{\mathrm{D}} (X, Y)$ implies that, for each (x, y), $Z_n^{x,y} \xrightarrow{\mathrm{D}} Z^{x,y}$. Now

$$0 \le Z_n^{x,y} \le X_n \qquad \text{for all } n \ge 1.$$

Since $\mathrm{E}X_n = \mathrm{E}X$ for all $n \ge 1$, by dominated convergence theorem, we have $\mathrm{E}Z_n^{x,y} \to \mathrm{E}Z^{x,y}$ as $n \to \infty$. Consequently $G_n(x, y) \to G(x, y)$ for all $x, y \in \mathbf{R}_+$. □

8.3. *Completing the proof of convergence of BP on the PWIT.* With endogeny in hand, we conclude that given a realization of $\mathcal{T}$, almost surely, the resulting stationary configuration of the X process of Lemma 1 is unique. Also, the following lemma will show that if the initial messages are i.i.d. random variables with the fixed point distribution μ_*, then the message process (35) converges, and the limit configuration is unique (almost surely).

LEMMA 6. *If the initial messages $X_{\mathcal{T}}^0(\dot{v}, v)$ are i.i.d. random variables with distribution μ_*, then the message process (35) converges in L^2 to the process X as $k \to \infty$.*

PROOF. Consider the evolution of bivariate messages according to (35), starting from $(X^0_{\mathcal{T}}(\cdot), X(\cdot))$. The second component will remain unchanged because the X process satisfies (16). The distribution of $(X^0_{\mathcal{T}}(\cdot), X(\cdot))$ is $\mu_* \otimes \mu_*$. We have

$$\mathrm{Law}\big(X^{k+1}_{\mathcal{T}}(\cdot), X(\cdot)\big) = T^{(2)}\big(\mathrm{Law}\big(X^k_{\mathcal{T}}(\cdot), X(\cdot)\big)\big).$$

Here $T^{(2)}$ is as defined in Theorem 12. By Theorem 11(c), $(X^k_{\mathcal{T}}(\cdot), X(\cdot))$ converges to $(X(\cdot), X(\cdot))$ in distribution as $k \to \infty$. Since $(X^k_{\mathcal{T}} - X)^2 \leq 2(X^k_{\mathcal{T}})^2 + 2X^2$, and $\mathrm{E}[2(X^k_{\mathcal{T}})^2 + 2X^2] = 4\mathrm{E}[X^2]$, the dominated convergence theorem gives $\mathrm{E}[(X^k_{\mathcal{T}} - X)^2] \to 0$ as $k \to \infty$. $\square$

We now prove that if the initial values are i.i.d. random variables with some arbitrary distribution (not necessarily μ_*), then the message process (35) does indeed converge to the unique stationary configuration. Of course, the initial condition of particular interest to us is the all zero initial condition (2), but we will prove a more general result.

The following lemma will allow us to interchange limit and minimization while working with the updates on $\mathcal{T}$.

LEMMA 7. *Let $X^0_{\mathcal{T}}(\dot{v}, v)$ be initialized to i.i.d. random variables with arbitrary distribution F on $\mathbf{R}_+$. Then the map*

$$\pi^k_{\mathcal{T}}(v) = \arg\min_{u \sim v}\big\{\big(\xi_{\mathcal{T}}(v,u) - X^k_{\mathcal{T}}(v,u)\big)^+\big\}$$

is a.s. well defined and finite for all $k \geq 1$, and

$$\sup_{k\geq 1} \mathrm{P}\Big\{\max \arg\min_{i\geq 1}\big\{\big(\xi_{\mathcal{T}}(v, v.i) - X^k_{\mathcal{T}}(v, v.i)\big)^+\big\} \geq i_0\Big\} \to 0 \qquad \text{as } i_0 \to \infty.$$

PROOF. Fix k. If $j \in \arg\min_{i\geq 1}\{(\xi_{\mathcal{T}}(v, v.i) - X^k_{\mathcal{T}}(v, v.i))^+\}$ and $j \geq 2$, then

$$\xi(v, v.j) - X^k_{\mathcal{T}}(v, v.j) \leq \big(\xi_{\mathcal{T}}(v, v.1) - X^k_{\mathcal{T}}(v, v.1)\big)^+.$$

Now

$$\begin{aligned} &\mathrm{P}\big\{\xi(v, v.j) - X^k_{\mathcal{T}}(v, v.j) \leq \big(\xi_{\mathcal{T}}(v, v.1) - X^k_{\mathcal{T}}(v, v.1)\big)^+\big\} \\ &\quad \leq \mathrm{P}\big\{\xi(v, v.j) \leq X^k_{\mathcal{T}}(v, v.j)\big\} \\ &\qquad + \mathrm{P}\big\{\xi(v, v.j) - X^k_{\mathcal{T}}(v, v.j) \leq \xi(v, v.1) - X^k_{\mathcal{T}}(v, v.1)\big\}. \end{aligned} \tag{39}$$

The updates are such that $\{X^k_{\mathcal{T}}(v, v.i), i \geq 1\}$ remain i.i.d. and independent of the Poisson process $\{\xi(v, v.i)\}$. Thus the probability on the right-hand side of (39) equals

$$\mathrm{P}\big\{\xi_j \leq X^k_1\big\} + \mathrm{P}\big\{\xi_{j-1} \leq X^k_2 - X^k_1\big\},$$

where $\{\xi_i\}$ is a Poisson process and X_1^k, X_2^k are independent random variables with same distribution as $X_{\mathcal{T}}^k(v, v.1)$. Then

$$\begin{aligned}
&\sum_{j=2}^{\infty} \mathrm{P}\Big\{ j \in \arg\min_{i \ge 1} \{ (\xi_{\mathcal{T}}(v, v.i) - X_{\mathcal{T}}^k(v, v.i))^+ \} \Big\} \\
&\quad \le \sum_{j=2}^{\infty} (\mathrm{P}\{\xi_j \le X_1^k\} + \mathrm{P}\{\xi_{j-1} \le X_2^k - X_1^k\}) \\
&\quad \le \sum_{j=1}^{\infty} \mathrm{P}\{\xi_j \le X_1^k\} + \sum_{j=1}^{\infty} \mathrm{P}\{\xi_j \le X_2^k - X_1^k\} \qquad (40) \\
&\quad = \mathrm{E}X_1^k + \mathrm{E}|X_1^k - X_2^k| \\
&\quad \le 3\mathrm{E}X_1^k.
\end{aligned}$$

From the proof of Theorem 10 it follows that $\mathrm{E}X_1^k$ converges, and hence it is bounded. This proves that the arg min is a.s. finite and the probability in the statement of the lemma, being upper bounded by the tail sum of the left-hand side of (40), converges uniformly to 0. □

We are now in a position to prove the required convergence.

THEOREM 13. *The recursive tree process defined by* (35) *with i.i.d. initial messages converges to the unique stationary configuration in the following sense. For every* $v \in \mathcal{V}$,

$$X_{\mathcal{T}}^k(v, v.i) \xrightarrow{L^2} X(v, v.i) \qquad \text{as } k \to \infty.$$

Also, the decisions at the root converge, that is, $\mathrm{P}\{\pi_{\mathcal{T}}^k(\phi) \ne \mathcal{C}_{\mathrm{opt}}(\phi)\} \to 0$ *as* $k \to \infty$.

PROOF. The proof is essentially identical to the proof of Theorem 5.2 of [14]. We present it here for completeness.

Let F be the c.d.f. of the initial distribution. Let $\theta_t, t \in \mathbf{R}$ denote the t-shift operator on $\mathcal{D}$, that is, $\theta_t \overline{F} : x \mapsto \overline{F}(x - t)$. Since $T^n \overline{F} \to \overline{F}_*$, and $T^n \overline{F}$ are of the form $y \mapsto c_n e^{-y}$, $y \ge 0$ for $n \ge 1$, for any $\varepsilon > 0$ there exists $k_\varepsilon \in \mathbf{N}$ such that

$$\theta_{-\varepsilon} \overline{F}_* \le T^{k_\varepsilon} \overline{F} \le \theta_\varepsilon \overline{F}_*.$$

By Strassen's theorem, probability measures satisfying such an ordering can be coupled in a pointwise monotone manner. In other words, there exists a probability space $E' = (\Omega', \mathscr{F}', P')$, possibly differing from the original space $E =$

$(\Omega, \mathscr{F}, P)$, on which we can define a random variable X^ε with complementary c.d.f. $T^{k_\varepsilon}\overline{F}$ and two random variables X^- and X_+ with distribution $\overline{F}_*$ such that almost surely

$$X^- - \varepsilon \le X^\varepsilon \le X^+ + \varepsilon. \tag{41}$$

We now define over the product space $(\bigotimes_{v\in\mathcal{V}} E') \otimes E$ the PWIT $\mathcal{T}$ and independent copies $(X_v^-, X_v^\varepsilon, X_v^+)_{v\in\mathcal{V}}$ of the triple $(X^-, X^\varepsilon, X^+)$.

On $\mathcal{T}$, we look at the message process with three different initializations:

$$X_{\mathcal{T}}^{0,-}(\dot{v}, v) = X_v^-, \qquad X_{\mathcal{T}}^{0,\varepsilon}(\dot{v}, v) = X_v^\varepsilon \quad \text{and} \quad X_{\mathcal{T}}^{0,+}(\dot{v}, v) = X_v^+ \qquad \forall v \in \mathcal{V}.$$

From the update rule (35) one can readily verify that the ordering between the messages is preserved in the following sense. For any $v \in \mathcal{V}$ and $k \ge 0$,

$$X_{\mathcal{T}}^{2k,-}(\dot{v}, v) - \varepsilon \le X_{\mathcal{T}}^{2k,\varepsilon}(\dot{v}, v) \le X_{\mathcal{T}}^{2k,+}(\dot{v}, v) + \varepsilon;$$
$$X_{\mathcal{T}}^{2k+1,+}(\dot{v}, v) - \varepsilon \le X_{\mathcal{T}}^{2k+1,\varepsilon}(\dot{v}, v) \le X_{\mathcal{T}}^{2k+1,-}(\dot{v}, v) + \varepsilon.$$

Now fix a $v \in \mathcal{V}$, and observe that

$$\big(X_{\mathcal{T}}^{k+k_\varepsilon}(\dot{v}, v)\big)_{k\ge 0} \overset{\mathrm{D}}{=} \big(X_{\mathcal{T}}^{k,\varepsilon}(\dot{v}, v)\big)_{k\ge 0}.$$

It follows that for every $k \ge k_\varepsilon$,

$$\begin{aligned}
&\sup_{s,t\ge k} \big\| X_{\mathcal{T}}^s(\dot{v}, v) - X_{\mathcal{T}}^t(\dot{v}, v)\big\|_{L^2} \\
&\qquad = \sup_{s,t\ge k-k_\varepsilon} \big\| X_{\mathcal{T}}^{s,\varepsilon}(\dot{v}, v) - X_{\mathcal{T}}^{t,\varepsilon}(\dot{v}, v)\big\|_{L^2} \\
&\qquad \le 2 \sup_{t\ge k-k_\varepsilon} \big\| X_{\mathcal{T}}^{t,\pm}(\dot{v}, v) - X(\dot{v}, v)\big\|_{L^2} + 2\varepsilon.
\end{aligned}$$

From endogeny and Lemma 6, it follows that

$$\sup_{t\ge k-k_\varepsilon} \big\| X_{\mathcal{T}}^{t,\pm}(\dot{v}, v) - X(\dot{v}, v)\big\|_{L^2} \to 0 \qquad \text{as } k \to \infty.$$

Thus the sequence $(X_{\mathcal{T}}^k(\dot{v}, v))_{k\ge 0}$ is Cauchy in L^2, and hence convergent. Now, Lemma 7 allows us to interchange limit and minimization in (35) to conclude that the limit process has to be a fixed point of (35). By endogeny there is a unique stationary configuration a.s. on any realization of the PWIT. Hence the limit configuration has to be identical to the X process.

Again by Lemma 7, for any $\varepsilon > 0$, we can choose an i_0 such that

$$\mathrm{P}\{\pi_{\mathcal{T}}^k(\phi) \nsubseteq \{1, 2, \dots, i_0\}\} < \varepsilon/3$$

for all $k \ge 1$, and $\mathrm{P}\{\mathcal{C}_{\mathrm{opt}}(\phi) \nsubseteq \{1, 2, \dots, i_0\}\} < \varepsilon/3$. Now, the convergence of $X_{\mathcal{T}}^k$ to X implies that for k sufficiently large, when $\pi_{\mathcal{T}}^k(\phi)$ and $\mathcal{C}_{\mathrm{opt}}(\phi)$ are contained in $\{1, 2, \dots, i_0\}$, the probability that the two maps differ is less than $\varepsilon/3$. This proves the second statement of the theorem. □

9. Belief propagation on $\overline{K}_n$.

9.1. *Convergence of the update rule on $\overline{K}_n$ to the update rule on $\mathcal{T}$.* We use from [14] the modified definition of local convergence applied to geometric networks with edge labels, that is, networks in which each directed edge (v, w) has a label $\lambda(v, w)$ taking values in some Polish space. For local convergence of a sequence of such labeled networks $G_1, G_2, \ldots$ to a labeled geometric network G_∞, we add the additional requirement that the rooted graph isomorphisms $\gamma_{n,\rho}$ satisfy

$$\lim_{n\to\infty} \lambda_{G_n}\big(\gamma_{n,\rho}(v, w)\big) = \lambda_{G_\infty}(v, w)$$

for each directed edge (v, w) in $\mathcal{N}_\rho(G_\infty)$.

Now we view the configuration of BP on a graph G at the kth iteration as a labeled geometric network with the label on edge (v, w) given by the pair

$$\big(X_G^k(v, w), \mathbf{1}_{\{v\in\pi_G^k(w)\}}\big).$$

With this definition, our convergence result can be written as the following theorem.

THEOREM 14. *For every fixed $k \geq 0$, the kth step configuration of BP on $\overline{K}_n$ converges in the local weak sense to the kth step configuration of BP on $\mathcal{T}$.*

$$\big(\overline{K}_n, X_{\overline{K}_n}^k(v, w), \mathbf{1}_{\{v\in\pi_{\overline{K}_n}^k(w)\}}\big) \xrightarrow{\text{l.w.}} \big(\mathcal{T}, X_{\mathcal{T}}^k(v, w), \mathbf{1}_{\{v\in\pi_{\mathcal{T}}^k(w)\}}\big). \tag{42}$$

PROOF. The proof of this theorem proceeds along the lines of the proof of Theorem 4.1 of [14].

Consider an almost sure realization of the convergence $\overline{K}_n \to \mathcal{T}$.

Recall from Section 3 the labeling of the vertices of $\mathcal{T}$ from the set $\mathcal{V}$. We now recursively apply multiple labels from $\mathcal{V}$ to the vertices of $\overline{K}_n$. Label the root as ϕ. If $v \in \mathcal{V}$ denotes a vertex x of $\overline{K}_n$, then $(v.1, v.2, \ldots, v.(n-1)$ denote the neighbors of x in $\overline{K}_n$ ordered by increasing lengths of the corresponding edge with x. Then the convergence in (42) is shown if we argue that

$$\forall\{v, w\} \in \mathcal{E} \qquad X_{\overline{K}_n}^k(v, w) \xrightarrow{\text{P}} X_{\mathcal{T}}^k(v, w) \quad \text{and}$$
$$\forall v \in \mathcal{V} \qquad \pi_{\overline{K}_n}^k(v) \xrightarrow{\text{P}} \pi_{\mathcal{T}}^k(v) \qquad \text{as } n \to \infty.$$

The above is trivially true for $k = 0$. Writing the update and decision rules as

$$X_{\overline{K}_n}^{k+1}(w, v) = \min_{u\in\{v.1,\ldots,v.(n-1),\dot{v}\}\setminus\{w\}} \big\{\big(\xi_{\overline{K}_n}(v, u) - X_{\overline{K}_n}^k(v, u)\big)^+\big\} \quad \text{and}$$
$$\pi_{\overline{K}_n}^k(v) = \mathop{\arg\min}_{u\in\{v.1,\ldots,v.(n-1),\dot{v}\}} \big\{\big(\xi_{\overline{K}_n}(v, u) - X_{\overline{K}_n}^k(v, u)\big)^+\big\},$$

we may try to use the convergence of each term on the right-hand side inductively to conclude the convergence of the term on the left. This is not directly possible as the minimum is over an unbounded number of terms as $n \to \infty$. However the following lemma allows us to restrict attention to a uniformly bounded number of terms for each n with probability as high as desired, and hence obtain convergence in probability for each $k \geq 0$. □

LEMMA 8. *For all $v \in \mathcal{V}$ and $k \geq 0$,*

$$\lim_{i_0\to\infty} \limsup_{n\to\infty} \mathrm{P}\Big\{\max \operatorname*{arg\,min}_{1\leq i\leq n-1} \{(\xi_{\overline{K}_n}(v, v.i) - X^k_{\overline{K}_n}(v, v.i))^+\} \geq i_0\Big\} = 0.$$

PROOF. The proof is the same as the proof of Lemma 4.1 of [14]. The only thing to keep in mind is arg min is a set, and we target the largest index, but the same proof applies. □

9.2. *Completing the upper bound*: *Proof of Theorem* 2. By Theorem 13, $\pi^k_{\mathcal{T}}(\phi) \xrightarrow{\mathrm{P}} \mathcal{C}_{\mathrm{opt}}(\phi)$ as $k \to \infty$. It follows that

$$\sum_{v\in\pi^k_{\mathcal{T}}(\phi)} \xi_{\mathcal{T}}(\phi, v) \xrightarrow{\mathrm{P}} \sum_{v\in\mathcal{C}_{\mathrm{opt}}(\phi)} \xi_{\mathcal{T}}(\phi, v) \qquad \text{as } k\to\infty. \tag{43}$$

We now prove convergence in expectation. Observe that

$$v \in \pi^k_{\mathcal{T}}(\phi) \quad\Longrightarrow\quad \xi_{\mathcal{T}}(\phi, v) - X^k_{\mathcal{T}}(\phi, v) \leq (\xi_{\mathcal{T}}(\phi, 1) - X^k_{\mathcal{T}}(\phi, 1))^+ \leq \xi_{\mathcal{T}}(\phi, 1).$$

By (35), $X^k_{\mathcal{T}}(\phi, v) \leq \xi_{\mathcal{T}}(v, v.1)$. Thus

$$v \in \pi^k_{\mathcal{T}}(\phi) \quad\Longrightarrow\quad \xi_{\mathcal{T}}(\phi, v) \leq \xi_{\mathcal{T}}(\phi, 1) + \xi_{\mathcal{T}}(v, v.1). \tag{44}$$

This implies

$$\sum_{v\in\pi^k_{\mathcal{T}}(\phi)} \xi_{\mathcal{T}}(\phi, v) \leq \xi_{\mathcal{T}}(\phi, 1) + \sum_{i\geq 2} \xi_{\mathcal{T}}(\phi, i)\mathbf{1}_{\{\xi_{\mathcal{T}}(\phi,i)\leq\xi_{\mathcal{T}}(\phi,1)+\xi_{\mathcal{T}}(i,i.1)\}}.$$

It can be verified that the sum on the right-hand side in the above equation is an integrable random variable. Equation (43) and the dominated convergence theorem give

$$\begin{aligned}\lim_{k\to\infty} \mathrm{E}\Big[\sum_{v\in\pi^k_{\mathcal{T}}(\phi)} \xi_{\mathcal{T}}(\phi, v)\Big] &= \mathrm{E}\Big[\sum_{v\in\mathcal{C}_{\mathrm{opt}}(\phi)} \xi_{\mathcal{T}}(\phi, v)\Big] \\ &= 2W(1) + W(1)^2,\end{aligned} \tag{45}$$

where the last equality follows from Theorem 7.

By Theorem 14 and Lemma 8, using the definition of local weak convergence, we have

$$\sum_{v\in\pi^k_{\overline{K}_n}(\phi)} \xi_{\overline{K}_n}(\phi, v) \xrightarrow{\mathrm{P}} \sum_{v\in\pi^k_{\mathcal{T}}(\phi)} \xi_{\mathcal{T}}(\phi, v) \qquad \text{as } n\to\infty. \tag{46}$$

We now apply the arguments that lead to (44) to the edge covers $\pi^k_{\overline{K}_n}(\phi)$, and obtain

$$v\in\pi^k_{\overline{K}_n}(\phi) \quad\Longrightarrow\quad \xi_{\overline{K}_n}(\phi, v) \le \xi_{\overline{K}_n}(\phi, 1) + \xi_{\overline{K}_n}(v, v.1).$$

For any two vertices u, v of $\overline{K}_n$, define $S_n(u, v) = \min_{w\neq u,v} \xi_{\overline{K}_n}(u, w)$. Then for a vertex v of $\overline{K}_n$, $\xi_{\overline{K}_n}(\phi, 1) \le S_n(\phi, v)$ and $\xi_{\overline{K}_n}(v, v.1) \le S_n(v, \phi)$. This gives

$$v\in\pi^k_{\overline{K}_n}(\phi) \quad\Longrightarrow\quad \xi_{\overline{K}_n}(\phi, v) \le S_n(\phi, v) + S_n(v, \phi).$$

Consequently,

$$\sum_{v\in\pi^k_{\overline{K}_n}(\phi)} \xi_{\overline{K}_n}(\phi, v) \le \sum_{v} \xi_{\overline{K}_n}(\phi, v)\mathbf{1}_{\{\xi_{\overline{K}_n}(\phi,v)\le S_n(\phi,v)+S_n(v,\phi)\}}. \tag{47}$$

Observe that $\xi_{\overline{K}_n}(\phi, v)$, $S_n(\phi, v)$ and $S_n(v, \phi)$ are independent exponential random variables with means $n, n/(n-2)$ and $n/(n-2)$, respectively. So we can write

$$\begin{aligned}
&\mathrm{E}\big[\xi_{\overline{K}_n}(\phi, v)\mathbf{1}_{\{\xi_{\overline{K}_n}(\phi,v)\le S_n(\phi,v)+S_n(v,\phi)\}}\big]\\
&\quad= \int_0^\infty \int_0^x \frac{t}{n} e^{-t/n}\,\mathrm{d}t \left(\frac{n-2}{n}\right)^2 x e^{-((n-2)/n)x}\,\mathrm{d}x\\
&\quad= \frac{3n^2-5n}{(n-1)^3}.
\end{aligned}$$

Summing over all neighbors of ϕ, we get

$$\mathrm{E}\Big[\sum_{v} \xi_{\overline{K}_n}(\phi, v)\mathbf{1}_{\{\xi_{\overline{K}_n}(\phi,v)\le S_n(\phi,v)+S_n(v,\phi)\}}\Big] = \frac{3n^2-5n}{(n-1)^2}, \tag{48}$$

which converges to 3 as $n\to\infty$.

Using local weak convergence, we can see that

$$\begin{aligned}
&\sum_{v} \xi_{\overline{K}_n}(\phi, v)\mathbf{1}_{\{\xi_{\overline{K}_n}(\phi,v)\le S_n(\phi,v)+S_n(v,\phi)\}}\\
&\qquad \xrightarrow{\mathrm{P}} \xi_{\mathcal{T}}(\phi, 1) + \sum_{i\ge 2} \xi_{\mathcal{T}}(\phi, i)\mathbf{1}_{\{\xi_{\mathcal{T}}(\phi,i)\le \xi_{\mathcal{T}}(\phi,1)+\xi_{\mathcal{T}}(i,i.1)\}}.
\end{aligned}$$

It can be verified that the expectation of the random variable on the right-hand side above equals 3. Using this with (46), (47) and (48), the generalized dominated convergence theorem yields

$$\lim_{n\to\infty} \mathrm{E}\Bigg[\sum_{v\in\pi^k_{\overline{K}_n}(\phi)} \xi_{\overline{K}_n}(\phi, v)\Bigg] = \mathrm{E}\Bigg[\sum_{v\in\pi^k_{\mathcal{T}}(\phi)} \xi_{\mathcal{T}}(\phi, v)\Bigg]. \tag{49}$$

Combining (49) and (45) gives

$$\lim_{k\to\infty}\lim_{n\to\infty} \mathrm{E}\Bigg[\sum_{v\in\pi^k_{\overline{K}_n}(\phi)} \xi_{\overline{K}_n}(\phi, v)\Bigg] = 2W(1) + W(1)^2. \tag{50}$$

The expectation in the statement of Theorem 2 can be written as

$$\begin{aligned}\mathrm{E}\Bigg[\sum_{e\in\mathcal{C}(\pi^k_{K_n})} \xi_{K_n}(e)\Bigg] &= \frac{1}{2}\mathrm{E}\Bigg[\sum_v \sum_{w\in\pi^k_{K_n}(v)} \xi_{K_n}(v, w)\Bigg]\\ &= \frac{1}{2}\mathrm{E}\Bigg[\sum_v \frac{1}{n} \sum_{w\in\pi^k_{\overline{K}_n}(v)} \xi_{\overline{K}_n}(v, w)\Bigg]\\ &= \frac{1}{2}\mathrm{E}\Bigg[\sum_{v\in\pi^k_{\overline{K}_n}(\phi)} \xi_{\overline{K}_n}(\phi, v)\Bigg].\end{aligned} \tag{51}$$

In the first equality above we count the contribution of the edges of the cover incident at each vertex of K_n. The factor of $1/2$ appears because each edge in the edge cover appears twice, once for each of its endpoints. The $1/n$ in the second equality accounts for the scaling of edge costs from K_n to $\overline{K}_n$. The third equality holds because the root ϕ in $\overline{K}_n$ is chosen uniformly at random from the n vertices. Equation (50) now completes the proof of Theorem 2.

9.3. *Completing the proof of Theorem* 1. Applying the scaling in (51) to the optimal edge covers in K_n and $\overline{K}_n$, we get

$$\mathrm{E}C_n = \frac{1}{2}\mathrm{E}\Bigg[\sum_{\{\phi,v\}\in C^*_n} \xi_{\overline{K}_n}(\phi, v)\Bigg].$$

Theorem 9 gives the lower bound

$$\liminf_{n\to\infty} \mathrm{E}C_n \geq W(1) + \frac{W(1)^2}{2}.$$

By Theorem 2 for any $\varepsilon > 0$, we can find k large such that

$$\lim_{n\to\infty} \mathrm{E}\Bigg[\sum_{e\in\mathcal{C}(\pi^k_{K_n})} \xi_{K_n}(e)\Bigg] \leq W(1) + \frac{W(1)^2}{2} + \varepsilon.$$

This gives

$$\limsup_{n\to\infty} \mathrm{E}C_n \le W(1) + \frac{W(1)^2}{2} + \varepsilon.$$

Since ε is arbitrary, we get the upper bound

$$\limsup_{n\to\infty} \mathrm{E}C_n \le W(1) + \frac{W(1)^2}{2}.$$

This completes the proof of Theorem 1.

Observe that for any $\varepsilon > 0$, there exist K_ε and N_ε such that for all $k \ge K_\varepsilon$ and $n \ge N_\varepsilon$, we have

$$\mathrm{E}\biggl[\sum_{e\in\mathcal{C}(\pi^k_{K_n})} \xi_{K_n}(e)\biggr] \le W(1) + \frac{W(1)^2}{2} + \varepsilon.$$

Thus for large n the BP algorithm gives a solution with cost within ε of the optimal value in K_ε iterations. In an iteration, the algorithm requires $O(n)$ computations at every vertex. This gives an $O(K_\varepsilon n^2)$ running time for the BP algorithm to compute an ε-approximate solution. The worst case complexity of the edge-cover problem is $O(n^3)$, a result due to Edmonds and Johnson (1970); see [15], Theorem 27.2.

10. More results. Our main results for the edge-cover problem were the proof of the limit of the expected minimum cost (Theorem 1) and the means to obtain an asymptotically optimal solution using the BP algorithm (Theorem 2). The use of objective method as the proof technique allows us to obtain several auxiliary results about the structure of the optimal solution, through calculations for the edge cover $\mathcal{C}_{\mathrm{opt}}$ on the PWIT. In this section we state and prove, as examples, results for the distribution of the degree of the root and the probability that the least cost edge at the root is part of the optimal edge cover $\mathcal{C}_{\mathrm{opt}}$. It is easy to show using local weak convergence and the results of Sections 8 and 9 that these quantities arise as limits of the quantities corresponding to the edge covers $\pi^k_{K_n}$.

THEOREM 15.

$$\mathrm{P}\{|\mathcal{C}_{\mathrm{opt}}(\phi)| = 1\} = e^{-W(1)}\bigl(1 + W(1)\bigr).$$

For $k \ge 2$,

$$\mathrm{P}\{|\mathcal{C}_{\mathrm{opt}}(\phi)| = k\} = e^{-W(1)}\frac{W(1)^k}{k!}.$$

PROOF. As in the proof of Theorem 6, $\{(\xi_j, X_j), j \ge 1\}$ is a Poisson process on $\mathbf{R}_+ \times \mathbf{R}_+$ with intensity $\mathrm{d}z\,\mathrm{d}F_*(x)$.

From the definition of $\mathcal{C}_{\rm opt}$,

$$\begin{aligned}\mathrm{P}\{|\mathcal{C}_{\rm opt}(\phi)|=1\} &= \mathrm{P}\{\text{at most one point of } \{(\xi_j, X_j)\} \text{ in } \{(z,x): z-x\le 0\}\}\\ &= e^{-A}(1+A),\end{aligned}$$

where

$$\begin{aligned}A &= \int_{z=0}^{\infty}\int_{x=z}^{\infty} \mathrm{d}F_*(x)\,\mathrm{d}z\\ &= \int_{z=0}^{\infty} W(1)e^{-z}\,\mathrm{d}z\\ &= W(1).\end{aligned}$$

Thus

$$\mathrm{P}\{|\mathcal{C}_{\rm opt}(\phi)|=1\} = e^{-W(1)}\big(1+W(1)\big).$$

For $k \ge 2$,

$$\begin{aligned}\mathrm{P}\{|\mathcal{C}_{\rm opt}(\phi)|=k\} &= \mathrm{P}\{k \text{ points of } \{(\xi_j, X_j)\} \text{ in } \{(z,x): z-x\le 0\}\}\\ &= e^{-A}\frac{A^k}{k!}\\ &= e^{-W(1)}\frac{W(1)^k}{k!}.\end{aligned}$$

□

THEOREM 16.

$$\mathrm{P}\{1 \in \mathcal{C}_{\rm opt}(\phi)\} = \frac{W(1)}{2} + \frac{1}{W(1)} - W(1)^2 - 1.$$

PROOF. The event $\{1 \in \mathcal{C}_{\rm opt}(\phi)\}$ equals the union of two disjoint events:

(a) $\xi(\phi,1) - X(\phi,1) < 0$ and
(b) $0 \le \xi(\phi,1) - X(\phi,1) \le \xi(\phi,i) - X(\phi,i)$ for all $i \ge 2$.

The probability of the first event is

$$\begin{aligned}\mathrm{P}\{\xi(\phi,1) - X(\phi,1) < 0\} &= \int_{z=0}^{\infty}\int_{x=z}^{\infty} \mathrm{d}F_*(x)e^{-z}\,\mathrm{d}z\\ &= \int_{z=0}^{\infty} W(1)e^{-z}e^{-z}\,\mathrm{d}z\\ &= \frac{W(1)}{2}.\end{aligned}$$

For the second event, write $\xi(\phi,i) = \xi(\phi,1) + \xi_i'$, where obviously $\{\xi_i', i \ge 2\}$ is a rate 1 Poisson process independent of $\{X(\phi,i), i \ge 2\}$. For $i \ge 2$, $\xi(\phi,1) -$

$X(\phi,1) \le \xi(\phi,i) - X(\phi,i)$ if and only if $-X(\phi,1) \le \xi_i' - X(\phi,i)$. The probability of the second event can be written as

$$
\begin{aligned}
&\mathrm{P}\{0 \le \xi(\phi,1) - X(\phi,1) \le \xi(\phi,i) - X(\phi,i) \text{ for all } i \ge 2\} \\
&\quad = \int_{x_1=0}^{\infty}\int_{z_1=x_1}^{\infty} P\{\text{no point of } \{(\xi_i', X(\phi,i), i \ge 2)\} \\
&\qquad\qquad\qquad\qquad \text{in } \{(z,x): z - x \le -x_1\}\} e^{-z_1}\,\mathrm{d}z_1\,\mathrm{d}F_*(x_1) \\
&\quad = \int_{x_1=0}^{\infty} e^{-x_1}\exp\biggl(-\int_{z=0}^{\infty}\int_{x=z+x_1}^{\infty}\mathrm{d}F_*(x)\,\mathrm{d}z\biggr)\,\mathrm{d}F_*(x_1) \\
&\quad = \int_{x_1=0}^{\infty} e^{-x_1}\exp\biggl(-\int_{z=0}^{\infty} W(1)e^{-z}e^{-x_1}\,\mathrm{d}z\biggr)\,\mathrm{d}F_*(x_1) \\
&\quad = \int_{x_1=0}^{\infty} e^{-x_1}\exp(-W(1)e^{-x_1})\,\mathrm{d}F_*(x_1) \\
&\quad = W(1)\bigl(1 - W(1)\bigr) + \int_{x_1=0}^{\infty} W(1)e^{-2x_1}\exp(-W(1)e^{-x_1})\,\mathrm{d}x_1 \\
&\quad = W(1)\bigl(1 - W(1)\bigr) + \frac{1}{W(1)} - W(1) - 1 \\
&\quad = \frac{1}{W(1)} - W(1)^2 - 1. \qquad \square
\end{aligned}
$$

11. Summary. In a nutshell, we have implemented Aldous's program based on [4] to solve the random edge-cover problem. Aldous's program serves as a rigorous mathematical alternative to the cavity method applied to mean-field combinatorial optimization problems. Aldous and Bandyopadhyay [5], Section 7.5, outline the steps of this rigorous methodology, highlighting the role of RDEs and endogeny. See below.

But first, we must indicate another way in which the complete graph with i.i.d. edge weights arises. Combinatorial optimization problems involving n random points on $\mathbb{R}^d$ are of interest in many physical settings, but are typically difficult to analyze because of dependence of the random variables representing the $\binom{n}{2}$ distances. A more tractable *mean-field model* ignores the underlying d-dimensional space, and simply models the interpoint distances as i.i.d. random variables. This resulting model is then the complete graph on n vertices with i.i.d. edge weights. The case of exponential mean 1 edge weights models the $d = 1$ setting. There are other distributions to model the $d > 1$ settings. Though we did not deal with $d > 1$ in this paper, we expect the extension to hold (as for matching).

Let us return to Aldous's program, as summarized by Aldous and Bandyopadhyay [5], Section 7.5, and reproduced below.

> "Start with a combinatorial optimization problem over some size-n random structure.

- Formulate a "size-∞" random structure, the $n \to \infty$ limit in the sense of local weak convergence.
- Formulate a corresponding combinatorial optimization problem on the size-∞ structure.
- Heuristically define relevant quantities on the size-∞ structure via additive renormalization ...
- If the size-∞ structure is treelike (the only case where one expects exact asymptotic solutions), observe that the relevant quantities satisfy a problem dependent RDE.
- Solve the RDE. Use the unique solution to find the value of the optimization problem on the size-∞ structure.
- Show that the RTP associated with the solution is endogenous.
- Endogeny shows that the optimal solution is a measurable function of the data, in the infinite-size problem. Since a measurable function is almost continuous, we can pull back to define almost-feasible solutions of the size-n problem with almost the same cost.
- Show that in the size-n problem one can patch an almost-feasible solution into a feasible solution for asymptotically negligible cost." [5], Section 7.5.

The size-n random structure is the complete graph on n-vertices $\overline{K}_n$ with independent exponential mean-n edge weights. The following points elaborate on how we addressed the steps above:

- The size-∞ random structure is the PWIT.
- The corresponding optimization problem on the size-∞ structure is simply the minimum-cost edge cover on the PWIT. While this step is easy for the edge-cover problem, in general some subtleties are involved. For example, the limiting size-∞ problem for Frieze's size-n problem of minimal spanning tree on $\overline{K}_n$ [8] is a minimal spanning forest with certain requirements on the included edges. See [3], Definition 4.2, for details.
- We then heuristically provided the quantities relevant to the edge-cover problem on the PWIT in Section 4. The additive renormalization measured the reduction in cost arising from the relaxation of the requirement that the root be hit.
- Using the tree structure of the limiting object, we obtained the RDE (13) associated with the edge-cover problem.
- We solved the RDE in Theorem 6, showed that it had a unique solution, and found the value of the optimization problem on the PWIT in Theorem 7. Another important step is Theorem 8 which proves that the edge cover $\mathcal{C}_{\text{opt}}$, based on the heuristic relation (10), is optimal among involution invariant edge covers on the PWIT. Our method for establishing this nontrivial step may have some bearing on other similar combinatorial optimization problems. This step eventually established a lower bound for the liminf of size-n optimal values.

- Theorem 12 established endogeny of the RTP associated with the solution of (13). Theorem 2 corresponding to the BP algorithm on K_n replaces the procedure of Aldous's program for obtaining solutions of the size-n problem from the solution of the size-∞ problem. The key steps for this are based on Salez and Shah's approach [14] and is as follows. Using endogeny, we argued that BP (with i.i.d. initializations) converges to the RDE-based stationary configuration on the PWIT. We then established that, at a particular node of $\overline{K}_n$, the BP update for large n depends essentially only on messages from its local neighborhood (Lemma 8). This is then used to express BP on the PWIT as the limit of BP on $\overline{K}_n$. The BP iterates on $\overline{K}_n$ were then the candidate solutions for the size-n problem.
- No corrective patch-up was needed for the size-n problem, since at each iteration of the BP algorithm, every vertex was covered by the corresponding selection of edges. Simple dominated convergence arguments then established the convergence of the expected optimal costs to the correct value.

It is worth noting that the upper bound result in Theorem 1 can be obtained via a simpler proof of Theorem 2 for a version of BP algorithm, where the messages are initialized as i.i.d. random variables from the fixed-point distribution F_*. In this case Lemma 6, which follows from endogeny, establishes the convergence result on the PWIT. The more general result of Theorem 13 shows that BP works when messages are initialized as i.i.d. random variables from any arbitrary distribution.

Finally, we must mention that Aldous [4] proved a strong property called *asymptotic essential uniqueness* for matching, which is roughly the property that if a matching on $\overline{K}_n$ is almost optimal, then it coincides with the optimal matching, except on a small proportion of edges. The question of whether this property holds for the edge-cover problem is one that we hope to address in the near future.

Acknowledgments. Part of this work was carried out when Rajesh Sundaresan was on sabbatical leave at the University of Illinois at Urbana–Champaign whose support is gratefully acknowledged.

REFERENCES

[1] ALDOUS, D. (1992). Asymptotics in the random assignment problem. *Probab. Theory Related Fields* **93** 507–534. MR1183889

[2] ALDOUS, D. and LYONS, R. (2007). Processes on unimodular random networks. *Electron. J. Probab.* **12** 1454–1508. MR2354165

[3] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 1–72. Springer, Berlin. MR2023650

[4] ALDOUS, D. J. (2001). The $\zeta(2)$ limit in the random assignment problem. *Random Structures Algorithms* **18** 381–418. MR1839499

[5] ALDOUS, D. J. and BANDYOPADHYAY, A. (2005). A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* **15** 1047–1110. MR2134098

[6] BANDYOPADHYAY, A. (2011). Endogeny for the logistic recursive distributional equation. *Z. Anal. Anwend.* **30** 237–251. MR2793003

[7] BAYATI, M., SHAH, D. and SHARMA, M. (2008). Max-product for maximum weight matching: Convergence, correctness, and LP duality. *IEEE Trans. Inform. Theory* **54** 1241–1251. MR2445062

[8] FRIEZE, A. M. (1985). On the value of a random minimum spanning tree problem. *Discrete Appl. Math.* **10** 47–56. MR0770868

[9] HAJEK, B. (1990). Performance of global load balancing by local adjustment. *IEEE Trans. Inform. Theory* **36** 1398–1414. MR1080823

[10] HESSLER, M. and WÄSTLUND, J. (2010). Edge cover and polymatroid flow problems. *Electron. J. Probab.* **15** 2200–2219. MR2748403

[11] MÉZARD, M. and PARISI, G. (1987). On the solution of the random link matching problem. *J. Physique* **48** 1451–1459.

[12] PADÓ, S. and LAPATA, M. (2006). Optimal constituent alignment with edge covers for semantic projection. In *Proceedings of the* 21*st International Conference on Computational Linguistics and the* 44*th Annual Meeting of the Association for Computational Linguistics* 1161–1168. Association for Computational Linguistics, Stroudsburg, PA.

[13] PADÓ, S. and LAPATA, M. (2009). Cross-lingual annotation projection for semantic roles. *J. Artificial Intelligence Res.* **36** 307–340.

[14] SALEZ, J. and SHAH, D. (2009). Belief propagation: An asymptotically optimal algorithm for the random assignment problem. *Math. Oper. Res.* **34** 468–480. MR2554069

[15] SCHRIJVER, A. (2003). *Combinatorial Optimization. Polyhedra and Efficiency. Vol. A. Algorithms and Combinatorics* **24**. *Paths*, *Flows*, *Matchings*. Springer, Berlin. MR1956924

[16] WÄSTLUND, J. (2009). Replica symmetry and combinatorial optimization. Available at arXiv:0908.1920 [math.PR].

[17] WÄSTLUND, J. (2010). The mean field traveling salesman and related problems. *Acta Math.* **204** 91–150. MR2600434

[18] WÄSTLUND, J. (2012). Replica symmetry of the minimum matching. *Ann. of Math.* (2) **175** 1061–1091. MR2912702

[19] WIKIPEDIA (2012). Semantic role labeling. 22 March 2012. Available at http://en.wikipedia.org/wiki/Semantic_role_labeling.

DEPARTMENT OF ELECTRICAL COMMUNICATION ENGINEERING
INDIAN INSTITUTE OF SCIENCE
BANGALORE 560012
INDIA
E-MAIL: mustafa@ece.iisc.ernet.in
rajeshs@ece.iisc.ernet.in