Belief propagation: an asymptotically optimal algorithm for the random assignment problem

Justin Salez
École Normale Supérieure de Paris
email: justin.salez@ens.fr

Devavrat Shah
EECS, Massachusetts Institute of Technology
email: devavrat@mit.edu

The random assignment problem concerns finding the minimum cost assignment or matching in a complete bipartite graph with edge weights being i.i.d. with some distribution, say exponential(1) distribution. In a remarkable result by Aldous (2001), it was shown that the average cost of such an assignment converges to $\zeta(2) = \pi^2/6$ as the size of bipartite graph increases to ∞; thus proving conjecture of Mézard and Parisi (1987) based on replica method arising from statistical physics insights. This conjecture also suggested a heuristic for finding such an assignment, which is an instance of the well-known heuristic Belief Propagation (BP) discussed by Pearl (1987). In a recent work by Bayati, Shah and Sharma (2005), BP was shown to find correct solution in $O(n^3)$ time for the instance of assignment problem over graph of size n with arbitrary weights. In contrast, in this paper we establish that the BP finds an asymptotically correct assignment in $O(n^2)$ time with high probability for the random assignment problem for a large class of edge weight distributions. Thus, BP is essentially an optimal algorithm for the assignment problem under random setup.

Our result utilizes result of Aldous (2001) and the notion of local weak convergence. Key non-trivial steps in establishing our result involve proving attractiveness (aka decay of correlation) of an operator acting on space of distributions corresponding to the min-cost matching on Poisson Weighted Infinite Tree (PWIT) and establishing uniform convergence of dynamics of BP on bipartite graph to an appropriately defined dynamics on PWIT.

Key words: Belief propagation; random assignment problem; local weak convergence; correlation decay; Poisson weighted infinite tree.

MSC2000 Subject Classification: Primary: 60C05, 68T20, 68W40; Secondary: 05C80, 82B44.

OR/MS subject classification: Primary: Network/graphs matchings, Probability stochastic model; Secondary: Mathematics combinatorics and fixed points.

1. Introduction

1.1 Background. Consider an $n \times n$ complete bipartite graph $G = (V = V_1 \cup V_2, E)$, with $|V_1| = |V_2| = n$ and $E = \{(i, j) : i \in V_1, j \in V_2\}$. An edge $(i, j) \in E$ is assigned non-negative cost $(X_{i,j})$. The assignment problem consists of determining a permutation (or matching) π of $\{1, \ldots, n\}$ so that its total cost, $\sum_{i=1}^n X_{i,\pi(i)}$ is minimized. This is equivalent to finding a minimum-weight perfect matching in G. Recall that, a perfect matching on a graph $G = (V_1 \cup V_2, E)$ is a subset $M \subseteq E$ of pairwise disjoint edges that cover all vertices in $V_1 \cup V_2$. In what follows, we will be interested in the random assignment problem where edge weight $X_{i,j}$ are i.i.d. with some distribution, say exponential(1) or uniform on $[0, 1]$. Such randomly weight $n \times n$ bipartite graph will be denoted as $\mathcal{K}_{n,n}$ and the minimum cost assignment will be denoted by $\pi^*_{\mathcal{K}_{n,n}}$. The goal is to find $\pi^*_{\mathcal{K}_{n,n}}$ as quickly as possible. As the main result of this paper, we establish that the BP algorithm finds asymptotically (in n) correct $\pi^*_{\mathcal{K}_{n,n}}$ with minimal possible computation cost of $O(n^2)$.

The assignment problem, though seems cunningly simple has led to rich development in combinatorial probability and algorithm design since early 1960s. To understand mathematical properties of $\pi^*_{\mathcal{K}_{n,n}}$, partly motivated to obtain insights for better algorithm design, the question of finding asymptotic limit of the average cost of $\pi^*_{\mathcal{K}_{n,n}}$ became of great interest (see [19, 9, 12, 13, 17, 11, 8]). In 1987, Mézard and Parisi [15] through replica method based calculations conjectured that

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^n X_{i,\pi^*_{\mathcal{K}_{n,n}}(i)}\right] = \zeta(2).$$

More than a decade later, in 2001 this was rigorously established by Aldous [2]. This work by Aldous led to the formalism of "the objective method" (see survey by Aldous and Steele [4]). The finite version of the above conjecture (for exponential weight distribution),

$$\mathbb{E}\left[\sum_{i=1}^{n} X_{i,\pi^*_{\mathcal{K}_{n,n}}(i)}\right] = \sum_{i=1}^{n} \frac{1}{i^2},$$

was independently established by Nair, Prabhakar and Sharma [16] and Linusson and Wåstlund [14] in 2003.

On the algorithmic aspect, the consideration of assignment problem laid foundations for network flow algorithms. Specifically, the best known (strongly polynomial time) algorithm is by Edmonds and Karp [10] that takes $O(n^3)$ operations to find minimum cost matching for arbitrary instance. The statistical physics based approach (cavity method) suggested a heuristic for finding the min. cost matching for the random instance of the problem. This is an instance of the Belief Propagation(BP) heuristic that is popular in the artificial intelligence (see, book by Pearl [18] and work by Yedidia, Freeman and Weiss [20]). In a recent work, one of the author of this paper, Shah along with Bayati and Sharma [6] studied this heuristic for the maximum instead of minimum version of the above problem (i.e. find matching with the maximum cost). They established the correctness of the heuristic for arbitrary (not only random) instance as long as the maximum cost matching is unique. They showed that the algorithm takes $O(n^3)$ operations to find the solution with constant dependent on the $\max_{i,j} X_{i,j}$ and difference of weight between the maximum and second maximum weight matching.

The BP algorithm seem to have much better empirical performance than the worst case result established in [6]. Motivated by this, here we consider the question of analyzing BP for random assignment problem. Specifically, we will establish that the BP finds almost optimal assignment within constant iterations for random instance of problem with high probability. Thus, the total computation cost of the algorithm for random instance scales as $O(n^2)$ to find almost optimal assignment. This is in sharp contrast to the best known (adversarial) bound of $O(n^3)$. Clearly, no algorithm can perform better than $\Omega(n^2)$. That is, BP is essentially optimal for random instance of the problem.

Remark 1 *There is a lot of work on analyzing performance of various heuristics for finding (variants of) assignment problem under various restrictions. Since the literature is vast on this topic and it is not possible to recall all the known results and hence we refrain from listing all of them.*

1.2 BP algorithm. We describe the BP algorithm for finding minimum cost matching in arbitrary graph $G = (V, E)$. It naturally applies to the bipartite graph. We will use the BP algorithm for general graph and its specialized form for bipartite graph interchangeably in this paper; the specific use will be clear from the context. For graph G, we use notation that the cost or length of an $e = (u, v) \in E$ is $\|e\|_G > 0$ or $\|u, v\|_G$. By $w \sim v$, we denote that w is a neighbor of v in G.

The BP algorithm is a distributed and iterative: in each iteration, it involves sending a message (real number) in both direction along each edge of the graph. Specifically, in iteration $k \geq 0$ every vertex $v \in V$ sends a message $\langle v \to w\rangle_G^k$ to each of its neighbor $w \sim v$. These messages are calculated as follows:

$$\begin{aligned} \langle v \to w\rangle_G^0 &:= 0; \\ \langle v \to w\rangle_G^k &:= \min_{u \sim v, u \neq w} \left\{ \left\|u, v\right\|_G - \langle u \to v\rangle_G^{k-1} \right\}, \quad \forall\, k \geq 1. \end{aligned} \tag{1}$$

Every vertex $v \in V$ estimates the minimum cost matching, $\pi_G^k : V \to V$ as follows:

$$\pi_G^k(v) := \operatorname*{arg\,min}_{w \sim v} \left\{ \left\|v, w\right\|_G - \langle w \to v\rangle_G^k \right\}. \tag{2}$$

Note that when the above algorithm is specialized to a bipartite graph, nodes in each partition attempts to compute a permutation; if algorithm's estimate converges to the right answer then the permutation estimate of both the partitions much converge to the same answer. For that reason, in the context of bipartite graph $\mathcal{K}_{n,n}$, by $\pi^k_{\mathcal{K}_{n,n}}$ we denote the estimated assignment of nodes in both the partition and now onwards we shall abuse the notation $\pi^*_{\mathcal{K}_{n,n}}$ for the optimal assignment of nodes in both partitions as well. We define fraction-difference between the estimated

$$d\big(\pi^k_{\mathcal{K}_{n,n}}, \pi^*_{\mathcal{K}_{n,n}}\big) := \frac{1}{2n}\,\mathrm{card}\Big\{x \in \mathcal{K}_{n,n}, \pi^k_{\mathcal{K}_{n,n}}(x) \neq \pi^*_{\mathcal{K}_{n,n}}(x)\Big\}.$$

It should be noted that each iteration of the algorithm requires $O(n^2)$ operations for $G = \mathcal{K}_{n,n}$. Therefore, if algorithm finds good estimate (i.e. $d\big(\pi^k_{\mathcal{K}_{n,n}}, \pi^*_{\mathcal{K}_{n,n}}\big)$ small) after constant number of iterations, then the overall cost is $O(n^2)$.

1.3 Result. We analyze the running time of BP for random assignment problem. We consider stochastic model with edge weights being i.i.d. with cumulative distribution function represented by H. Observe that the continuity of H will be sufficient for any two distinct perfect matchings have a.s. distinct weights, allowing us to consider *the* minimal one $\pi^*_{\mathcal{K}_{n,n}}$. We establish that for a large class of distribution H, the algorithm finds asymptotically correct solution in finite number of iterations with high probability. Formal statement of our result is as follows:

Theorem 1 *Let the cumulative distribution function of edge weights, H satisfy the following:*

A1. Regularity at 0: $H(0) = 0$, H is right differentiable at 0, $H'(0) \neq 0$;

A2. Light-tail property : $1 - H(t) = O\left(e^{-\beta t}\right)$ as $t \to \infty$ for some $\beta > 0$.

Then,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{E}\Big[d\big(\pi^k_{\mathcal{K}_{n,n}}, \pi^*_{\mathcal{K}_{n,n}}\big)\Big] = 0. \tag{3}$$

In establishing this result, we strongly utilize the frame-work of local weak convergence developed by Aldous [2]. As a reader will find, the above result is far from being an implication of the result by Aldous. Specifically, to establish the above result we need to take two non-trivial steps: (1) Establishing that the dynamics of Belief Propagation on $\mathcal{K}_{n,n}$ converges to dynamics of BP on an appropriately defined limiting object; and (2) characterizing attractiveness of an operator related to BP. It is worth noting that (2) was left as an open problem by Aldous and Bandopadhay [3, Open Problem # 62].

1.4 Implication. Theorem 1 implies that for any $\varepsilon > 0$, there exists $n(\varepsilon), k(\varepsilon)$ such that for all $n \geq n(\varepsilon)$, the BP algorithm finds correct assignment to $1-\varepsilon$ fraction of nodes in $k(\varepsilon)$ iterations with probability at least $1 - \varepsilon$. Thus, total computation cost is $O(n^2)$, constant depending on ε, for finding good approximation. This result applies for large class of weight distribution function including uniform over $[0, 1]$ or exponential distribution.

1.5 Organization. The remaining paper is dedicated to proving Theorem 1. Our result utilizes the machinery of local weak convergence introduced by Aldous. The Figure 1.5 illustrates the three main steps of the proof of Theorem 1 – which corresponds to establishing the top horizontal arrow.

1. First, we show that BP's behavior on $\mathcal{K}_{n,n}$ "converges" to its behavior on the so-called Poisson Weighted Infinite Tree (PWIT) $\mathcal{T}$ – corresponding to the left vertical arrow in the Figure 1.5 and formally stated as Theorem 3. This is done in Section 3.
2. Second, we establish strong convergence of the recursive distributional tree process corresponding to BP's execution on $\mathcal{T}$ – corresponding to the bottom horizontal arrow in Figure 1.5 and summarized as Theorem 5. We note that the Theorem 5 resolves an open problem stated in [2, 3] related to the assignment problem, which was necessary for establishing convergence of algorithm. However, it was not necessary for evaluating the limiting expected cost as $\zeta(2)$. This is done in Section 4.
3. Third, the connection between the limiting estimate of BP and the optimal solution on $\mathcal{K}_{n,n}$, $\pi^*_{\mathcal{K}_{n,n}}$ is provided by the work by Aldous [2] – corresponding to the vertical right arrow and stated as Theorem 2 and Theorem 6. This is used in Section 5 to complete the proof of Theorem 1.

2. Preliminaries

In this section, we describe the necessary background on the notion of local weak convergence introduced by Aldous [2]. Consider a rooted, edge-weighted and connected graph $G = (V, \varnothing, E)$, where V is the set of vertices, E the set of edges and $\varnothing \in V$ represents it's root. Define distance between any two vertices of the graph as the infimum over weights of all paths connecting them.

$$\begin{tikzcd}
\pi^k_{\mathcal{K}_{n,n}} \arrow[r, dotted, "?", "k\to\infty"'] \arrow[d, "n\to\infty"'] & \pi^*_{\mathcal{K}_{n,n}} \arrow[d, "n\to\infty"] \\
\pi^k_{\mathcal{T}} \arrow[r, "k\to\infty"'] & \pi^*_{\mathcal{T}}
\end{tikzcd}$$

Figure 1: Theorem 1 corresponds to establishing top-horizontal arrow; which is done by establishing the other three arrows in the above diagram.

Define the $\varrho-$restriction of G as the sub-graph $\lceil G \rceil_\varrho$ of G obtained by deleting all vertices that are at distance more than ϱ from the root of G. Such a rooted, edge-weighted and connected graph G is called *geometric graph* if it's $\varrho-$restrictions are finite for any finite $\varrho > 0$. Now, we define notion of local weak convergence for geometric graphs.

Definition 1 (local convergence of geometric graphs) *A sequence $(G_n)_{n\geq 1}$ of geometric graphs is said to converge locally to a geometric graph G if for any $\varrho > 0$ such that no vertex in G is at distance exactly ϱ from the root, the following is satisfied:*

1. *There exists $n_\varrho \in \mathbb{N}^*$ such that all the $\lceil G_n \rceil_\varrho, n \geq n_\varrho$ are isomorphic[1] to $\lceil G \rceil_\varrho$;*
2. *The corresponding isomorphisms $\gamma_n^\varrho \colon \lceil G \rceil_\varrho \rightleftharpoons \lceil G_n \rceil_\varrho, n \geq n_\varrho$ can be chosen so that for any edge e in $\lceil G \rceil_\varrho$:*
$$\left\|\gamma_n^\varrho(e)\right\|_{G_n} \xrightarrow[n\to\infty]{} \left\|e\right\|_G.$$

Consider the following mapping between pair of geometric graphs, G, G':

$$G, G' \mapsto \int_0^\infty e^{-\varrho}\left(1 \wedge \inf_{\gamma\colon \lceil G\rceil_\varrho \rightleftharpoons \lceil G'\rceil_\varrho} \max_{e\in\lceil G\rceil_\varrho} \left|\ln \frac{\|\gamma(e)\|_{G'}}{\|e\|_G}\right|\right) d\varrho \tag{4}$$

With little work, it can be shown that the above mapping defines a metric on $\mathcal{G}^*$, the space of geometric graphs. Further, the above definition of local weak convergence is equivalent to the convergence with respect to the above defined metric. Thus resulting metric space is complete and separable. As a consequence, we can import machinery related to the weak convergence of distribution on Polish space. Specifically, we will often use the Skorohod's representation theorem (see, [7, Theorem 6.7]): it essentially allows one to assume (due to existence of appropriate joint probability space) almost sure convergence when there is a distribution or local weak convergence.

Next, we recall result by Aldous that showed that $\mathcal{K}_{n,n}$ convergences to the Poisson Weighted Infinite Tree under the topology of local weak convergence. Before we state the result, we will need some notation that will be useful throughput the paper. Let $\mathcal{V}$ denote the set of all finite words over the alphabet $\mathbb{N}^*$, $\varnothing$ the empty word, "$\cdot$" the usual concatenation operation on $\mathcal{V}$ and for any $v \in \mathcal{V}^* := \mathcal{V}\setminus\{\varnothing\}$, $\dot{v}$ the word obtained from v by simply deleting the last letter. Set also $\mathcal{E} := \{\{v, v.i\}, v \in \mathcal{V}, i \geq 1\}$. The graph $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denotes an infinite tree with $\varnothing$ as root, all words of length 1 as the nodes at depth 1, words of length 2 as the nodes at depth 2, and so on.

Theorem 2 (Convergence to PWIT[1, 2]) *Given a collection $(\xi^v = \xi_1^v, \xi_2^v \ldots)_{v\in\mathcal{V}}$ of independent, ordered Poisson point processes with intensity 1 on $\mathbb{R}^+$, consider the infinite tree $\mathcal{T} := (\mathcal{V}, \mathcal{E})$ rooted at $\varnothing$ and with edge lengths $\left\|v, v.i\right\|_\mathcal{T} := \xi_i^v, v \in \mathcal{V}, i \geq 1$. This defines the law of a random rooted geometric graph called the Poisson Weighted Infinite Tree (PWIT). Under the assumption $A1$ on H, we have:*

$$nH'(0)\mathcal{K}_{n,n} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{T}, \tag{5}$$

with respect to the topology of local weak convergence.

[1] An isomorphism from $G = (V, \varnothing, E)$ to $G' = (V', \varnothing', E')$, denoted $\gamma\colon G \rightleftharpoons G'$, is simply a bijection from V to V' preserving the root $\gamma(\varnothing) = \varnothing'$) and the structure $(\forall(x, y) \in V, \{\gamma(x), \gamma(y)\} \in E' \Leftrightarrow \{x, y\} \in E)$.

For simplicity of notation, we will get rid of the scaling factor $nH'(0)$ from the remaining of the paper by the following transformation: the edge lengths in $\mathcal{K}_{n,n}$ as distributed according to $H\left(\frac{\cdot}{nH'(0)}\right)$ instead of H. Note that, under this scaling (transformation), both the optimal matching $\pi^*_{\mathcal{K}_{n,n}}$ and the BP computatoin $\pi^k_{\mathcal{K}_{n,n}}, k \geq 0$, remain invariant.

3. First step: convergence of dynamics of BP

The goal of this section is to deduce from Theorem 2 that the behavior of BP when running on $\mathcal{K}_{n,n}$ *converges* as $n \to \infty$ to its behavior when running on $\mathcal{T}$. In order to make this notion precise, we first re-label the vertices of $\mathcal{K}_{n,n}$ by words of $\mathcal{V}$. We would like this re-labeling done in a manner that yields to consistent comparison between the messages of BP on $\mathcal{K}_{n,n}$ and those on $\mathcal{T}$. Such a re-labeling is explained next. To begin with, the empty word $\varnothing$ will represent a fixed root of $\mathcal{K}_{n,n}$ (alternatively, one may choose one of the vertex uniformly at random; but symmetry of the problem makes this distributionally equivalent to the choice of fixed root node). Assign words $1, 2, \cdots, n$ to the n immediate neighbors of the chosen root ordered as per their edge weights (lengths) in an increasing manner. Now inductively, if word $v \in \mathcal{V}^*$ represents some vertex $x \in \mathcal{K}_{n,n}$ and $\dot{v}$ some $y \in \mathcal{K}_{n,n}$, then let the words $v.1, v.2, \cdots, v.(n-1)$ represent the $n-1$ neighbors of x distinct from y in $\mathcal{K}_{n,n}$, again ordered by edge weights (lenghts) in an increasing manner. Note that this definition makes almost surely since the edge lengths are pairwise distinct by continuity of H. Note also that any $\{u, v\} \in \mathcal{E}$ represents an edge of $\mathcal{K}_{n,n}$ for large enough n. Now, we state and prove the main result of this section. In what follows, consider $\mathcal{K}_{n,n}$ converging to $\mathcal{T}$ almost surely due to Theorem 2 and Skorohod's representation theorem.

Theorem 3 (Continuity of BP) *Consider BP operating on $\mathcal{K}_{n,n}$, where $\mathcal{K}_{n,n}$ converges to $\mathcal{T}$ almost surely. Then, for all $k \geq 0$, the k^{th} step messages of BP on $\mathcal{K}_{n,n}$ converge to those on $\mathcal{T}$ in probability. That is,*

$$\forall v \in \mathcal{V}^*, \langle v \to \dot{v} \rangle^k_{\mathcal{K}_{n,n}} \xrightarrow[n\to\infty]{proba} \langle v \to \dot{v} \rangle^k_{\mathcal{T}}. \tag{6}$$

Further, the estimates at the root nodes converge in probability:

$$\mathbb{P}\Big(\pi^k_{\mathcal{K}_{n,n}}(\varnothing) \neq \pi^k_{\mathcal{T}}(\varnothing)\Big) \xrightarrow[n\to\infty]{} 0. \tag{7}$$

Proof. We will prove (6) by induction over $k \geq 0$. The base case of $k = 0$ is trivial. Now suppose the convergence holds for some $k \geq 0$ and fix $v \neq \varnothing$. We need to show

$$\langle v \to \dot{v} \rangle^{k+1}_{\mathcal{K}_{n,n}} \xrightarrow[n\to\infty]{proba} \langle v \to \dot{v} \rangle^{k+1}_{\mathcal{T}}. \tag{8}$$

Equivalently,

$$\min_{1\leq i<n} \Big\{ \big\| v, v.i \big\|_{\mathcal{K}_{n,n}} - \langle v.i \to v \rangle^k_{\mathcal{K}_{n,n}} \Big\} \xrightarrow[n\to\infty]{proba} \min_{i\geq 1} \Big\{ \big\| v, v.i \big\|_{\mathcal{T}} - \langle v.i \to v \rangle^k_{\mathcal{T}} \Big\}. \tag{9}$$

Note that for every fixed $i \geq 1$, $\big\| v, v.i \big\|_{\mathcal{K}_{n,n}} \xrightarrow[n\to\infty]{a.s.} \big\| v, v.i \big\|_{\mathcal{T}}$ since we have $\mathcal{K}_{n,n}$ convering to $\mathcal{T}$ a.s.. Hence, using the induction hypothesis it follows that

$$\big\| v, v.i \big\|_{\mathcal{K}_{n,n}} - \langle v.i \to v \rangle^k_{\mathcal{K}_{n,n}} \xrightarrow[n\to\infty]{proba} \big\| v, v.i \big\|_{\mathcal{T}} - \langle v.i \to v \rangle^k_{\mathcal{T}}.$$

It is temping to complete the proof here by adding "minimum over i" in the above. However, it needs additional justification as we have infinite number of terms. In order to complete the proof, we will show that only a large finite number of terms matter for minimization in the following sense :

$$\lim_{i_0\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\arg\min_{1\leq i<n} \Big\{ \big\| v, v.i \big\|_{\mathcal{K}_{n,n}} - \langle v.i \to v \rangle^k_{\mathcal{K}_{n,n}} \Big\} \geq i_0 \right) = 0. \tag{10}$$

Given (10), the desired conclusion of (8) will be obtained as follows: for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\Big| \langle v \to \dot{v} \rangle^{k+1}_{\mathcal{K}_{n,n}} - \langle v \to \dot{v} \rangle^{k+1}_{\mathcal{T}} \Big| \geq \varepsilon \Big) &\leq \mathbb{P}\left(\arg\min_{1\leq i<n} \Big\{ \big\| v, v.i \big\|_{\mathcal{K}_{n,n}} - \langle v.i \to v \rangle^k_{\mathcal{K}_{n,n}} \Big\} \geq i_0 \right) \\
&+ \quad \mathbb{P}\left(\arg\min_{i\geq 1} \Big\{ \big\| v, v.i \big\|_{\mathcal{T}} - \langle v.i \to v \rangle^k_{\mathcal{T}} \Big\} \geq i_0 \right) \\
&+ \quad \mathbb{P}\left(\Big| \min_{1\leq i<i_0} \Big\{ \big\| v, v.i \big\|_{\mathcal{K}_{n,n}} - \langle v.i \to v \rangle^k_{\mathcal{K}_{n,n}} \Big\} - \min_{1\leq i<i_0} \Big\{ \big\| v, v.i \big\|_{\mathcal{T}} - \langle v.i \to v \rangle^k_{\mathcal{T}} \Big\} \Big| \geq \varepsilon \right).
\end{aligned}$$

Now, taking lim sup with respect to $n \to \infty$ and then taking limit as $i_0 \to \infty$ yields the desired result. Here, we have used the fact that $\arg\min_{i\geq 1}\left\{\left\|v, v.i\right\|_{\mathcal{T}} - \langle v.i \to v\rangle^k_{\mathcal{T}}\right\}$ is a well-defined finite r.v. with probability 1. This will become obvious in the next section. The proof of (10) will follow from Lemma 3 stated and proved later in this section. Finally, the convergence claimed in (7) follows using similar arguments as explained in the following sequence of inequalities:

$$\begin{aligned}\mathbb{P}\Big(\pi^k_{\mathcal{K}_{n,n}}(\varnothing) \neq \pi^k_{\mathcal{T}}(\varnothing)\Big) &\leq \mathbb{P}\left(\arg\min_{1\leq i<n}\left\{\left\|\varnothing, i\right\|_{\mathcal{K}_{n,n}} - \langle i \to \varnothing\rangle^k_{\mathcal{K}_{n,n}}\right\} \geq i_0\right)\\ &+ \mathbb{P}\left(\arg\min_{i\geq 1}\left\{\left\|\varnothing, i\right\|_{\mathcal{T}} - \langle i \to \varnothing\rangle^k_{\mathcal{T}}\right\} \geq i_0\right)\\ &+ \mathbb{P}\left(\arg\min_{1\leq i<i_0}\left\{\left\|\varnothing, i\right\|_{\mathcal{K}_{n,n}} - \langle i \to \varnothing\rangle^k_{\mathcal{K}_{n,n}}\right\} \neq \arg\min_{1\leq i<i_0}\left\{\left\|\varnothing, i\right\|_{\mathcal{T}} - \langle i \to \varnothing\rangle^k_{\mathcal{T}}\right\}\right).\end{aligned}$$

This completes the proof of Theorem 3. □

It now remains to prove assertion (10) which is stated as Lemma 3. In order to prove it, we will need two Lemmas 1 and 2 stated below will provide the desired uniform controls over edge-lengths and messages under re-labeling. In these two Lemmas, we will use the following notation: for a word $v \in \mathcal{V}$, by $|v|$ we denote number of letters in it (e.g. for $v = \varnothing$, $|v| = 0$); its letters will be represented as $v = (v_1, \ldots, v_{|v|})$ (e.g. $v = 1.2.1.3$ then $|v| = 4$ and $v_1 = 1, v_2 = 2, v_3 = 1, v_4 = 3$). For $0 \leq h \leq |v|$, we will write $v_{\leq h}$ for the prefix $v_1 \cdots v_h$.

Lemma 1 (Uniform control on edge-lengths) *There exist constants $(M_h)_{h\geq 1}$, α and $\beta > 0$ such that for all $v \in \mathcal{V}, i \geq 1, t \in \mathbb{R}^+$ and n large enough so that for $v.i \in \mathcal{K}_{n,n}$,*

$$\mathbb{P}\Big(\left\|v, v.i\right\|_{\mathcal{K}_{n,n}} \leq t\Big) \leq M_{|v|}\frac{(\alpha t)^i}{i!}e^{\alpha t} \quad \text{and} \quad \mathbb{P}\Big(\left\|v, v.1\right\|_{\mathcal{K}_{n,n}} \geq t\Big) \leq M_{|v|}e^{-\beta t}.$$

Proof. Suppose $\|v, v.i\|_{\mathcal{K}_{n,n}} \leq t$. Then by construction, the sequence of words $(v_{\leq 0}, \ldots, v_{\leq |v|})$ represents a path in $\mathcal{K}_{n,n}$ starting from the root and ending at a vertex from which at least i incident edges have length at most t (property of re-labeling). Following down this path while deleting every cycle, we obtain a cycle-free path $x = (x_0, \ldots, x_k)$ $(0 \leq k \leq |v| \wedge 2n-1)$ starting from the root satisfying

$$\operatorname{card}\Big\{y \neq x_{k-1}, \left\|x_k, y\right\|_{\mathcal{K}_{n,n}} \leq t\Big\} \geq i-1. \tag{11}$$

For $0 \leq j < k$, (x_j, x_{j+1}) corresponds to some $(v_{\leq p-1}, v_{\leq p})$, $1 \leq p \leq |v|$. By construction the number of edges that are incident on $v_{\leq p-1}$ and shorter than $\{v_{\leq p-1}, v_{\leq p}\}$ is precisely $v_p - 1$ or v_p, depending on the parent-edge. Therefore, there exists $p \in \{1, \ldots, |v|\}$ such that

$$v_p - \left\lceil \frac{k}{2} \right\rceil \leq \operatorname{card}\Big\{y \notin \{x_1, \ldots, x_k\}, \left\|x_j, y\right\|_{\mathcal{K}_{n,n}} < \left\|x_j, x_{j+1}\right\|_{\mathcal{K}_{n,n}}\Big\} \leq v_p. \tag{12}$$

In above, we used the fact that only half of the $x_1, \ldots, x_k$ are neighbors of x_j, and hence the bound of $\lceil \frac{k}{2} \rceil$. Thus, we have shown that

$$\mathbb{P}\Big(\left\|v, v.i\right\|_{\mathcal{K}_{n,n}} \leq t\Big) \leq \sum_{k=0}^{|v|}\sum_{x=(x_0,\ldots x_k)}\mathbb{P}\left(A_{n,x} \cap \bigcap_{j=0}^{k-1} B^j_{n,x}\right),$$

where event $A_{n,x}$ corresponds to (11) and $B^j_{n,x}$ correspond to (12) for $0 \leq j < k$. The summation in the above inequality is over all possible cycle-free paths $x = (x_0, ...x_k)$ starting from the root in $\mathcal{K}_{n,n}$. Now since all the edges involved are pairwise distinct, the events $A_{n,x}, B^0_{n,x}, ..., B^{k-1}_{n,x}$ are independent. Therefore, their probabilities can be computed as folllows:

$$\begin{aligned}\mathbb{P}\left(B^j_{n,x}\right) &= \sum_{p=1}^{|v|}\sum_{q=v_p-\lceil\frac{k}{2}\rceil}^{v_p}\frac{1}{n+1-\lceil\frac{k}{2}\rceil} \leq \frac{(|v|+1)^3}{n};\\ \mathbb{P}\left(A_{n,x}\right) &= \sum_{q=i-1}^{n-1}\binom{n-1}{q}H\left(\frac{t}{nH'(0)}\right)^q\left(1-H\left(\frac{t}{nH'(0)}\right)\right)^{n-1-q} \leq \frac{(\alpha t)^i}{i!}e^{\alpha t},\end{aligned}$$

where we have used assumption A1 and the following notation:

$$\alpha := \frac{1}{H'(0)}\sup_{\varrho\in\mathbb{R}^+}\frac{H(\varrho)}{\varrho} < +\infty.$$

This yields the first uniform bound since the number of cycle-free paths $x = (x_0, ..., x_k)$ starting from the root in $\mathcal{K}_{n,n}$ is clearly bounded above by n^k. For the second one, the event $A_{n,x}$ is simply replaced by $\mathrm{card}\Big\{y \neq x_{k-1}, \|x_k, y\|_{\mathcal{K}_{n,n}} \leq t\Big\} \leq 1$, whose probability is straightforwardly exponentially bounded using assumption A2. This completes the proof of Lemma 1. □

Lemma 2 (Uniform control on messages) *There exist constants $(M_{k,h}, \beta_{k,h})_{k,h\geq 0} > 0$ such that for all $v \in \mathcal{V}^*$, and for all $t \in \mathbb{R}^+$,*

$$\mathbb{P}\Big(\Big|\langle v \to \dot{v}\rangle^k_{\mathcal{K}_{n,n}}\Big| \geq t\Big) \leq M_{k,|v|}e^{-\beta_{k,|v|}t}, \tag{13}$$

uniformly in n, as long as n is large enough so that $v \in \mathcal{K}_{n,n}$.

Proof. The proof is by induction over k. The base case of $k = 0$ follows trivially. Now, as per induction hypothesis for a given $k \in \mathbb{N}$ suppose (13) is true. By Lemma 1 we can write for all $v \in \mathcal{V}^*$ and $t \in \mathbb{R}^+$:

$$\begin{aligned}\mathbb{P}\Big(\langle v \to \dot{v}\rangle^{k+1}_{\mathcal{K}_{n,n}} \geq t\Big) &= \mathbb{P}\Big(\min_{1\leq i<n}\big\{\|v, v.i\|_{\mathcal{K}_{n,n}} - \langle v.i \to v\rangle^k_{\mathcal{K}_{n,n}}\big\} \geq t\Big)\\ &\leq \mathbb{P}\Big(\|v, v.1\|_{\mathcal{K}_{n,n}} \geq \frac{t}{2}\Big) + \mathbb{P}\Big(\langle v.1 \to v\rangle^k_{\mathcal{K}_{n,n}} \leq -\frac{t}{2}\Big)\\ &\leq M_{|v|}e^{-\frac{\beta}{2}t} + M_{k,|v|+1}e^{-\frac{\beta_{k,|v|+1}}{2}t}.\end{aligned}$$

The other side is harder to obtain which we do next. For this, again by Lemma 1 :

$$\begin{aligned}\mathbb{P}\Big(\langle v \to \dot{v}\rangle^{k+1}_{\mathcal{K}_{n,n}} \leq -t\Big) &= \mathbb{P}\Big(\min_{1\leq i<n}\big\{\|v, v.i\|_{\mathcal{K}_{n,n}} - \langle v.i \to v\rangle^k_{\mathcal{K}_{n,n}}\big\} \leq -t\Big)\\ &\leq \sum_{i=1}^{n-1}\mathbb{P}\Big(\|v, v.i\|_{\mathcal{K}_{n,n}} \leq r_i(t)\Big) + \sum_{i=1}^{n-1}\mathbb{P}\Big(\langle v.i \to v\rangle^k_{\mathcal{K}_{n,n}} \geq t + r_i(t)\Big),\\ &\leq M_{|v|}\sum_{i=1}^{\infty}\frac{\big(\alpha r_i(t)\big)^i e^{\alpha r_i(t)}}{i!} + M_{k,|v|+1}\sum_{i=1}^{\infty}e^{-\beta_{k,|v|+1}(t+r_i(t))},\end{aligned}$$

where the inequalities hold for any choice of the quantities $r_i(t) \geq 0$. Our proof thus boils down to the following simple question: can we choose the $r_i(t)$ such that

1. $r_i(t)$ is large enough to ensure exponential vanishing of $\sum_{i=1}^{\infty} e^{-\beta_{k,|v|+1}(t+r_i(t))}$;
2. $r_i(t)$ is small enough to ensure exponential vanishing of $\sum_{i=1}^{\infty}\frac{\big(\alpha r_i(t)\big)^i e^{\alpha r_i(t)}}{i!}$.

The answer is yes. Indeed, taking $r_i(t) := \delta i e^{-\gamma t}$ with $\gamma, \delta > 0$ yields

$$\frac{1}{t}\log\sum_{i=1}^{\infty}e^{-\beta_{k,|v|+1}(t+r_i(t))} \xrightarrow[t\to+\infty]{} \gamma - \beta_{k,|v|+1}.$$

Therefore, selection of $\gamma < \beta_{k,|v|+1}$ is enough to ensure (1). As far as (2) is concerned, we have

$$\frac{1}{t}\log\sum_{i=1}^{\infty}\frac{\big(\alpha r_i(t)\big)^i e^{\alpha r_i(t)}}{i!} \leq -\gamma + \frac{1}{t}\log\sum_{i=1}^{\infty}\frac{\big(\alpha\delta e^{\alpha\delta}i\big)^i}{i!}.$$

The term in the infinite summation is equivalent to $\frac{(\alpha\delta e^{\alpha\delta-1})^i}{\sqrt{2\pi i}}$ by Stirling's formula. Hence the result follows as soon as δ is small enough for $\alpha\delta e^{\alpha\delta-1} < 1$. □

We now know enough to justify the crucial assertion we used in the proof of Theorem 3 :

Lemma 3 (Uniform control on essential messages) *For all $v \in \mathcal{V}$ and $k \geq 0$:*

$$\lim_{i_0\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\arg\min_{1\leq i<n}\Big\{\|v, v.i\|_{\mathcal{K}_{n,n}} - \langle v.i \to v\rangle^k_{\mathcal{K}_{n,n}}\Big\} \geq i_0\right) = 0.$$

Proof. Choose $\delta > 0$ small enough to ensure $\alpha\delta e^{\alpha\delta-1} < 1$ and use result of Lemmas 1 and 2 to conclude that for $t \in \mathbb{R}^+$, we have

$$\begin{aligned}
&\mathbb{P}\left(\operatorname*{arg\,min}_{1\le i<n}\left\{\left\|v,v.i\right\|_{\mathcal{K}_n} - \langle v.i \to v\rangle^k_{\mathcal{K}_{n,n}}\right\} \ge i_0\right) \\
\le\;& \mathbb{P}\left(\langle v \to \dot{v}\rangle^{k+1}_{\mathcal{K}_{n,n}} \ge t\right) + \sum_{i=i_0}^{n-1}\mathbb{P}\left(\left\|v,v.i\right\|_{\mathcal{K}_{n,n}} \le \delta i\right) + \sum_{i=i_0}^{n-1}\mathbb{P}\left(\langle v.i \to v\rangle^k_{\mathcal{K}_{n,n}} \ge \delta i - t\right) \\
\le\;& M_{k+1,|v|}e^{-\beta_{k+1,|v|}t} + M_{|v|}\sum_{i=i_0}^{\infty}\frac{(\alpha\delta e^{\alpha\delta} i)^i}{i!} + M_{k,|v|+1}\sum_{i=i_0}^{\infty}e^{-\beta_{k,|v|+1}(\delta i - t)}.
\end{aligned}$$

Letting $i_0 \to \infty$ and finally $t \to \infty$ yields the desired result. □

4. Second step: convergence of BP on PWIT

In the previous section, we established that the dynamics of BP on $\mathcal{K}_{n,n}$ converges to that on the limiting PWIT $\mathcal{T}$ as $n \to \infty$. This convergence happens with respect to the local weak convergence. Therefore, in order to understand the behavior of BP on $\mathcal{K}_{n,n}$ for large n, we need to study the dynamics of BP on the infinite random structure PWIT. That is, we are interested in understanding the following random message-process on $\mathcal{T}$: $\forall v \in \mathcal{V}^*$ and $k \ge 0$,

$$\langle v \to \dot{v}\rangle^{k+1}_{\mathcal{T}} \quad := \quad \min_{i\ge 1}\left\{\left\|v,v.i\right\|_{\mathcal{T}} - \langle v.i \to v\rangle^k_{\mathcal{T}}\right\}, \tag{14}$$

and initial messages $\left(\langle v \to \dot{v}\rangle^0_{\mathcal{T}}\right)_{v\in\mathcal{V}^*}$ are i.i.d. (in our algorithm, messages are initialized to 0).

First observe that at any given time k all $\langle v \to \dot{v}\rangle^k_{\mathcal{T}}, v \in \mathcal{V}^*$ share the same distribution, owing to the natural spatial invariance of the PWIT. Moreover, if F denotes the corresponding common anti-c.d.f.[2] at a given time, a straightforward computation (see for instance [2]) shows that the new anti-c.d.f. obtained after a single application of update rule (14) is :

$$TF : x \mapsto \exp\left(-\int_{-x}^{+\infty} F(t)\, dt\right). \tag{15}$$

This defines an operator T on the space $\mathcal{D}$ of anti-c.d.f.'s, (i.e. left-continuous non-increasing functions $F\colon \mathbb{R} \to [0,1]$). This operator T is known to have a unique fix-point (see [2]), which is the so-called *logistic* distribution:

$$F^* : x \mapsto \frac{1}{1+e^x}. \tag{16}$$

Our first step will naturally consist in studying the dynamics of T on $\mathcal{D}$.

4.1 Attractiveness of T: weak convergence on $\mathcal{T}$. Finding the domain of attraction of F^* under operator T is not known and has been listed as open problem by Aldous and Bandyopadhyay ([3, Open Problem # 62]). In what follows, we will answer this question and more. We will fully characterize the asymptotical behavior of the successive iterates $(T^kF)_{k\ge 0}$ for any initial distribution $F \in \mathcal{D}$. Due to matching constraints, the operator T is non-decreasing in the following sense:

$$F \le F' \text{ on all } \mathbb{R} \Rightarrow TF' \ge TF \text{ on all } \mathbb{R}. \tag{17}$$

This suggests considering the monotonic (non-decreasing) second iterate T^2. However, unlike T the second iterate T^2 admits an infinite number of fix-points. For this, let θ_t ($t \in \mathbb{R}$) be the shift operator defined on $\mathcal{D}$ by $\theta_t F\colon x \mapsto F(x-t)$. Then,

$$T \circ \theta_t = \theta_{-t} \circ T, \tag{18}$$

Therefore, it follows that $T^2(\theta_t F^*) = \theta_t(T^2F^*) = \theta_t F^*$ for all $t \in \mathbb{R}$. Thus, $\theta_t F^*$ is fixed points of T^2 for all $t \in \mathbb{R}$. On first instance, this fact may seem a bit disappointing. But as we shall see, these fixed-points will play a crucial role in our study of the operator T's iterates. First, a useful transformation.

Definition 2 *For $F \in \mathcal{D}$, define the transform $\widehat{F}$ as follows :*

$$\begin{array}{rccl}\widehat{F} & : & \mathbb{R} & \to & [-\infty,+\infty] \\ & & x & \mapsto & x + \ln\left(\frac{F(x)}{1-F(x)}\right).\end{array} \tag{19}$$

[2]The anti-c.d.f. of a real r.v. X is the function $F\colon x \to \mathbb{P}(X > x)$.

The reason behind considering this transform is the following straightforward fact.

Lemma 4 *For any given $F \in \mathcal{D}$ and $x \in \mathbb{R}$, $F(x) = \theta_{\widehat{F}(x)} F^*(x)$. Further, $F \equiv \theta_x F^*$ if and only if $\widehat{F}$ is constant on $\mathbb{R}$ with value x.*

The above Lemma suggests that the maximal amplitude of the variations of $\widehat{F}$ on $\mathbb{R}$ tells about the distance between F and the family of fix-points $\{\theta_t F^*, t \in \mathbb{R}\}$. Therefore, we will consider behavior of the maximal amplitude of F when T acts on F.

Lemma 5 *Let $F \in \mathcal{D} \setminus \{0\}$ be integrable at $+\infty$. Then, $\widehat{T^4 F}$ is uniformly bounded on $\mathbb{R}$.*

Proof. The non-increasing property of $F \colon \mathbb{R} \to [0,1]$ and $F \neq 0$ implies that there exists $\beta > 0$ and $x_0 \in \mathbb{R}$ such that

$$\forall x \leq x_0, \beta \leq F(x) \leq 1.$$

An application of T on F implies the following: for all $x \geq -x_0$,

$$A_0 e^{-x} \leq TF(x) \quad = \quad \exp\left(-\int_{-x}^{+\infty} F(u)du\right) \leq B_0 e^{-\beta x}, \tag{20}$$

where $A_0 := \exp\left(-x_0 - \int_{x_0}^{+\infty} F(u)du\right) > 0$ and $B_0 := \exp\left(-\beta x_0 - \int_{x_0}^{+\infty} F(u)du\right) > 0$ since F is integrable. Now, the (20) implies that the TF is integrable at $+\infty$ as well. Thus, by an inductive application $T^k F$ is integrable at $+\infty$ for all $k \geq 1$. Now, consider the following.

$$1 - T^2 F(x) \quad = \quad 1 - \exp\left(-\int_{-x}^{+\infty} TF(u)du\right). \tag{21}$$

Since TF is integrable at $+\infty$, $\int_{-x}^{\infty} TF(u)du \to 0$. Therefore, as $x \to -\infty$,

$$1 - T^2 F(x) \quad \approx \quad \int_{-x}^{+\infty} TF(u)du.$$

Therefore, using (20), we obtain for all x small enough, say smaller than some $x_1 \in \mathbb{R}$:

$$1 - A_1 e^{\beta x} \leq T^2 F(x) \leq 1 - B_1 e^{x}, \tag{22}$$

for some constants $A_1, B_1 > 0$. Applying T again yields for all $x \geq -x_1$:

$$A_2 e^{-x} \leq T^3 F(x) = \exp\left(-\int_{-x}^{+\infty} T^2 F(u)du\right) \leq B_2 e^{-x}, \tag{23}$$

with the new positive constants $A_2 := \exp\left(-x_1 - \int_{x_1}^{+\infty} T^2 F(u)du + \int_{-\infty}^{x_1} B_1 e^{u} du\right)$ and $B_2 := \exp\left(-x_1 - \int_{x_1}^{+\infty} T^2 F(u)du + \int_{-\infty}^{x_1} A_1 e^{\beta u} du\right)$. Of course, since TF satisfies the same assumptions as F this domination will also hold for $T^4 F$. Finally, by the same argument as (22) above, one easily deduce from (23) that for x small enough :

$$1 - A_3 e^{x} \leq T^4 F(x) \leq 1 - B_3 e^{x}, \tag{24}$$

where A_3, B_3 are positive constants.

Using arguments similar to those used for (23) and (24), we obtain that there exists x_4 such that for all $x \geq x_4$,

$$A_4 e^{-x} \;\leq\; T^4 F(x) \;\leq\; B_4 e^{-x}. \tag{25}$$

And, there exists x_4' such that for $x \leq x_4'$,

$$1 - A_4' e^{-x} \;\leq\; T^4 F(x) \;\leq\; 1 - B_4' e^{-x}. \tag{26}$$

Now, recall definition of $\widehat{T}^4 F$ defined over $\mathbb{R}$:

$$\widehat{T}^4 F \colon x \mapsto x + \ln\left(\frac{T^4 F(x)}{1 - T^4 F(x)}\right).$$

Now, as $x \to \infty$ $T^4 F(x) \to 0$. Therefore, for large enough x using (25), the following holds:

$$\begin{aligned} x + \ln\left(\frac{T^4 F(x)}{1 - T^4 F(x)}\right) &= O(1) + x + \ln T^4 F(x) \\ &= O(1) + x - x \;=\; O(1). \end{aligned} \tag{27}$$

Similarly, using (26) we obtain that as $x \to -\infty$, the $\widehat{T}^4 F(x)$ is uniformly bounded by an $O(1)$ term. Therefore, invoking continuity of $\widehat{T}^4 F$ over $\mathbb{R}$, we obtain that it is uniformly bounded over the $\mathbb{R}$. □

Lemma 6 *If $F \in \mathcal{D}$ is such that $\widehat{F}$ is bounded, then $\widehat{T^2F}$ is bounded too and :*

$$\begin{aligned} \sup_{\mathbb{R}} \widehat{T^2F} &\leq \sup_{\mathbb{R}} \widehat{F}; \\ \inf_{\mathbb{R}} \widehat{T^2F} &\geq \inf_{\mathbb{R}} \widehat{F}. \end{aligned}$$

Further, the above inequalities are strict if and only if $\widehat{F}$ is not constant on $\mathbb{R}$.

Proof. The proof of inequalities is imminent from the non-decreasing property of T^2 we noted earlier. Specifically, consider the following. Let,

$$\inf_{\mathbb{R}} \widehat{F} \triangleq m \leq \widehat{F} \leq M \triangleq \sup_{\mathbb{R}} \widehat{F}.$$

That is,

$$\theta_m F^* \leq F \leq \theta_M F^*.$$

Now, non-decreasing property of T^2 and F^* being fixed point implies

$$\theta_m F^* \leq T^2 F \leq \theta_M F^* \;\Rightarrow\; m \leq \widehat{T^2F} \leq M.$$

Thus, we obtain the desired inequalities

$$\sup_{\mathbb{R}} \widehat{F} \;\leq\; \inf_{\mathbb{R}} \widehat{T^2F} \;\leq\; \sup_{\mathbb{R}} \widehat{T^2F} \;\leq\; \sup_{\mathbb{R}} \widehat{F}.$$

Next, we show that the inequality is strict if and only if $\widehat{F}$ is not a constant. Note that since $\theta_t F^*$ is fixed point for T^2 for any constant t, if $\widehat{F}$ is constant then so is $\widehat{T}^2F$ and equal to $\widehat{F}$. Thus, we only need to show that if $\widehat{F}$ is not a constant over $\mathbb{R}$ then the inequality is strict.

To this end, note that the left-continuity of F implies that of $\widehat{F}$. Therefore if $\widehat{F}$ is not constant on $\mathbb{R}$ then there exists an open interval (a, b) such that $M' = \sup_{(a,b)} \widehat{F} < \sup_{\mathbb{R}} \widehat{F} = M$. Therefore, for $x \geq -a$,

$$\begin{aligned} TF(x) &= \exp\left(-\int_{-x}^{+\infty} F(u)du\right) \\ &\geq \exp\left(-\int_{-x}^{a} \theta_M F^*(u)du\right) \exp\left(-\int_{a}^{b} \theta_{M'} F^*(u)du\right) \exp\left(-\int_{b}^{\infty} \theta_M F^*(u)du\right) \\ &= \kappa \times T(\theta_M F^*)(x) \quad \text{with} \quad \kappa := \exp\left(\int_a^b \left(\theta_M F^* - \theta_{M'} F^*\right)(u)du\right) > 1. \end{aligned}$$

Applying T again implies that for every $x \in \mathbb{R}$,

$$\begin{cases} x \leq a \Rightarrow T^2F(x) \leq \exp\left(-\kappa \int_{-x}^{+\infty} T(\theta_M F^*)(u)du\right) = \left(\theta_M F^*(x)\right)^{\kappa}; \\ x \geq a \Rightarrow T^2F(x) \leq \exp\left(-\int_{-x}^{-a} T(\theta_M F^*)(u)du\right) \left(\theta_M F^*(a)\right)^{\kappa} = \kappa' \times \theta_M F^*(x), \end{cases}$$

where $\kappa' := \left(\theta_M F^*(a)\right)^{\kappa-1} < 1$. Therefore, the right hand side of both the terms can be strictly upper bounded by $\theta_M F^*(x)$ implying that $\widehat{T^2F}(x) < M$ for all $x \in \mathbb{R}$. In order to make sure that the supremum also remains strictly lower bounded, we need to worry about case when $x \to \pm\infty$. For this, observe that (recall definition of F^*)

$$\begin{cases} x \leq a \Rightarrow \widehat{T^2F}(x) \leq x + \ln\left(\dfrac{\left(\theta_M F^*(x)\right)^{\kappa}}{1 - \left(\theta_M F^*(x)\right)^{\kappa}}\right) \xrightarrow[x \to -\infty]{} M - \ln\kappa < M; \\ x \geq a \Rightarrow \widehat{T^2F}(x) \leq x + \ln\left(\dfrac{\kappa' \times \theta_M F^*(x)}{1 - \kappa' \times \theta_M F^*(x)}\right) \xrightarrow[x \to +\infty]{} M + \ln\kappa' < M. \end{cases}$$

The other inequality can be obtained in exactly the same manner; therefore we skip the details.
□

Lemma 7 *Consider $F \in \mathcal{D}$ such that $\widehat{F}$ is bounded. Then,*

1. *$\widehat{T^kF}$ is continuously differentiable for $k \geq 2$, and*

2. *$\{(\widehat{T^kF})', k \geq 3\}$ is uniformly integrable on $\mathbb{R}$.*

PROOF. First, we prove the claim (1) about continous differentiability. For this, consider $k \geq 2$. Now, $T^{k-2}F$ is bounded by 1 on $\mathbb{R}$, $T^{k-1}F\colon x \mapsto \exp\left(-\int_x^\infty T^{k-2}F(u)du\right)$ is Lipschitz continuous with Lipschitz constant 1. Therefore, $T^kF\colon x \mapsto \exp\left(-\int_x^\infty T^{k-1}F(u)du\right)$ is continuously differentiable on $\mathbb{R}$ such that for all $x \in \mathbb{R}$,

$$(T^kF)'(x) = -T^kF(x)\,T^{k-1}F(-x). \tag{28}$$

This in turn implies that $\widehat{T^kF}\colon x \to x + \ln\left(\frac{T^kF(x)}{1-T^kF(x)}\right)$ is indeed continuously differentiable on $\mathbb{R}$ and for all $x \in \mathbb{R}$,

$$(\widehat{T^kF})'(x) = \frac{1 - T^kF(x) - T^{k-1}F(-x)}{1 - T^kF(x)}. \tag{29}$$

Now, the claim (2) about uniform integrability of $\{(\widehat{T}^2F)', k \geq 3\}$. Recall that Lemma 6 ensures uniform boundedness of the family $\{\widehat{T^kF}, k \geq 0\}$ on $\mathbb{R}$ since $\widehat{F}$ is bounded. Let one such bound be $M \geq 0$. This allows us to bound the above fraction, for given x by $\frac{e^{2M}-1}{1+e^{x+M}}$ independently of k. This already yields uniform integrability towards $+\infty$. For uniform integrability towards $-\infty$, note that the numerator in (29) is continuously differentiable on $\mathbb{R}$ for $k \geq 3$ by (28). Therefore, for $x \in \mathbb{R}$,

$$\begin{aligned}\left|1 - T^kF(x) - T^{k-1}F(-x)\right| &= \left|\int_{-\infty}^{x} T^{k-1}F(-u)\big(T^kF(u) - T^{k-2}F(u)\big)\,du\right| \\ &\leq \int_{-\infty}^{x} \theta_M F^*(-u)\big(\theta_M F^*(u) - \theta_{-M}F^*(u)\big)du.\end{aligned}$$

This is enough to ensure uniform integrability at $-\infty$: because, the above integral is bounded above by $O(e^{2x})$ whereas the denominator (in 29) scales as $\Theta(e^x)$, thus yielding the overall bound on $(\widehat{T^kF})'(x)$ as $O(e^x)$ as $x \to -\infty$. This completes the proof of Lemma 7. □

Now, we are ready to state the main result of this sub-section, which is characterizes the asymptotic behavior of T^k.

Theorem 4 (Dynamics of T on $\mathcal{D}$) *Consider any $F \in \mathcal{D} \setminus \{0\}$ that is integrable at $+\infty$. Then*

$$\sup_{\mathbb{R}} \left|\widehat{T^kF} - (-1)^k\gamma\right| \underset{k\to\infty}{\searrow} 0,$$

for some constant $\gamma \in \mathbb{R}$ (dependent on F). Specifically, the following convergence happens uniformly on $\mathbb{R}$:

$$T^{2k}F \xrightarrow[k\to\infty]{} \theta_\gamma F^* \text{ and } T^{2k+1}F \xrightarrow[k\to\infty]{} \theta_{-\gamma}F^*.$$

Remark 2 *The above characterization is optimal in the following sense: $F \equiv 0$ or $\int_0^\infty F = +\infty$ then the sequence $(T^kF)_{k\geq 1}$ simply alternates between the 0 function and the 1 function.*

PROOF. Lemma 5 ensures existence of $t \geq 0$ such that for all $k \in \mathbb{N}$,

$$\theta_{-t}F^* \leq T^kF \leq \theta_t F^*. \tag{30}$$

By Lemma 6, the bounded real sequences $(\inf_{\mathbb{R}} \widehat{T^{2k}F})_{k\geq 2}$ and $(\sup_{\mathbb{R}} \widehat{T^{2k}F})_{k\geq 2}$ are respectively non-decreasing and non-increasing. Hence they converge to say m and M respectively as $k \to \infty$. By Arzela-Ascoli theorem, the family of pointwise bounded and equi-continuous functions $(T^{2k}F)_{k\geq 2}$ is relatively compact with respect to the topology of uniform convergence on every compact set of $\mathbb{R}$. Therefore, there exists a convergent subsequence, $\varphi(k)$ such that

$$T^{2\varphi(k)}F \xrightarrow[k\to\infty]{} F_\infty.$$

This uniform convergence implies that $\widehat{T^{2\varphi(k)}F}$ converges to $\widehat{F_\infty}$ since on every fixed compact set of $\mathbb{R}$ the uniform bound (30) keeps all the values of the $T^{2\varphi(k)}F, k \geq 0$ within a compact set of $]0,1[$ over which the mapping $y \mapsto \ln\frac{y}{1-y}$ is uniformly continuous. Therefore, in the limit we obtain that $m \leq \widehat{F_\infty} \leq M$. Moreover these inequalities are tight, in the sense that

$$m = \inf_{\mathbb{R}} \widehat{F_\infty}; \text{ and } M = \sup_{\mathbb{R}} \widehat{F_\infty}.$$

To see this, consider any $\varepsilon > 0$. Then, Lemma 7 implies existence of a compact set K_ε such that for all k large enough

$$\begin{aligned} \inf_{K_\varepsilon} \widehat{T^{\varphi(k)}F} &\leq \inf_{\mathbb{R}} \widehat{T^{\varphi(k)}F} + \varepsilon; \\ \sup_{K_\varepsilon} \widehat{T^{\varphi(k)}F} &\geq \sup_{\mathbb{R}} \widehat{T^{\varphi(k)}F} - \varepsilon. \end{aligned}$$

Now, using the above as $k \to \infty$ we obtain that $\inf_{\mathbb{R}} \widehat{F_\infty} \leq m + \varepsilon$ and $\sup_{\mathbb{R}} \widehat{F_\infty} \geq M - \varepsilon$. Since $\varepsilon > 0$ is abribtrary, we conclude that

$$\inf_{\mathbb{R}} \widehat{F_\infty} = m \text{ and } \sup_{\mathbb{R}} \widehat{F_\infty} = M.$$

Now, the restriction of T to the subset $\{F \in \mathcal{D}, -t \leq \widehat{F} \leq t\}$ is clearly continuous with respect to the topology of uniform convergence on every compact set. Therefore,

$$T^{2(\varphi(k)+1)}F \xrightarrow[k\to\infty]{} T^2 F_\infty.$$

But using exactly the same arguments as above, we again obtain a similar conclusion that

$$\inf_{\mathbb{R}} \widehat{T^2 F_\infty} = m \text{ and } \sup_{\mathbb{R}} \widehat{T^2 F_\infty} = M.$$

Therefore, Lemma 6 implies that it must be that $m = M$. That is, we have proved uniform convergence of $(\widehat{T^{2k}F})_{k\geq 0}$ to a constant function on $\mathbb{R}$. Finally property (18) ensures convergence of $(\widehat{T^{2k+1}F})_{k\geq 0}$ to the opposite constant. □

4.2 Attractiveness of messages: strong convergence on $\mathcal{T}$. So far, we have established the distributional convergence of messages by establishing the convergence of operator T. To complete the algorithm analysis, we need to prove sample-path wise convergence of the message process – which is stronger than that established in the last section. To this end, we will construct a stochastic coupling on $\mathcal{T}$ guaranteeing the desired strong convergence of the recursive tree process defined by (14). We note that Aldous and Bandyhopadhyay [3, 5] have studied the special case where the initial messages $(\langle v \to \dot{v}\rangle^0_{\mathcal{T}})_{v\in\mathcal{V}^*}$ i.i.d. with distribution being the fix-point F^*. They established L^2-convergence of the message process to the (almost surely) unique stationary configuration independent of the initial message realization $(\langle v \to \dot{v}\rangle^0_{\mathcal{T}})_{v\in\mathcal{V}^*}$. They call this property as *endogenity.* We want to establish such an *endogenity* property when initial messages are generated in an i.i.d. manner with distribution F as long as F is *reasonable* (precise conditions below). Now, we state the main result of this sub-section.

Theorem 5 (Strong convergence on $\mathcal{T}$) *Let initial messages $\left(\langle v \to \dot{v}\rangle^0_{\mathcal{T}}\right)_{v\in\mathcal{V}^*}$ be i.i.d. satisfying*

$$\mathbb{P}\left(\langle v \to \dot{v}\rangle^0_{\mathcal{T}} > -\infty\right) > 0 \text{ and } \mathbb{E}\left[\left(\langle v \to \dot{v}\rangle^0_{\mathcal{T}}\right)^+\right] < \infty.$$

Then, there exists a constant $\gamma \in \mathbb{R}$ (depending upon the initial distribution), such that the recursive tree process $\left((\langle v \to \dot{v}\rangle^k_{\mathcal{T}})_{v\in\mathcal{V}^*}\right)_{k\geq 0}$ defined on $\mathcal{T}$ by (14) converges to the almost sure unique stationary configuration $((\langle v \to \dot{v}\rangle^*_{\mathcal{T}})_{v\in\mathcal{V}^*})_{k\geq 0}$ in the following sense:

$$\forall v \in \mathcal{V}^*, \langle v \to \dot{v}\rangle^k_{\mathcal{T}} - (-1)^k\gamma \xrightarrow[k\to\infty]{L^1} \langle v \to \dot{v}\rangle^*_{\mathcal{T}}.$$

Further, the estimate of assignment at the root of $\mathcal{T}$ corresponding to these messages converge in probability to the estimate at the root based on the stationary configuration:

$$\pi^k_{\mathcal{T}}(\varnothing) \stackrel{def}{=} \arg\min_{i\geq 1}\left\{\left\|\varnothing, i\right\|_{\mathcal{T}} - \langle i \to \varnothing\rangle^k_{\mathcal{T}}\right\} \xrightarrow[k\to\infty]{proba} \pi^*_{\mathcal{T}}(\varnothing) \stackrel{def}{=} \arg\min_{i\geq 1}\left\{\left\|\varnothing, i\right\|_{\mathcal{T}} - \langle i \to \varnothing\rangle^*_{\mathcal{T}}\right\}.$$

Remark 3 *Our assumptions on the initial message distribution is necessary as otherwise the message process can become infinite everywhere on $\mathcal{T}$ after the first iteration, almost surely.*

Proof. Let F denote the anti-cdf of the initial message distribution and γ be the corresponding constant appearing in Theorem 4. First, observe that if we add a constant to all the initial values then under the dynamics (14), the same constant is added to every even message $\langle v \to \dot{v}\rangle^{2k}_{\mathcal{T}}$ and the negative of the constant is added to every odd message $\langle v \to \dot{v}\rangle^{2k+1}_{\mathcal{T}}$. Therefore, without loss of generality we assume $\gamma = 0$. That is, for any $\varepsilon > 0$ there exists $k_\varepsilon \in \mathbb{N}$ so that

$$\theta_{-\varepsilon}F^* \leq T^{k_\varepsilon}F \leq \theta_\varepsilon F^*.$$

By the Skorohod's representation theorem, there exists joint probability space $E' = (\Omega', \mathcal{F}', P')$, possibly different from the original space $E = (\Omega, \mathcal{F}, P)$ on which PWIT $\mathcal{T}$ is defined, and a random variable X^ε defined in it with distribution $T^{k_\varepsilon} F$ along with two other random variables X^- and X^+ with distribution F^* such that

$$X^- - \varepsilon \le X^\varepsilon \le X^+ + \varepsilon, \quad \text{with probability 1.}$$

Now consider the product space $(\bigotimes_{v\in\mathcal{V}} E') \otimes E$ over which the PWIT $\mathcal{T}$ and independent copies of $(X_v^-, X_v^\varepsilon, X_v^+)_{v\in\mathcal{V}}$ of the triple (X^-, X, X^+) are defined for each vertex $v \in \mathcal{V}$. On $\mathcal{T}$, let us compare the message configurations $\left(\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k,-}\right)_{v\in\mathcal{V}^*, k\ge 0}$, $\left(\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k,\varepsilon}\right)_{v\in\mathcal{V}^*, k\ge 0}$ and $\left(\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k,+}\right)_{v\in\mathcal{V}^*, k\ge 0}$ obtained by three versions of our recursive tree process differing only in their initial configurations as follows: for each $v \in \mathcal{V}^*$,

$$\begin{aligned}
\langle v \to \dot{v}\rangle_{\mathcal{T}}^{0,-} &:= X_v^-; \\
\langle v \to \dot{v}\rangle_{\mathcal{T}}^{0,\varepsilon} &:= X_v^\varepsilon; \\
\langle v \to \dot{v}\rangle_{\mathcal{T}}^{0,+} &:= X_v^+.
\end{aligned}$$

Due to the anti-monotony and 'homogeneity' of the update rule (14), it can be easily checked that the inequalities $X_v^- - \varepsilon \le X_v^* \le X_v^+ + \varepsilon$ propagate in the sense that for any $k \ge 0$ and $v \in \mathcal{V}^*$,

$$\begin{aligned}
\langle v \to \dot{v}\rangle_{\mathcal{T}}^{2k,-} - \varepsilon \quad &\le \quad \langle v \to \dot{v}\rangle_{\mathcal{T}}^{2k,\varepsilon} \quad \le \quad \langle v \to \dot{v}\rangle_{\mathcal{T}}^{2k,+} + \varepsilon\ ; \\
\langle v \to \dot{v}\rangle_{\mathcal{T}}^{2k+1,+} - \varepsilon \quad &\le \quad \langle v \to \dot{v}\rangle_{\mathcal{T}}^{2k+1,\varepsilon} \quad \le \quad \langle v \to \dot{v}\rangle_{\mathcal{T}}^{2k+1,-} + \varepsilon.
\end{aligned}$$

By construction, for every $v \in \mathcal{V}^*$, the sequences $\left(\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k+k_\varepsilon}\right)_{k\ge 0}$ and $\left(\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k,\varepsilon}\right)_{k\ge 0}$ are distributionally equivalent. Therefore, for every $k \ge k_\varepsilon$ and $v \in \mathcal{V}^*$, we have

$$\begin{aligned}
\sup_{s,t\ge k} \left\| \langle v \to \dot{v}\rangle_{\mathcal{T}}^{s} - \langle v \to \dot{v}\rangle_{\mathcal{T}}^{t} \right\|_{L^2} &= \sup_{s,t\ge k-k_\varepsilon} \left\| \langle v \to \dot{v}\rangle_{\mathcal{T}}^{s,\varepsilon} - \langle v \to \dot{v}\rangle_{\mathcal{T}}^{t,\varepsilon} \right\|_{L^2} \\
\le \sup_{t\ge k-k_\varepsilon} \left\| \langle v \to \dot{v}\rangle_{\mathcal{T}}^{t,+} - \langle v \to \dot{v}\rangle_{\mathcal{T}}^{*} \right\|_{L^2} &+ \sup_{t\ge k-k_\varepsilon} \left\| \langle v \to \dot{v}\rangle_{\mathcal{T}}^{t,-} - \langle v \to \dot{v}\rangle_{\mathcal{T}}^{*} \right\|_{L^2} + 2\varepsilon.
\end{aligned}$$

The endogeneity property of the logistic recursive tree process established by Aldous and Bandyhopadhyay [3, 5] implies that the first two terms vanish as $k \to \infty$. Thus, the sequence $\left(\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k}\right)_{k\ge 0}$ is cauchy in the L^2 space and therefore it is convergent. Clearly, Theorem 4 implies that the marginal distribution of limiting message is F^*. However, it is not sufficient for our purpose: we need to show that the limiting (joint) message configuration over $\mathcal{T}$ is stationary, i.e. a fixed point of the dynamics as it corresponds to the minimum cost matching (assignment) as established by Aldous [2]. This is the precise stronger notion of convergence we are establishing here. Note that the stationary configuratin is almost surely unique as established by Aldous and therefore establishing stationarity of the limiting configuration relates to the minimum cost matching assignment.

To this end, recall that for all $k \ge 0$ and $v \in \mathcal{V}^*$, almost surely

$$\langle v \to \dot{v}\rangle_{\mathcal{T}}^{k+1} = \min_{i\ge 1} \left\{ \left\| v, v.i \right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^{k} \right\}.$$

Since the left side tends to $\langle v \to \dot{v}\rangle_{\mathcal{T}}^{\infty}$, it is enough to show that the right side tends in probability to $\min_{i\ge 1} \left\{ \left\| v, v.i \right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^{\infty} \right\}$ despite the infinite number of terms involved in the min. Indeed, for every $\varepsilon > 0$, $i_0 \ge 1$ and $k \ge 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left| \min_{i\ge 1} \left\{ \left\| v, v.i \right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^{k} \right\} - \min_{i\ge 1} \left\{ \left\| v, v.i \right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^{\infty} \right\} \right| \ge \varepsilon \right) \\
&\le \sum_{i=1}^{i_0} \mathbb{P}\left(\left| \langle v.i \to v\rangle_{\mathcal{T}}^{k} - \langle v.i \to v\rangle_{\mathcal{T}}^{\infty} \right| \ge \varepsilon \right) \\
&+ \sum_{i=i_0+1}^{\infty} \mathbb{P}\left(\left\| v, v.i \right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^{\infty} \le \left\| v, v.1 \right\|_{\mathcal{T}} - \langle v.1 \to v\rangle_{\mathcal{T}}^{\infty} \right) \\
&+ \sum_{i=i_0+1}^{\infty} \mathbb{P}\left(\left\| v, v.i \right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^{k} \le \left\| v, v.1 \right\|_{\mathcal{T}} - \langle v.1 \to v\rangle_{\mathcal{T}}^{k} \right).
\end{aligned}$$

Now, the first term vanishes as $k \to \infty$ because L^2 convergence implies the convergence in probability. The second term can be made arbitrarily small by choosing i_0 large enough since the infinite sum is convergent as we shall show next. The i^{th} term in the summation of the second term is precisely $\mathbb{P}(\xi_{i-1} \leq X_1^* - X_2^*)$ where X_1^* and X_2^* are i.i.d. with distribution F^* and $(\xi_i)_{i\geq 1}$ is a Poisson point process with rate 1 independent of X_1^*, X_2^*. Therefore,

$$\sum_{i=1}^{\infty} \mathbb{P}(\xi_i \leq X_1^* - X_2^*) = \sum_{i=1}^{\infty} \mathbb{E}\left[\int_0^{(X_1^*-X_2^*)^+} e^{-x} \frac{x^{i-1}}{(i-1)!}\, dx\right] = \mathbb{E}\left[\left(X_1^* - X_2^*\right)^+\right] < +\infty. \tag{31}$$

Now, for the third term, Theorem 5 provides the uniform bound of

$$\theta_{-M} F^* \leq T F^k \leq \theta_M F^*,$$

for some $M \geq 0$. Therefore, the infinite sum is bounded above uniformly in k by the infinite sum of the $\mathbb{P}(\xi_{i-1} \leq X_1^* - X_2^* + 2M)$. Similar calculation as that done for second term imply that this is convergent. By uniqueness of the stationary configuration (c.f. [5]), we obtain that almost surely, for all $v \in \mathcal{V}^*$,

$$\langle v \to \dot{v} \rangle_{\mathcal{T}}^{\infty} = \langle v \to \dot{v} \rangle_{\mathcal{T}}^{*}.$$

To conclude the proof, we need to show that the estimation at the root of $\mathcal{T}$ converges as well. For this, observe that by the Borel-Cantelli's Lemma convergence (31) ensures that the minimum of the $\left\|\varnothing, i\right\|_{\mathcal{T}} - \langle i \to \varnothing\rangle_{\mathcal{T}}^*, i \geq 1$ is achieved. Now, since all of these quantities are pairwise distinct with probability 1, their argmin $\pi_{\mathcal{T}}^*(\varnothing)$ is singleton and well-defined. The same holds for $\pi_{\mathcal{T}}^k(\varnothing)$ due to the uniform domination $\theta_{-M} F^* \leq T F^k \leq \theta_M F^*$. Putting all together, we obtain that for every $i_0 \geq 1$ and $k \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left(\pi_{\mathcal{T}}^k(\varnothing) \neq \pi_{\mathcal{T}}^*(\varnothing)\right) \\
&\leq \mathbb{P}\left(\operatorname*{arg\,min}_{1\leq i<i_0}\left\{\left\|\varnothing, i\right\|_{\mathcal{T}} - \langle i \to \varnothing\rangle_{\mathcal{T}}^k\right\} \neq \operatorname*{arg\,min}_{1\leq i<i_0}\left\{\left\|\varnothing, i\right\|_{\mathcal{T}} - \langle i \to \varnothing\rangle_{\mathcal{T}}^*\right\}\right) \\
&\quad + \sum_{i=i_0+1}^{\infty} \mathbb{P}\left(\left\|v, v.i\right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^* \leq \left\|v, v.1\right\|_{\mathcal{T}} - \langle v.1 \to v\rangle_{\mathcal{T}}^*\right) \\
&\quad + \sum_{i=i_0+1}^{\infty} \mathbb{P}\left(\left\|v, v.i\right\|_{\mathcal{T}} - \langle v.i \to v\rangle_{\mathcal{T}}^k \leq \left\|v, v.1\right\|_{\mathcal{T}} - \langle v.1 \to v\rangle_{\mathcal{T}}^k\right),
\end{aligned}$$

As before, the two infinite summations can be arbitrarily small (uniformly in k) by choosing i_0 large enough; the first term vanishes as $k \to \infty$ due to the continuity of arg min under convergence of messages. This completes the proof of Theorem 5. □

5. Third step: completing proof of Theorem 1

Here, we complete the proof of Theorem 1. As stated earlier, we have followed the three-step proof plan. The first two steps are proved in the previous two sections. The third step utilizes the following remarkable result of Aldous [2].

Theorem 6 *Let $\pi_{\mathcal{T}}^*$ be the assignment associated to the almost sure unique stationary configuration $(\langle v \to \dot{v}\rangle_{\mathcal{T}}^*)_{v\in\mathcal{V}^*}$. Then, $\pi_{\mathcal{T}}^*$ is almost surely a perfect matching on $\mathcal{T}$ such that,*

$$\left(\mathcal{K}_{n,n}, \pi_{\mathcal{K}_{n,n}}^*\right) \xrightarrow{\mathcal{D}} \left(\mathcal{T}, \pi_{\mathcal{T}}^*\right),$$

with respect to the topology of local weak convergence.

Proof. **(Theorem 1)** Now, we are ready for completing the proof of our main result. For this, we will need some additional useful formalism which is developed next. Suppose, we are given a complete separable metric space (Λ, d_Λ). There is a natural way to extend the notion of local weak convergence while in-corporating such a metric-space as follows. Suppose, the rooted geometric graph $G = (V, E, \varnothing, \|\cdot\|_G)$ is labeled by points of metric space Λ, i.e. every vertex $v \in V$ is assigned a label $\lambda(v) \in \Lambda$; and/or every oriented edge $(v, w) \in \overrightarrow{E}$ is also assigned a label $\lambda_G(v, w) \in \Lambda$. Under such setup, a sequence $(G_n)_{n\geq 1}$ of such *labeled rooted geometric graphs are said to converge locally to a labeled rooted geometric graph* G if the following two conditions hold:

1. G_n converge to G as per the local weak convergence, and
2. As in definition of local weak convergence, for each $\rho > 0$ such that no node in G is at distance ρ from its root, for $n \geq n_\rho$ there exists isomorphisms

$$\gamma_n^\varrho \colon \lceil G \rceil_\varrho \rightleftharpoons \lceil G_n \rceil_\varrho, n \geq n_\varrho$$

such that for all vertex v in $\lceil G \rceil_\varrho$,

$$\lambda_{G_n}\left(\gamma_n^\varrho(v)\right) \xrightarrow[n\to\infty]{} \lambda_G\left(v\right),$$

and/or similarly, for any oriented edge (v, w) in $\lceil G \rceil_\varrho$,

$$\lambda_{G_n}\left(\gamma_n^\varrho(v), \gamma_n^\varrho(w)\right) \xrightarrow[n\to\infty]{} \lambda_G\left(v, w\right).$$

In order to metrize the topology induced by the above definition of convergence for 'labelled' geometric graphs, we can incorporate d_Λ into the metric defined by (4). This will make the above space complete and separable metric space. Within thus developed framework, Theorem 3 can be restated as follows: for all $k \geq 0$,

$$\left(\mathcal{K}_{n,n}, \langle \cdot \to \cdot \rangle^k_{\mathcal{K}_{n,n}}, \pi^k_{\mathcal{K}_{n,n}}\right) \xrightarrow[n\to\infty]{\mathcal{D}} \left(\mathcal{T}, \langle \cdot \to \cdot \rangle^k_{\mathcal{T}}, \pi^k_{\mathcal{T}}\right), \tag{32}$$

where the mapping $\pi^k_{\mathcal{K}_{n,n}}$ is a $\{0,1\}$-valued edge-labeling function $(v, w) \mapsto \mathbf{1}_{\{w = \pi^k_{\mathcal{K}_{n,n}}(v)\}}$. Similarly, Theorem 6 implies the following: for $k \geq 0$,

$$\left(\mathcal{K}_{n,n}, \langle \cdot \to \cdot \rangle^k_{\mathcal{K}_{n,n}}, \pi^k_{\mathcal{K}_{n,n}}, \pi^*_{\mathcal{K}_{n,n}}\right) \xrightarrow[n\to\infty]{\mathcal{D}} \left(\mathcal{T}, \langle \cdot \to \cdot \rangle^k_{\mathcal{T}}, \pi^k_{\mathcal{T}}, \pi^*_{\mathcal{K}_{n,n}}\right). \tag{33}$$

Therefore, the error in the BP algorithm's estimation on $\mathcal{K}_{n,n}$ converges to BP algorithm's estimation on $\mathcal{T}$: for all $k \geq 0$,

$$\mathbb{E}\left[d(\pi^k_{\mathcal{K}_{n,n}}, \pi^*_{\mathcal{K}_{n,n}})\right] = \mathbb{P}\left(\pi^k_{\mathcal{K}_{n,n}}(\varnothing) \neq \pi^*_{\mathcal{K}_{n,n}}(\varnothing)\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\pi^k_{\mathcal{T}}(\varnothing) \neq \pi^*_{\mathcal{T}}(\varnothing)\right). \tag{34}$$

Now, the Theorem (5) imples that

$$\lim_{k\to\infty} P\left(\pi^k_{\mathcal{T}}(\varnothing) \neq \pi^*_{\mathcal{T}}(\varnothing)\right) = 0.$$

This completes the proof of Theorem 1. □

6. Conclusion

In this paper, we established that the BP algorithm finds almost optimal solution to a random assignment problem in $O(n^2)$ time for a problem of size n with high probability. The natural lower bound of $\Omega(n^2)$ due to it being the input-size of the problem, makes BP an (order) optimal algorithm for finding minimum cost matching in a bipartite graph. This result significantly improves over the $O(n^3)$ bound proved by Bayati, Shah and Sharma [6] for BP for bipartite graph with arbitrary weights; or for that matter the best known worst case bound on performance of algorithm by Edmonds and Karp [10].

Beyond the obvious practical interest of such an extremely efficient distributed algorithm for locally solving huge instances of the optimal assignment problem, we hope that the method used here – essentially replacing the asymptotical analysis of the iteration as the size of the underlying graph tends to infinity by its exact study on the infinite limiting structure revealed via local weak convergence – will become a powerful tool in the fascinating quest for a general mathematical understanding of loopy belief propagation. To the best of our knowledge, this is the first non-trivial use of local weak convergence frame-work for analyzing performance of algorithm.

References

[1] D. Aldous. Asymptotics in the random assignment problem. *Probability Theory and Related Fields*, 93:507–534, 1992.

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

[3] D. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *Annals of Applied Probability*, 15(2):1047–1110, 2005.

[4] D. Aldous and J. M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. *Discrete and Combinatorial Probability*, 2003.

[5] A. Bandyopadhyay. Bivariate uniqueness in the logistic recursive distributional equation. Technical Report 629, Department of Statistics, University of California, Berkeley, CA, 2002.

[6] M. Bayati, D. Shah, and M. Sharma. Max-product for maximum weight matching: convergence, correctness and lp duality. *IEEE Transactions on Information Theory*, March 2008.

[7] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, 1999.

[8] D. Coppersmith and G. Sorkin. Constructive bounds and exact expectations for the random assignment problem. *Random Structures Algorithms*, 15:113–144, 1999.

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

[10] J. Edmonds and R. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Jour. of the ACM*, 19:248–264, 1972.

[11] M.X. Goemans and M.S. Kodialam. A lower bound on the expected cost of an optimal assignment. *Math. Oper. Res.*, 18:267–274, 1993.

[12] R.M. Karp. An upper bound on the expected cost of an optimal assignment. *In Discrete Algorithms and Complexity: Proceedings of the Japan- U.S. Joint Seminar, Academic Press*, pages 1–4, 1987.

[13] A. Lazarus. Certain expected values in the random assignment problem. *Oper. Res. Lett.*, 14:207–214, 1993.

[14] S. Linusson and J. Wåstlund. A proof of parisiŠs conjecture on the random assignment problem. *Probab. Theory Relat. Fields*, 128:419–440, 2004.

[15] M. Mézard and G. Parisi. On the solution of the random link matching problem. *Journal of Physics*, 48:1451–1459, 1987.

[16] C. Nair, B. Prabhakar, and M. Sharma. Proofs of the parisi and coppersmith-sorkin random assignment conjectures. *Random Structures Algorithms*, 27(4):413–444, 2005.

[17] B. Olin. Asymptotic properties of random assignment problems. *PhD thesis, Kungl Tekniska H́ogskolan, Stockholm, Sweden*, 1992.

[18] J. Pearl. *TProbabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA, 1988.

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

[20] J. Yedidia, W. Freeman, and Y. Weiss. Generalized belief propagation. Technical Report TR-2000-26, Mitsubishi Elect. Res. Lab., Cambridge, MA, 2000.