

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

Justin Salez, Devavrat Shah

► **To cite this version:**

Justin Salez, Devavrat Shah. Belief propagation : an asymptotically optimal algorithm for the random assignment problem. Mathematics of Operations Research, INFORMS, 2009. <inria-00358331>

HAL Id: inria-00358331

<https://hal.inria.fr/inria-00358331>

Submitted on 3 Feb 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Justin Salez

INRIA and 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 asks for the minimum-cost perfect matching in the complete $n \times n$ bipartite graph \mathcal{K}_{nn} with i.i.d. edge weights, say uniform on $[0, 1]$. In a remarkable work by Aldous (2001), the optimal cost was shown to converge to $\zeta(2)$ as $n \rightarrow \infty$, as conjectured by Mézard and Parisi (1987) through the so-called cavity method. The latter also suggested a non-rigorous decentralized strategy for finding the optimum, which turned out to be an instance of the Belief Propagation (BP) heuristic discussed by Pearl (1987). In this paper we use the objective method to analyze the performance of BP as the size of the underlying graph becomes large. Specifically, we establish that the dynamic of BP on \mathcal{K}_{nn} converges in distribution as $n \rightarrow \infty$ to an appropriately defined dynamic on the Poisson Weighted Infinite Tree, and we then prove correlation decay for this limiting dynamic. As a consequence, we obtain that BP finds an asymptotically correct assignment in $O(n^2)$ time only. This contrasts with both the worst-case upper bound for convergence of BP derived by Bayati, Shah and Sharma (2005) and the best-known computational cost of $\Theta(n^3)$ achieved by Edmonds and Karp's algorithm (1972).

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. Given a matrix of n^2 costs $(X_{i,j})_{1 \leq i,j \leq n}$, the assignment problem consists of determining a permutation π of $\{1, \dots, n\}$ whose total cost $\sum_{i=1}^n X_{i,\pi(i)}$ is minimal. This is equivalent to finding a minimum-weight complete matching in the $n \times n$ complete bipartite graph whose n^2 edges are weighted by the $(X_{i,j})$. Recall that a complete matching on a graph is a subset of pairwise disjoint edges covering all vertices. Here we consider the so-called random assignment problem where the $(X_{i,j})$ are i.i.d. with cumulative distribution function denoted by H , i.e. $H(t) = \mathbb{P}(X_{i,j} \leq t)$. We let \mathcal{K}_{nn} denote the resulting randomly weighted $n \times n$ bipartite graph and $\pi_{\mathcal{K}_{nn}}^*$ its optimal matching. Observe that the continuity of H is a necessary and sufficient condition for $\pi_{\mathcal{K}_{nn}}^*$ to be a.s. unique. We are interested in the convergence of the BP heuristic for finding $\pi_{\mathcal{K}_{nn}}^*$ as n increases to infinity.

1.1 Related Work. Although it seems cunningly simple, the assignment problem has led to rich development in combinatorial probability and algorithm design since the early 1960s. Partly motivated to obtain insights for better algorithm design, the question of finding asymptotics of the average cost of $\pi_{\mathcal{K}_{nn}}^*$ became of great interest (see [20, 9, 13, 14, 18, 11, 8]). In 1987, through cavity method based calculations, Mézard and Parisi [16] conjectured that, for Exponential(1) edge weights,

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

This was rigorously established by Aldous [2] more than a decade later, leading to the formalism of “the objective method” (see survey by Aldous and Steele [4]). In 2003, an exact version of the above conjecture was independently established by Nair, Prabhakar and Sharma [17] and Linusson and Wästlund [15].

On the algorithmic aspect, the assignment problem has been extremely well studied and its consideration laid foundations for the rich theory of network flow algorithms. The best known algorithm is by

Edmonds and Karp [10] and takes $O(n^3)$ operations in the worst-case for arbitrary instance. For i.i.d. random edge weights, Karp [12] designed a special implementation of the augmenting path approach using priority queues that works in expected time $O(n^2 \log n)$. Concurrently, the statistical physics-based approach mentioned above suggested a non-rigorous decentralized strategy which turned out to be an instance of the more general BP heuristic, popular in artificial intelligence (see, book by Pearl [19] and work by Yedidia, Freeman and Weiss [21]). In a recent work, one of the authors of the present paper, Shah along with Bayati and Sharma [6], established correctness of this iterative scheme for any instance of the assignment problem, as long as the optimal solution is unique. More precisely, they showed exact convergence within at most $\lceil \frac{2n \max_{i,j} X_{i,j}}{\varepsilon} \rceil$ iterations, where ε denotes the difference of weights between optimum and second optimum. This upper bound is always greater than n , and can be shown to scale like $\Theta(n^2)$ as n goes to infinity in the random model. Since each iteration of the BP algorithm needs $\Theta(n^2)$ operations to be performed, one is left with an upper bound of $O(n^4)$ for the total computation cost. However, simulation studies tend to show much better performances on average than what is suggested by this worst-case analysis.

1.2 Our contribution. Motivated by the above discussion, we consider here the question of determining the convergence rate of BP for the random assignment problem. We establish that, for a large class of edge-weight distributions, the number of iterations required in order to find an almost optimal assignment remains in fact bounded as $n \rightarrow \infty$. Thus, the total computation cost scales as $O(n^2)$ only, in sharp contrast with both the worst-case upper bound for exact convergence of BP derived in [6] and the $\Theta(n^3)$ bound achieved by Edmonds and Karp's algorithm. Clearly, no algorithm can perform better than $\Omega(n^2)$, since it is the size of the input. That is, BP is an asymptotically optimal algorithm on average.

2. Result and organization.

2.1 BP algorithm. As we shall see later, the dynamics of BP on \mathcal{K}_{nn} happens to converge to the dynamics of BP on a limiting infinite tree. Therefore, we define the BP algorithm for an arbitrary weighted graph $G = (V, E)$. We use notation that the weight of $\{v, w\} \in E$ is $\|v, w\|_G$. By $w \sim v$, we denote that w is a neighbor of v in G . Note that a complete matching on G can be equivalently seen as an involutive mapping π_G connecting each vertex v to one of its neighbors $\pi_G(v)$. We shall henceforwards use this mapping representation rather than the edge set description.

The BP algorithm is distributed and iterative. Specifically, in each iteration $k \geq 0$, every vertex $v \in V$ sends a real-valued message $\langle v \rightarrow w \rangle_G^k$ to each of its neighbor $w \sim v$ as follows:

- *initialization rule:*

$$\langle v \rightarrow w \rangle_G^0 = 0; \quad (1)$$

- *update rule:*

$$\langle v \rightarrow w \rangle_G^{k+1} = \min_{u \sim v, u \neq w} \{ \|u, v\|_G - \langle u \rightarrow v \rangle_G^k \}. \quad (2)$$

Based on those messages, every vertex $v \in V$ estimates the neighbor $\pi_G^k(v)$ to which it connects as follows:

- *decision rule:*

$$\pi_G^k(v) = \arg \min_{u \sim v} \{ \|u, v\|_G - \langle u \rightarrow v \rangle_G^k \}. \quad (3)$$

When $G = \mathcal{K}_{nn}$, [6] ensures convergence of $\pi_{\mathcal{K}_{nn}}^k$ to the optimum $\pi_{\mathcal{K}_{nn}}^*$ as long as the latter is unique, which holds almost surely if and only if H is continuous. The present paper asks about the typical rate of such a convergence, and more precisely its dependency upon n as n increases to ∞ .

2.2 Result. In order to state our main result, we introduce the normalized Hamming distance between two given assignments π, π' on a graph $G = (V, E)$:

$$d(\pi, \pi') = \frac{1}{|V|} \text{card}\{v \in V, \pi(v) \neq \pi'(v)\}.$$

THEOREM 2.1 *Assume the cumulative distribution function H satisfies:*

- A1. *Regularity : H is continuous and $H'(0^+)$ exists and is non-zero;*
- A2. *Light-tail property : as $t \rightarrow \infty$, $H(t) = 1 - O(e^{-\beta t})$ for some $\beta > 0$.*

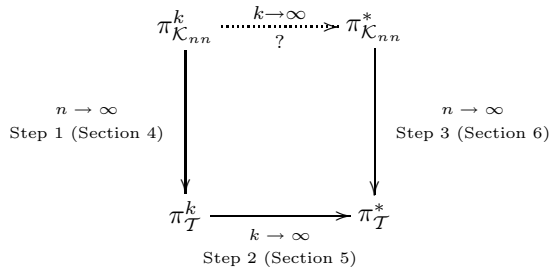
Then,

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[d(\pi_{\mathcal{K}_{nn}}^k, \pi_{\mathcal{K}_{nn}}^*) \right] \xrightarrow{k \rightarrow \infty} 0.$$

In other words, given any $\varepsilon > 0$, there exists $k(\varepsilon), n(\varepsilon)$ such that the expected fraction of non-optimal row-to-column assignments after $k(\varepsilon)$ iterations of the BP algorithm on a random $n \times n$ cost array is less than ε , no matter how large $n \geq n(\varepsilon)$ is. Consequently, the probability to get more than any given fraction of errors can be made as small as desired within finitely many iterations, independently of n . Since each iteration requires $O(n^2)$ operations, the overall computation cost scales as $O(n^2)$ only, with constant depending on the admissible error. This applies for a wide class of cost distributions, including uniform over $[0, 1]$ or Exponential.

REMARK 2.1 *It may be the case that the ε fraction of wrong row-to-column assignments results in local violations of the matching property. Depending on the context of application, this might be quite unsatisfactory. However, such an “ ε -feasible matching” can easily be modified in order to produce an honest matching without substantially increasing the total cost (see [1, Proposition 2] for details).*

2.3 Organization. The remaining of the paper is dedicated to proving Theorem 2.1. Although it is far from being an implication of the result by Aldous [2], it utilizes the machinery of local convergence, and in particular the Poisson Weighted Infinite Tree \mathcal{T} appearing as the limit of $(\mathcal{K}_{nn})_{n \geq 1}$. These notions are recalled in Section 3. The diagram below illustrates the three steps of our proof : Theorem 2.1 corresponds to establishing the top-horizontal arrow, which is done by establishing the three others.



1. First (Section 4), we prove that BP’s behavior on \mathcal{K}_{nn} “converges” as $n \rightarrow \infty$ to its behavior on \mathcal{T} . This is formally stated as Theorem 4.1 and corresponds to the left vertical arrow above.
2. Second (Section 5), we establish convergence of BP on \mathcal{T} . This is summarized as Theorem 5.2 and corresponds to the bottom horizontal arrow in the above diagram. We note that Theorem 5.1 resolves an open problem stated by Aldous and Bandyopadhyay ([3, Open Problem # 62]).
3. Third (Section 6), the connection between the fixed point on \mathcal{T} and the optimal matching on \mathcal{K}_{nn} is provided by the work by Aldous [2] – corresponding to the vertical right arrow and stated as Theorem 6.1. We use it to complete our proof.

3. Preliminaries. We recall here the necessary framework introduced by Aldous in [2]. Consider a rooted, edge-weighted and connected graph G , with distance between two vertices being defined as the infimum over all paths connecting them of the sum of edge weights along that path. For any $\varrho > 0$, define the ϱ -restriction of G as the subgraph $\lceil G \rceil_\varrho$ induced by the vertices lying within distance ϱ from the root. Call G a *geometric graph* if $\lceil G \rceil_\varrho$ is finite for every $\varrho > 0$.

DEFINITION 3.1 (LOCAL CONVERGENCE) *Let G, G_1, G_2, \dots be geometric graphs. We say that $(G_n)_{n \geq 1}$ converges to G if for every $\varrho > 0$ such that no vertex in G is at distance exactly ϱ from the root the following holds:*

1. $\exists n_\varrho \in \mathbb{N}$ s.t. the $\lceil G_n \rceil_\varrho, n \geq n_\varrho$ are all isomorphic¹ to $\lceil G \rceil_\varrho$;
2. The corresponding isomorphisms $\gamma_n^\varrho: \lceil G \rceil_\varrho \cong \lceil G_n \rceil_\varrho, n \geq n_\varrho$ can be chosen so that for every edge $\{v, w\}$ in $\lceil G \rceil_\varrho$:

$$\|\gamma_n^\varrho(v), \gamma_n^\varrho(w)\|_{G_n} \xrightarrow{n \rightarrow \infty} \|v, w\|_G.$$

In the case of labeled geometric graphs, each oriented edge (v, w) is also assigned a label $\lambda(v, w)$ taking values in some Polish space Λ . Then the isomorphisms $(\gamma_n^\varrho)_{n \geq n_\varrho}$ have to moreover satisfy the following:

3. For every oriented edge (v, w) in $\lceil G \rceil_\varrho$, $\lambda_{G_n}(\gamma_n^\varrho(v), \gamma_n^\varrho(w)) \xrightarrow{n \rightarrow \infty} \lambda_G(v, w)$.

The intuition behind this definition is the following: in any arbitrarily large but fixed neighborhood of the root, G_n should look very much like G for large n , in terms of structure (part 1), edge weights (part 2) and labels (part 3). With little work, one can define a distance that metrizes this notion of convergence and makes the space of (labeled) geometric graphs complete and separable. As a consequence, one can import the usual machinery related to the theory of weak convergence of probability measures. We refer the reader unfamiliar with these notions to the excellent book of Billingsley [7].

Now, consider our randomly weighted $n \times n$ bipartite graph \mathcal{K}_{nn} as a random geometric graph by fixing an arbitrary root, independently of the edge weights. Then the sequence $(\mathcal{K}_{nn})_{n \geq 1}$ happens to converge locally in distribution to an appropriately weighted infinite random tree. Before we formally state this result known as the ‘‘PWIT Limit Theorem’’ [2, 3], we introduce some notations that will be useful throughout the paper. We let \mathcal{V} denote the set of all finite words over the alphabet \mathbb{N}^* , \emptyset the empty word, ‘‘ \cdot ’’ the concatenation operation and for any $v \in \mathcal{V}^* = \mathcal{V} \setminus \{\emptyset\}$, \dot{v} the word obtained from v by deleting the last letter. We also set $\mathcal{E} = \{\{v, v.i\}, v \in \mathcal{V}, i \geq 1\}$. The graph $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ thus denotes an infinite tree with \emptyset as root, letters as the nodes at depth 1, words of length 2 as the nodes at depth 2, etc. Now, consider a collection $(\xi^v = \xi_1^v, \xi_2^v \dots)_{v \in \mathcal{V}}$ of independent, ordered Poisson point processes with intensity 1 on \mathbb{R}^+ , and assign to edge $\{v, v.i\} \in \mathcal{E}$ the weight $\|v, v.i\|_{\mathcal{T}} = \xi_i^v$. This defines the law of a random geometric graph \mathcal{T} called the ‘‘Poisson Weighted Infinite Tree’’ (PWIT).

THEOREM 3.1 (PWIT LIMIT THEOREM, ALDOUS [1, 2]) *Under assumption A1 on H :*

$$nH'(0^+) \mathcal{K}_{nn} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{T}, \quad (4)$$

in the sense of local weak convergence of geometric graphs.

REMARK 3.1 *To get rid of scaling factors, we will henceforth multiply all edge weights in \mathcal{K}_{nn} by $nH'(0^+)$. Observe that both the optimal matching $\pi_{\mathcal{K}_{nn}}^*$ and BP estimates $\pi_{\mathcal{K}_{nn}}^k, k \geq 0$ remain unaffected.*

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

4. First step: convergence to a limiting dynamic as $n \rightarrow \infty$. In this section we deduce from the PWIT Limit Theorem that the behavior of BP when running on \mathcal{K}_{nn} “converges” as $n \rightarrow \infty$ to its behavior when running on \mathcal{T} . To turn this idea into a rigorous statement, let us encode the execution of BP as labels attached to the oriented edges of the graph. Specifically, given a geometric graph G and an integer $k \geq 0$, we define the k^{th} -step configuration of BP on G , denoted by $(G, \langle \cdot \rightarrow \cdot \rangle_G^k, \pi_G^k)$, as the labeled geometric graph obtained by setting the label of any oriented edge (v, w) in G to be the couple $(\langle v \rightarrow w \rangle_G^k, \mathbf{1}_{\{w=\pi_G^k(v)\}})$. We can now state and prove the main theorem of the present section.

THEOREM 4.1 (CONTINUITY OF BP) *Consider an almost sure realization of the PWIT limit Theorem:*

$$\mathcal{K}_{nn} \xrightarrow[n \rightarrow \infty]{a.s.} \mathcal{T}. \quad (5)$$

Then for every fixed $k \geq 0$, the k^{th} -step configuration of BP on \mathcal{K}_{nn} converges locally in probability to the k^{th} -step configuration of BP on \mathcal{T} :

$$(\mathcal{K}_{nn}, \langle \cdot \rightarrow \cdot \rangle_{\mathcal{K}_{nn}}^k, \pi_{\mathcal{K}_{nn}}^k) \xrightarrow[n \rightarrow \infty]{P} (\mathcal{T}, \langle \cdot \rightarrow \cdot \rangle_{\mathcal{T}}^k, \pi_{\mathcal{T}}^k). \quad (6)$$

PROOF. Let us (redundantly) re-label the vertices of \mathcal{K}_{nn} by words of \mathcal{V} in a manner that yields to consistent comparison between the messages on \mathcal{K}_{nn} and those on \mathcal{T} . To begin with, let the empty word \emptyset represent the root of \mathcal{K}_{nn} and words $1, 2, \dots, n$ its immediate neighbors, ordered by increasing weight of the edge connecting them to the root. Then, inductively, if word $v \in \mathcal{V}^*$ represents some vertex $x \in \mathcal{K}_{nn}$ and \dot{v} some $y \in \mathcal{K}_{nn}$, then let the words $v.1, v.2, \dots, v.(n-1)$ represent the $n-1$ neighbors of x distinct from y in \mathcal{K}_{nn} , again ordered by increasing weight of the corresponding edge. Note that this definition makes almost surely sense since the edge weights are pairwise distinct (by continuity of H). In fact, it follows from an easy induction on $v \in \mathcal{V}$ that the vertex represented by v in \mathcal{K}_{nn} is nothing but $\gamma_n^\varrho(v)$ as soon as ϱ and n are large enough, where $\gamma_n^\varrho: [\mathcal{T}]_\varrho \xrightarrow{\cong} [\mathcal{K}_{nn}]_\varrho$ is the (random) isomorphism involved in the definition of the local convergence (5). In particular,

$$\forall \{v, w\} \in \mathcal{E}, \|v, w\|_{\mathcal{K}_{nn}} \xrightarrow[n \rightarrow \infty]{a.s.} \|v, w\|_{\mathcal{T}}. \quad (7)$$

With this relabeling in hand, the desired convergence (6) can now be written:

$$\forall \{v, w\} \in \mathcal{E}, \langle v \rightarrow w \rangle_{\mathcal{K}_{nn}}^k \xrightarrow[n \rightarrow \infty]{P} \langle v \rightarrow w \rangle_{\mathcal{T}}^k \text{ and } \forall v \in \mathcal{V}, \pi_{\mathcal{K}_{nn}}^k(v) \xrightarrow[n \rightarrow \infty]{P} \pi_{\mathcal{T}}^k(v). \quad (8)$$

The recursive nature of the messages almost compels one to think of proving (8) by induction over k . The base case of $k = 0$ is trivial. However, when trying to go from step k to step $k+1$ one soon gets confronted by a major hinder: the update and decision rules (1) and (3) are not continuous with respect to local convergence. Indeed, writing:

$$\langle v \rightarrow w \rangle_{\mathcal{K}_{nn}}^{k+1} = \min_{\substack{u \in \{v.1, \dots, v.(n-1), \dot{v}\} \\ u \neq w}} \left\{ \|u, v\|_{\mathcal{K}_{nn}} - \langle u \rightarrow v \rangle_{\mathcal{K}_{nn}}^k \right\}$$

$$\text{and } \pi_{\mathcal{K}_{nn}}^k(v) = \arg \min_{u \in \{v.1, \dots, v.(n-1), \dot{v}\}} \left\{ \|u, v\|_{\mathcal{K}_{nn}} - \langle u \rightarrow v \rangle_{\mathcal{K}_{nn}}^k \right\},$$

one can not simply invoke convergence of each term inside the min and arg min to conclude, because there are unboundedly many such terms as $n \rightarrow \infty$. Remarkably enough, it turns out that under assumption A2, we can in fact restrict ourselves to a uniformly bounded number of them with probability as high as desired, as stated in the following lemma. \square

LEMMA 4.1 (UNIFORM CONTROL ON ESSENTIAL MESSAGES) *For all $v \in \mathcal{V}$ and $k \geq 0$:*

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\arg \min_{1 \leq i < n} \left\{ \|v, v.i\|_{\mathcal{K}_{nn}} - \langle v \rightarrow v.i \rangle_{\mathcal{K}_{nn}}^k \right\} \geq i_0 \right) \xrightarrow{i_0 \rightarrow \infty} 0.$$

The proof of this Lemma is long and technical and hence is deferred to Appendix A

5. Second step: analysis of BP on PWIT. In light of Theorem 4.1, one can replace the asymptotic analysis of BP on \mathcal{K}_{nn} as n becomes large by the direct study of BP's dynamics on the limiting PWIT. Formally, we are interested in the limiting behavior of the random process defined for all $v \in \mathcal{V}^*$ by the recursion:

$$\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{k+1} = \min_{i \geq 1} \{ \|v, v.i\|_{\mathcal{T}} - \langle v \rightarrow v.i \rangle_{\mathcal{T}}^k \}, \quad (9)$$

where the initial values $(\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^0)_{v \in \mathcal{V}^*}$ are i.i.d. random variables independent of \mathcal{T} (0 in the case of our algorithm). The fact that the above min is a.s. well defined despite the infinite number of terms will become clear later (see Lemma 5.4). For the time being, it is sufficient to consider it as a $\overline{\mathbb{R}}$ -valued infimum. First observe that at any given time k all $\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^k, v \in \mathcal{V}^*$ share the same distribution, owing to the natural spatial invariance of the PWIT. Moreover, if F denotes the corresponding tail distribution function at a given time, a straightforward computation (see for instance [2]) shows that the tail distribution function TF obtained after a single application of update rule (9) is given by:

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

This defines an operator T on the space \mathcal{D} of tail distribution functions of $\overline{\mathbb{R}}$ -valued random variables, i.e. non-increasing corlol² functions $F : \mathbb{R} \rightarrow [0, 1]$. T is known to have a unique fixed point (see [2]), the so-called *logistic* distribution:

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

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

5.1 Weak attractiveness. Finding the domain of attraction of F^* under operator T is not known and has been listed as an open problem by Aldous and Bandyopadhyay ([3, Open Problem # 62]). In what follows, we answer this question and more. We fully characterize the asymptotic behavior of the successive iterates $(T^k F)_{k \geq 0}$ for any initial distribution $F \in \mathcal{D}$.

First observe that T is anti-monotone with respect to pointwise order:

$$F_1 \leq F_2 \implies TF_1 \geq TF_2. \quad (10)$$

This suggests considering the non-decreasing second iterate T^2 . Unlike T , T^2 admits infinitely many fixed points. To see this, let θ_t ($t \in \mathbb{R}$) be the t -shift operator defined on \mathcal{D} by $\theta_t F : x \mapsto F(x - t)$. Then a trivial change of variable gives:

$$T \circ \theta_t = \theta_{-t} \circ T. \quad (11)$$

Therefore, it follows that $T^2(\theta_t F^*) = \theta_t(T^2 F^*) = \theta_t F^*$ for all $t \in \mathbb{R}$. That is, the $\theta_t F^*, t \in \mathbb{R}$ are fixed points of T^2 . These considerations lead us to introduce the key tool of our analysis:

DEFINITION 5.1 For $F \in \mathcal{D}$, define the transform \widehat{F} as follows :

$$\forall x \in \mathbb{R}, \widehat{F}(x) = x + \ln \left(\frac{F(x)}{1 - F(x)} \right).$$

Intuitively, \widehat{F} represents the local shift (along the X-axis) between F and F^* . Indeed, it enables us to express any $F \in \mathcal{D}$ as a locally deformed version of F^* via the following straightforward inversion formula:

$$\forall x, F(x) = F^*(x - \widehat{F}(x)) = \theta_{\widehat{F}(x)} F^*(x).$$

In particular, $\theta_{t_1} F^* \leq F \leq \theta_{t_2} F^* \iff t_1 \leq \widehat{F} \leq t_2$, and $F = \theta_t F^*$ if and only if \widehat{F} is constant on \mathbb{R} with value t . In that sense, the maximal amplitude of the variations of \widehat{F} on \mathbb{R} tells something about the distance between F and the family of fixed points $\{\theta_t F^*, t \in \mathbb{R}\}$. Thus, the action of T on those variations appears to be of crucial importance and will be at the center of our attention for the rest of this sub-section. We now state three lemmas whose proofs are given in Appendix B.

²continuous on the right, limit on the left

LEMMA 5.1 Let $F \in \mathcal{D} \setminus \{0\}$ such that $\int_0^{+\infty} F < \infty$. Then, $\widehat{T^4 F}$ is bounded on \mathbb{R} .

LEMMA 5.2 If $F \in \mathcal{D}$ is such that \widehat{F} is bounded, then \widehat{TF} is bounded too, and moreover:

$$-\sup_{\mathbb{R}} \widehat{F} \leq \inf_{\mathbb{R}} \widehat{TF} \leq \sup_{\mathbb{R}} \widehat{TF} \leq -\inf_{\mathbb{R}} \widehat{F}.$$

Further, if \widehat{F} is not constant then this contraction becomes strict under a second iteration :

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

LEMMA 5.3 Let $F \in \mathcal{D}$ be such that \widehat{F} is bounded. Then, $\widehat{T^k F}$ is continuously differentiable for $k \geq 2$, and the family of derivatives $(\widehat{T^k F})'$, $k \geq 3$ is uniformly integrable:

$$\sup_{k \geq 3} \int_{|x| > M} |(\widehat{T^k F})'(x)| dx \xrightarrow{M \rightarrow \infty} 0.$$

We are now in position to provide a complete description of the dynamics of T on \mathcal{D} .

THEOREM 5.1 (DYNAMICS OF T ON \mathcal{D}) Let $F \in \mathcal{D}$. Assume F is not the 0 function and $\int_0^{+\infty} F < +\infty$ (otherwise $(T^k F)_{k \geq 1}$ trivially alternates between the 0 and 1 functions). Then, there exists a constant $\gamma \in \mathbb{R}$ dependent on F such that $\widehat{T^{2k} F} \xrightarrow{k \rightarrow \infty} \gamma$ and $\widehat{T^{2k+1} F} \xrightarrow{k \rightarrow \infty} -\gamma$, uniformly on \mathbb{R} . In particular,

$$T^{2k} F \xrightarrow{k \rightarrow \infty} \theta_{\gamma} F^* \text{ and } T^{2k+1} F \xrightarrow{k \rightarrow \infty} \theta_{-\gamma} F^*, \text{ uniformly on } \mathbb{R}.$$

PROOF. By Lemma 5.1, one can choose a large enough $M \geq 0$ for $T^4 F$ to lie in the subspace

$$\mathcal{D}_M = \{F \in \mathcal{D}, -M \leq \widehat{F} \leq M\} = \{F \in \mathcal{D}, \theta_{-M} F^* \leq F \leq \theta_M F^*\}.$$

Lemma 5.2 guarantees the stability of \mathcal{D}_M under the action of T , so the whole sequence $(T^k F)_{k \geq 4}$ remains in \mathcal{D}_M . Even better, 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 monotone, hence convergent, say to γ^- and γ^+ respectively. All we have to show is that $\gamma^- = \gamma^+$; convergence of $(\widehat{T^{2k+1} F})_{k \geq 2}$ to the opposite constant will then simply follow from property (11).

By Arzela-Ascoli theorem, the family of (clearly bounded and 1-Lipschitz) functions $(T^{2k} F)_{k \geq 2}$ is relatively compact with respect to compact convergence. Thus, there exists a convergent sub-sequence:

$$T^{2\varphi(k)} F \xrightarrow{k \rightarrow \infty} F_{\infty}. \quad (12)$$

From the uniform continuity of $y \mapsto \ln \frac{y}{1-y}$ on every compact subset of $]0, 1[$ (Heine's theorem), it follows that the restriction of the $\widehat{\cdot}$ transform to \mathcal{D}_M is continuous with respect to compact convergence. Hence,

$$\widehat{T^{2\varphi(k)} F} \xrightarrow{k \rightarrow \infty} \widehat{F_{\infty}}.$$

Even better, the uniform integrability of variations stated in Lemma 5.3 makes the above compact convergence perfectly equivalent to uniform convergence on all \mathbb{R} . In particular,

$$\inf_{\mathbb{R}} \widehat{F_{\infty}} = \lim_{k \rightarrow \infty} \uparrow \inf_{\mathbb{R}} \widehat{T^{2\varphi(k)} F} = \gamma^- \text{ and } \sup_{\mathbb{R}} \widehat{F_{\infty}} = \lim_{k \rightarrow \infty} \downarrow \sup_{\mathbb{R}} \widehat{T^{2\varphi(k)} F} = \gamma^+. \quad (13)$$

On the other hand, a straightforward use of the dominated convergence Theorem shows that the restriction of T to \mathcal{D}_M is continuous with respect to compact convergence. Therefore, (12) implies

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

But using exactly the same arguments as above (note that γ^-, γ^+ do not depend on φ), we obtain a similar conclusion :

$$\inf_{\mathbb{R}} \widehat{T^2 F_{\infty}} = \gamma^- \text{ and } \sup_{\mathbb{R}} \widehat{T^2 F_{\infty}} = \gamma^+. \quad (14)$$

By the second part of Lemma 5.2, having both (13) and (14) implies that $\gamma^- = \gamma^+$. \square

5.2 Strong attractiveness. So far, we have established the distributional convergence of the message process. To complete the algorithm analysis, we now need to prove sample-path wise convergence. We note that Aldous and Bandyopadhyay [3, 5] have studied the special case where the i.i.d. initial messages $(\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^0)_{v \in \mathcal{V}^*}$ are distributed according to the fixed point F^* . They established L^2 -convergence of the message process to some unique stationary configuration which is independent of $(\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^0)_{v \in \mathcal{V}^*}$. They call this the *bivariate uniqueness property*. This sub-section is dedicated to extending such a property to the case of F -distributed i.i.d. initial messages, where F is any tail distribution satisfying the assumption of Theorem 5.1, namely:

$$\int_0^\infty F < \infty, \text{ or equivalently } \mathbb{E} \left[(\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^0)^+ \right] < \infty. \quad (15)$$

Recall that, if (15) does not hold, then $(T^k F)_{k \geq 1}$ simply alternates between the 0 and 1 functions. In other words, all messages in \mathcal{T} become almost surely infinite after the very first iteration. Henceforth, we will assume (15) to hold, which is in particular the case if all initial messages are set to zero. We first state a Lemma that will allow us to fix the problem of non-continuity of the update and decision rules on \mathcal{T} caused by the infinite number of terms involved in the minimization.

LEMMA 5.4 *Under assumption (15), $\pi_{\mathcal{T}}^k(v) = \arg \min_{w \sim v} \{ \|w, v\|_{\mathcal{T}} - \langle w \rightarrow v \rangle_{\mathcal{T}}^k \}$ is a.s. well defined for every $k \geq 4, v \in \mathcal{V}$ despite the infinite number of terms involved in the argmin. Moreover,*

$$\sup_{k \geq 4} \mathbb{P} \left(\arg \min_{i \geq 1} \{ \|v.i, v\|_{\mathcal{T}} - \langle v.i \rightarrow v \rangle_{\mathcal{T}}^k \} \geq i_0 \right) \xrightarrow{i_0 \rightarrow \infty} 0. \quad (16)$$

With this uniform control in hand, we are now ready to prove the strong convergence of BP on \mathcal{T} .

THEOREM 5.2 (CONVERGENCE OF BP ON \mathcal{T}) *Assume the i.i.d. initial messages satisfy (15). Then, up to some additive constant $\gamma \in \mathbb{R}$, the recursive tree process defined by (9) converges to the unique stationary configuration $\langle \cdot \rightarrow \cdot \rangle_{\mathcal{T}}^*$ in the following sense: for every $v \in \mathcal{V}^*$,*

$$\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{2k} \xrightarrow[k \rightarrow \infty]{L^2} \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^* + \gamma \quad \text{and} \quad \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{2k+1} \xrightarrow[k \rightarrow \infty]{L^2} \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^* - \gamma.$$

Further, defining $\pi_{\mathcal{T}}^*$ as the assignment induced by $\langle \cdot \rightarrow \cdot \rangle_{\mathcal{T}}^*$ according to rule (3), we have convergence of decisions at the root:

$$\pi_{\mathcal{T}}^k(\emptyset) \xrightarrow[k \rightarrow \infty]{P} \pi_{\mathcal{T}}^*(\emptyset).$$

PROOF. Denote by F the tail distribution function of the initial messages. The idea is to construct an appropriate stochastic coupling between our F -initialized message process and the F^* -initialized version and then use the endogeneity of the latter to conclude. We let γ be the constant appearing in Theorem 5.1. First, observe that the dynamics (9) are “anti-homogeneous”: if we add the same constant to every initial message, then that constant is simply added to every even message $\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{2k}$ and subtracted from every odd message $\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{2k+1}$. Therefore, without loss of generality we may 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 a classical result often termed as Strassen’s Theorem, probability measures satisfying such a stochastic ordering can always be coupled in a pointwise monotone manner. Specifically, there exists a probability space $E' = (\Omega', \mathcal{F}', P')$, possibly differing from the original space $E = (\Omega, \mathcal{F}, P)$, on which can be defined a random variable X^ε with distribution $T^{k_\varepsilon} F$ and two random variables X^- and X^+ with distribution F^* , in such a way that almost surely,

$$X^- - \varepsilon \leq X^\varepsilon \leq X^+ + \varepsilon. \quad (17)$$

Now consider the product space $(\bigotimes_{v \in \mathcal{V}} E') \otimes E$ over which we can jointly define the PWIT \mathcal{T} and independent copies $(X_v^-, X_v^\varepsilon, X_v^+)_{v \in \mathcal{V}}$ of the triple (X^-, X, X^+) for each vertex $v \in \mathcal{V}$. On \mathcal{T} , let us compare the configurations $(\langle \cdot \rightarrow \cdot \rangle_{\mathcal{T}}^{k,-})_{k \geq 0}$, $(\langle \cdot \rightarrow \cdot \rangle_{\mathcal{T}}^{k,\varepsilon})_{k \geq 0}$ and $(\langle \cdot \rightarrow \cdot \rangle_{\mathcal{T}}^{k,+})_{k \geq 0}$ resulting from three different initial conditions, namely:

$$\forall v \in \mathcal{V}^*, \begin{cases} \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{0,-} = X_v^-; \\ \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{0,\varepsilon} = X_v^\varepsilon; \\ \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{0,+} = X_v^+. \end{cases}$$

Due to anti-monotony and anti-homogeneity of the update rule (9), inequality (17) ‘propagates’ in the sense that for any $k \geq 0$ and $v \in \mathcal{V}^*$,

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

Now fix $v \in \mathcal{V}^*$. By construction,

$$\left(\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{k+k_\varepsilon} \right)_{k \geq 0} \stackrel{\mathcal{D}}{=} \left(\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{k,\varepsilon} \right)_{k \geq 0}.$$

In particular, for every $k \geq k_\varepsilon$ we have

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

But from the bivariate uniqueness property established by Aldous and Bandyopadhyay [3, 5] for the logistic distribution, it follows that

$$\sup_{t \geq k-k_\varepsilon} \left\| \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^{t,\pm} - \langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^* \right\|_{L^2} \xrightarrow[k \rightarrow \infty]{} 0.$$

Thus, the sequence $(\langle v \rightarrow \dot{v} \rangle_{\mathcal{T}}^k)_{k \geq 0}$ is Cauchy in L^2 , hence convergent. Using Lemma 5.4 to justify the interchange between limit and minimization, it is not hard to check that the limiting configuration has to be stationary, i.e. is a fixed point for the recursion (9), and that the estimates $\pi_{\mathcal{T}}^k, k \geq 0$ do in turn converge (in probability) to the estimate $\pi_{\mathcal{T}}^*$ associated with the limiting configuration. Note that endogeneity implies uniqueness of the stationary configuration, and therefore $\pi_{\mathcal{T}}^*$ is nothing but the infinite optimal assignment studied in [2]. \square

6. Third step: putting things together. Finally, we are now in position to complete the proof of Theorem 2.1, using the following remarkable result by Aldous.

THEOREM 6.1 (ALDOUS, [2]) *Let $\pi_{\mathcal{T}}^*$ be the assignment associated with the unique stationary configuration $\langle \cdot \rightarrow \cdot \rangle_{\mathcal{T}}^*$. Then $\pi_{\mathcal{T}}^*$ is a perfect matching on \mathcal{T} , and*

$$(\mathcal{K}_{nn}, \pi_{\mathcal{K}_{nn}}^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} (\mathcal{T}, \pi_{\mathcal{T}}^*). \quad (18)$$

PROOF OF THEOREM 2.1. Using Theorem 4.1 and Skorokhod’s representation Theorem, the above convergence (18) can be extended to include BP’s answer at any fixed step k :

$$(\mathcal{K}_{nn}, \pi_{\mathcal{K}_{nn}}^k, \pi_{\mathcal{K}_{nn}}^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} (\mathcal{T}, \pi_{\mathcal{T}}^k, \pi_{\mathcal{T}}^*).$$

In particular, the probability of getting a wrong decision at the root of \mathcal{K}_{nn} converges as $n \rightarrow \infty$ to the probability of getting a wrong decision at the root of \mathcal{T} : for all $k \geq 0$,

$$\mathbb{P}(\pi_{\mathcal{K}_{nn}}^k(\emptyset) \neq \pi_{\mathcal{K}_{nn}}^*(\emptyset)) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(\pi_{\mathcal{T}}^k(\emptyset) \neq \pi_{\mathcal{T}}^*(\emptyset)).$$

Finally, the symmetry of \mathcal{K}_{nn} lets us rewrite the left-hand side as the expected fraction of errors $\mathbb{E}[d(\pi_{\mathcal{K}_{nn}}^k, \pi_{\mathcal{K}_{nn}}^*)]$, and Theorem 5.2 ensures that the right-hand side vanishes as $k \rightarrow \infty$. \square

7. Conclusion. In this paper we have established that the BP algorithm finds an almost optimal solution to the random $n \times n$ assignment problem in time $O(n^2)$ with high probability. The natural lower bound of $\Omega(n^2)$ makes BP an (order) optimal algorithm. This result significantly improves over both the worst-case upper bound for exact convergence of the BP algorithm proved by Bayati, Shah and Sharma [6] and the best-known computational time achieved by Edmonds and Karp's algorithm [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 asymptotic analysis of the algorithm 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 BP.

References

- [1] D. Aldous, *Asymptotics in the random assignment problem*, Pr. Th. Rel. Fields **93** (1992), 507–534.
- [2] _____, *The $\zeta(2)$ limit in the random assignment problem*, RSA **18** (2001), 381–418.
- [3] D. Aldous and A. Bandyopadhyay, *A survey of max-type recursive distributional equations*, Annals of Applied Probability **15(2)** (2005), 1047–1110.
- [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*, Tech. 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 (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** (1999), 113–144.
- [9] M. E. Dyer, A.M. Frieze, and C.J.H. McDiarmid, *On linear programs with random costs*, Math. Programming **35** (1986), 3–16.
- [10] J. Edmonds and R. Karp, *Theoretical improvements in algorithmic efficiency for network flow problems*, Jour. of the ACM **19** (1972), 248–264.
- [11] M.X. Goemans and M.S. Kodialam, *A lower bound on the expected cost of an optimal assignment*, Math. Oper. Res. **18** (1993), 267–274.
- [12] R. M. Karp, *An algorithm to solve the $m \times n$ assignment problem in expected time $O(mn \log n)$* , Networks **10** (1980), 143–152.
- [13] _____, *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 (1987), 1–4.
- [14] A. Lazarus, *Certain expected values in the random assignment problem*, Op. Res. Let. **14** (1993), 207–214.
- [15] S. Linusson and J. Wästlund, *A proof of Parisi's conjecture on the random assignment problem*, Probab. Theory Relat. Fields **128** (2004), 419–440.
- [16] M. Mézard and G. Parisi, *On the solution of the random link matching problem*, Journal of Physics **48** (1987), 1451–1459.
- [17] C. Nair, B. Prabhakar, and M. Sharma, *Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures*, Random Structures Algorithms **27(4)** (2005), 413–444.
- [18] B. Olin, *Asymptotic properties of random assignment problems*, PhD thesis, Kungl Tekniska Högskolan, Stockholm, Sweden (1992).

- [19] J. Pearl, *Probabilistic reasoning in intelligent systems: Networks of plausible inference*, Morgan Kaufmann, San Francisco, CA, 1988.
- [20] D. W. Walkup, *On the expected value of a random assignment problem*, SIAM J. Comp. **8** (1979), 440–442.
- [21] J. Yedidia, W. Freeman, and Y. Weiss, *Generalized belief propagation*, Tech. Report TR-2000-26, Mitsubishi Elect. Res. Lab., Cambridge, MA, 2000.

Appendix A. Proof of Lemma 4.1. The proof of Lemma 4.1 lays upon two technical lemmas stated below. Essentially, the picture is the following: when i gets large, the length $\|v, v.i\|_{\mathcal{K}_{nn}}$ of the i^{th} shortest edge attached to v in \mathcal{K}_{nn} becomes large too (Lemma A.1), whereas the message $\langle v.i \rightarrow v \rangle_{\mathcal{K}_{nn}}^k$ passing along that edge remains reasonably small (Lemma A.2). Therefore, the resulting contribution $\|v.i, v\|_{\mathcal{K}_{nn}} - \langle v.i \rightarrow v \rangle_{\mathcal{K}_{nn}}^k$ is too large to matter in the minimization. In what follows, $|v|$ will denote the number of letters of the word $v \in \mathcal{V}$, and $v_1, \dots, v_{|v|}$ its consecutive letters (e.g. if $v = 1.2.1.3$ then $|v| = 4$ and $v_1 = 1, v_2 = 2, v_3 = 1, v_4 = 3$). Also, we will write $v_{\leq h}$ for the prefix $v_1 \dots v_h$.

LEMMA A.1 (UNIFORM CONTROL ON EDGE WEIGHTS) *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 all n large enough for \mathcal{K}_{nn} to contain $v.i$,*

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

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

$$\text{card}\left\{y \sim x_k, y \neq x_{k-1}, \|x_k, y\|_{\mathcal{K}_{nn}} \leq t\right\} \geq i - 1. \quad (19)$$

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 definition of our relabeling, the number of edges in \mathcal{K}_{nn} that are incident to $v_{\leq p-1}$ 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 $p \in \{1, \dots, |v|\}$ such that

$$v_p - \left\lceil \frac{k}{2} \right\rceil \leq \text{card}\left\{y \notin \{x_1, \dots, x_k\}, \|x_j, y\|_{\mathcal{K}_{nn}} < \|x_j, x_{j+1}\|_{\mathcal{K}_{nn}}\right\} \leq v_p. \quad (20)$$

The $\left\lceil \frac{k}{2} \right\rceil$ above comes from the fact that only half of the x_1, \dots, x_k are neighbors of x_j in \mathcal{K}_{nn} . We thus have shown that

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

where the event $A_{n,x}$ corresponds to (19) and the event $B_{n,x}^j$ to (20). The summation in the above inequality is over all possible cycle-free paths $x = (x_0, \dots, x_k)$ starting from the root in \mathcal{K}_{nn} . Now since all the edges involved are pairwise distinct, the events $A_{n,x}, B_{n,x}^0, \dots, B_{n,x}^{k-1}$ are independent. Moreover,

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

where we have used assumption A1 to define $\alpha = \frac{1}{H'(0)} \sup_{\varrho \in \mathbb{R}^+} \frac{H(\varrho)}{\varrho} < +\infty$. This yields the first bound since there are less than n^k cycle-free paths $x = (x_0, \dots, x_k)$ starting from the root in \mathcal{K}_{nn} . For the second

one, the event $A_{n,x}$ is simply replaced by $\text{card}\{y \sim x_k, y \neq x_{k-1}, \|x_k, y\|_{\mathcal{K}_{nn}} \leq t\} \leq 1$, whose probability is straightforwardly exponentially bounded using assumption A2. \square

LEMMA A.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 $t \in \mathbb{R}^+$, uniformly in n (as long as n is large enough so that $v \in \mathcal{K}_{nn}$),*

$$\mathbb{P}\left(|\langle v \rightarrow \dot{v} \rangle_{\mathcal{K}_{nn}}^k| \geq t\right) \leq M_{k,|v|} e^{-\beta_{k,|v|} t}. \quad (21)$$

PROOF. The proof is by induction over k . The base case of $k = 0$ follows trivially. Now, assume (21) is true for a given $k \in \mathbb{N}$. By Lemma A.1 we can write for all $v \in \mathcal{V}^*$ and $t \in \mathbb{R}^+$:

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

The other side is slightly harder to obtain. Again by Lemma A.1 :

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

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

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

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

$$\frac{1}{t} \log f(t) \xrightarrow{t \rightarrow +\infty} \gamma - \beta_{k,|v|+1} \quad \text{and} \quad \frac{1}{t} \log g(t) \leq -\gamma + \frac{1}{t} \log \sum_{i=1}^{\infty} \frac{(\alpha \delta e^{\alpha \delta} i)^i}{i!}.$$

Therefore, choosing any $\gamma < \beta_{k,|v|+1}$ is enough to ensure (i), and taking δ small enough for $\alpha \delta e^{\alpha \delta - 1} < 1$ will guarantee (ii) since the right-hand summand is equivalent to $\frac{(\alpha \delta e^{\alpha \delta - 1})^i}{\sqrt{2\pi i}}$ by Stirling's formula. \square

We now know enough to prove Lemma 4.1.

PROOF OF LEMMA 4.1. Set $\delta > 0$ small enough to ensure $\alpha \delta e^{\alpha \delta - 1} < 1$. Then, for $t \in \mathbb{R}^+$,

$$\begin{aligned} &\mathbb{P}\left(\arg \min_{1 \leq i < n} \{\|v, v.i\|_{\mathcal{K}_{nn}} - \langle v \rightarrow v.i \rangle_{\mathcal{K}_{nn}}^k\} \geq i_0\right) \\ &\leq \mathbb{P}\left(\langle v \rightarrow \dot{v} \rangle_{\mathcal{K}_{nn}}^{k+1} \geq t\right) + \sum_{i=i_0}^{n-1} \mathbb{P}\left(\|v, v.i\|_{\mathcal{K}_{nn}} \leq \delta i\right) + \sum_{i=i_0}^{n-1} \mathbb{P}\left(\langle v.i \rightarrow v \rangle_{\mathcal{K}_{nn}}^k \geq \delta i - t\right) \\ &\leq M_{k+1,|v|} e^{-\beta_{k+1,|v|} t} + M_{|v|} \sum_{i=i_0}^{\infty} \frac{(\alpha \delta e^{\alpha \delta} i)^i}{i!} + M_{k,|v|+1} \sum_{i=i_0}^{\infty} e^{-\beta_{k,|v|+1}(\delta i - t)}, \end{aligned}$$

by Lemmas A.1 and A.2. Letting $i_0 \rightarrow \infty$ and finally $t \rightarrow \infty$ yields the desired result. \square

Appendix B. Proof of the Lemmas in Section 5. Here we prove Lemmas 5.1, 5.2, 5.3 and 5.4.

PROOF OF LEMMA 5.1. From $\int_0^{+\infty} F < \infty$ and the definition of $TF: x \mapsto e^{-\int_{-x}^{+\infty} F}$, it follows that:

(i) as $x \rightarrow +\infty$, $TF(x) = \Theta\left(e^{-\int_{-x_0}^{x_0} F}\right)$ for any fixed $x_0 \in \mathbb{R}$;

(ii) as $x \rightarrow -\infty$, $TF(x) = 1 - \Theta\left(\int_{-x}^{+\infty} F\right)$.

Now since F is non-zero and non-increasing, there exists $\alpha, \beta > 0$ (simply take $\beta = 1$) such that, for all small enough $x \in \mathbb{R}$, $\alpha \leq F(x) \leq \beta$. Replacing these inequalities into (i) above yields :

$$\text{as } x \rightarrow +\infty, TF(x) = O(e^{-\alpha x}) \text{ and } TF(x) = \Omega(e^{-\beta x}). \quad (22)$$

In particular, TF satisfies the assumptions made on F , so by induction $T^k F, k \geq 2$ also do, and we may therefore iteratively apply (i)/(ii) to $TF, T^2 F$ and $T^3 F$. This successively yields:

$$\text{as } x \rightarrow -\infty, \quad T^2 F(x) = 1 - O(e^{\alpha x}) \text{ and } T^2 F(x) = 1 - \Omega(e^{\beta x}); \quad (23)$$

$$\text{as } x \rightarrow +\infty, \quad T^3 F(x) = \Theta(e^{-x}); \quad (24)$$

$$\text{as } x \rightarrow -\infty, \quad T^4 F(x) = 1 - \Theta(e^x). \quad (25)$$

Replacing F by TF , we see that (24) also holds for $T^4 F$, so we end up with $T^4 F(x)$ being both $\Theta(e^{-x})$ as $x \rightarrow +\infty$ and $1 - \Theta(e^x)$ as $x \rightarrow -\infty$. Besides, on any compact set, $T^4 F$ takes values within a compact subset of $]0, 1[$ by monotonicity. Hence the boundedness of $\widehat{T^4 F}: x \mapsto x + \ln\left(\frac{T^4 F(x)}{1 - T^4 F(x)}\right)$ over \mathbb{R} . \square

PROOF OF LEMMA 5.2. It follows from the properties (10) and (11) of T that for every $m, M \in \mathbb{R}$,

$$\theta_m F^* \leq F \leq \theta_M F^* \implies \theta_{-M} F^* \leq TF \leq \theta_{-m} F^*.$$

Once rewritten in terms of the $\widehat{\cdot}$ transform, this becomes:

$$m \leq \widehat{F} \leq M \implies -M \leq \widehat{TF} \leq -m,$$

and the desired inequalities follow by taking $m = \inf_{\mathbb{R}} \widehat{F}$ and $M = \sup_{\mathbb{R}} \widehat{F}$. Now assume \widehat{F} is not constant on \mathbb{R} . The right-continuity (of F and hence) of \widehat{F} ensures existence of an open interval (a, b) such that $M' = \sup_{(a,b)} \widehat{F} < \sup_{\mathbb{R}} \widehat{F} = M$. Then, for $x \geq -a$,

$$\begin{aligned} TF(x) &= \exp\left(-\int_{-x}^{+\infty} F\right) \geq \exp\left(-\int_{-x}^a \theta_M F^* - \int_a^b \theta_{M'} F^* - \int_b^{+\infty} \theta_M F^*\right) \\ &= \kappa \times \theta_{-M} F^*(x) \text{ with } \kappa = \exp\left(\int_a^b (\theta_M F^* - \theta_{M'} F^*)\right) > 1. \end{aligned}$$

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

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

where $\kappa' = (\theta_M F^*(a))^{\kappa-1} < 1$. Now, simply observing that both $(\theta_M F^*(x))^{\kappa}$ and $\kappa' \times \theta_M F^*(x)$ are strictly less than $\theta_M F^*(x)$ is already enough for claiming that $\widehat{T^2 F}(x) < M$ for all $x \in \mathbb{R}$. In order to conclude that $\sup_{\mathbb{R}} \widehat{T^2 F} < M$, we only need to check that the inequality remains strict at $\pm\infty$:

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

The inequality $\inf_{\mathbb{R}} \widehat{T^2 F} > \inf_{\mathbb{R}} \widehat{F}$ can be obtained in exactly the same way; we skip the details. \square

PROOF OF LEMMA 5.3. Fix $k \geq 2$. From the inequality $|e^a - e^b| \leq |a - b|$ for all $a, b \leq 0$, and the fact that $0 \leq T^{k-2}F \leq 1$, it follows that $T^{k-1}F: x \mapsto \exp(-\int_x^\infty T^{k-2}F)$ is Lipschitz continuous with Lipschitz constant 1. Therefore, $T^kF: x \mapsto \exp(-\int_x^\infty T^{k-1}F)$ is differentiable on \mathbb{R} and for all $x \in \mathbb{R}$,

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

Hence, $\widehat{T^kF}: x \rightarrow x + \ln\left(\frac{T^kF(x)}{1-T^kF(x)}\right)$ is (continuously) differentiable on \mathbb{R} , and for all $x \in \mathbb{R}$,

$$(\widehat{T^kF})'(x) = \frac{1 - T^kF(x) - T^{k-1}F(-x)}{1 - T^kF(x)}. \quad (26)$$

It now remains to check the uniform integrability of $\{(\widehat{T^kF})', k \geq 3\}$. Recall that Lemma 5.2 ensures uniform boundedness of the family $\{\widehat{T^kF}, k \geq 0\}$. In other words, there exists $M \geq 0$ such that:

$$\forall k \geq 0, \theta_{-M}F^* \leq T^kF \leq \theta_MF^*.$$

Plugging it into (26) immediately yields the uniform bound $|(\widehat{T^kF})'(x)| \leq \frac{e^{2M}-1}{1+e^{x+M}}$, which is enough for uniform integrability on $(0, +\infty)$. For $(-\infty, 0)$ now, observe that the numerator in (26) vanishes as $x \rightarrow -\infty$ and is a continuously differentiable function of x as soon as $k \geq 3$, with derivative

$$x \mapsto T^{k-1}F(-x)(T^kF(x) - T^{k-2}F(x)).$$

Therefore, for all $k \geq 3$ and $x \in \mathbb{R}$,

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

Now, the above integrand is $O(e^{2u})$ as $u \rightarrow -\infty$, so the integral is $O(e^{2x})$ as $x \rightarrow -\infty$, whereas the denominator in (26) remains always above $1 - \theta_MF^*(x) = \Theta(e^x)$ as $x \rightarrow -\infty$. Thus the resulting bound on $\sup_{k \geq 3} |(\widehat{T^kF})'(x)|$ is $O(e^x)$ as $x \rightarrow -\infty$, which is enough for uniform integrability on $(-\infty, 0)$. \square

PROOF OF LEMMA 5.4. By the definition of \mathcal{T} and the fact that the k -step messages sent to v by all its children are i.i.d., we find that for every $i \geq 2$,

$$\mathbb{P}(\|v.i, v\|_{\mathcal{T}} - \langle v.i \rightarrow v \rangle_{\mathcal{T}}^k \leq \|v.1, v\|_{\mathcal{T}} - \langle v.1 \rightarrow v \rangle_{\mathcal{T}}^k) = \mathbb{P}(\xi_{i-1} \leq X_k - Y_k), \quad (27)$$

where X_k and Y_k are i.i.d. with distribution T^kF and $(\xi_i)_{i \geq 1}$ is a Poisson point process with rate 1 independent of X_k, Y_k . Now, observe that

$$\sum_{i=1}^{\infty} \mathbb{P}(\xi_i \leq X_k - Y_k) = \sum_{i=1}^{\infty} \mathbb{E} \left[\int_0^{(X_k - Y_k)^+} e^{-x} \frac{x^{i-1}}{(i-1)!} dx \right] = \mathbb{E} \left[(X_k - Y_k)^+ \right] \leq E[|X^*|] + \sup_{\mathbb{R}} |\widehat{T^kF}|, \quad (28)$$

where X^* is an F^* -distributed random variable. It follows from the previous sub-section that $\sup_{\mathbb{R}} |\widehat{T^kF}| < +\infty$ as soon as $k \geq 4$, so we can apply the Borel-Cantelli Lemma to get that

$$\|v.i, v\|_{\mathcal{T}} - \langle v.i \rightarrow v \rangle_{\mathcal{T}}^k \geq \|v.1, v\|_{\mathcal{T}} - \langle v.1 \rightarrow v \rangle_{\mathcal{T}}^k \text{ for all large enough } i$$

with probability one, and hence the argmin is well defined. Even better, the boundedness of $\widehat{T^kF}$ derived in the previous sub-section is in fact uniform in $k \geq 4$, and this is enough for (16) to hold. \square