have been proposed: MPLP [8], tree-reweighted max-product (TRMP) [28] or max-sum diffusion (MSD) [31]. As shown in [18], these algorithms provide bounds for the log-partition function and, under a suitable schedules of the updates, are guaranteed to converge to the optimum solution if there is a unique optimum. In this paper, we consider the vertex cover and matching problems and propose a new line of research leading to simple parallel BP algorithms for these problems.

We first review some theoretical works about the performances of BP algorithms for combinatorial optimization problems including matching, independent set and network flow. In [2], the max-product BP algorithm is shown to find in pseudo-polynomial time a maximum weight matching in a bipartite graph provided that the optimal matching is unique. [1] and [22] generalize this result by establishing convergence and correctness of the max-product BP when the LP relaxation has a unique optimum and this optimum is integral. [22] also shows that when this condition is not satisfied then max-product BP will give useless estimates for some edges. By setting all the weights to one, the results of [1], [22] apply to our setting of maximum cardinality matching: max-product BP converges and is correct only when the graph has a unique maximum matching which is optimum for the LP problem.

For the vertex cover problem, a one-sided relation between LP relaxation and BP is established: [23] shows that for the maximum weight independent set problem, if the max-product BP algorithm (started from the natural initial condition) converges then it is correct and the LP problem has a unique integral solution. Since a subset of vertices is a vertex cover if and only if its complement is an independent set, by setting all the weights to one, results of [23] apply to the minimum cardinality vertex cover: the tightness of the LP relaxation is necessary for the max-product optimality but it is not sufficient.

We should stress that [1] and [22] deal with a generalization of the matching problem namely with b-matchings. Also [6] extends [2] and analyzes the max-product BP applied to the minimum-cost network flow problem. In [2], [1], [22], [23] or [6], a crucial assumption is required for convergence and correctness of BP: uniqueness of the optimum solution. For the minimum cardinality vertex cover and maximum cardinality matching problems, this assumption is very restrictive.

In this paper, we will overcome this difficulty by a different approach: we introduce annealing, i.e. we study a relaxed version of the optimization problem parametrized by the parameter $z > 0$ (sometimes called the inverse temperature) such that in the limit $z \to \infty$, we recover the original optimization problem. Since the work of Heilmann and Lieb [11], it is well known that the matching problem has the 'correlation decay' property at positive temperature, i.e. for $z < \infty$. Building on [3], [15], [21], we can show that this property ensures the convergence of our BP algorithm on any finite graph as long as $z < \infty$. However, we are only interested in the limit $z \to \infty$. Our first main contribution shows that our BP algorithm computes in the limit $z \to \infty$ a maximum fractional matching. For the minimum vertex cover, this approach seems doomed to fail. It is well-known that there is no 'correlation decay' at low temperature (i.e. as $z \to \infty$) for the independent set problem making it very hard or even impossible to relate the minimum vertex cover and its relaxed version with $z < \infty$. However, the fractional vertex cover problem is the dual of the fractional matching problem. As a consequence, we will show our second main contribution: our BP algorithm computes in the limit $z \to \infty$ a minimum half-integral vertex cover, hence providing a 2-approximation. Surprisingly, if the graph is bipartite, it computes a minimum vertex cover.

To the best of our knowledge, our results are the first rigorous results in the regime $z \to \infty$ on arbitrary graphs showing the performances of BP algorithm. A similar approach based on the Bethe approximation was proposed in [4] but no convergence results were given. As noted in [30], [7], if convergence is proved, then it is easy to see that our BP algorithm solves the LP relaxation of the combinatorial optimization problem. Our paper shows rigorously that this approach is successful for the minimum vertex cover problem and the maximum matching problem. Also related to our approach is [27] which deals with the sum of weighted perfect matchings in complete bipartite graphs and shows that the Bethe free entropy is concave which easily implies that the same is true in our setting. We note that the Bethe approximation is only used in the analysis of BP for the maximum matching problem and not for the minimum vertex cover. The main technical contribution of this paper is in the analysis of the minimum vertex cover problem. As explained above, a naive direct approach studying BP for vertex cover fails. Instead we made a careful analysis of the BP algorithm for the dual relaxed problem (namely fractional matching) in order to be able to get results for the original vertex cover problem. This analysis requires original techniques based on subtle monotonicity arguments for the local iteration.

We end this introduction by a last motivation for this work. The increasing need to reason about large-scale graph-structured data in machine learning and data mining has driven the development of new graph-parallel abstractions such as Pregel [17], Graphlab [16] and Powergraph [10] that encode computation as vertex-programs which run in parallel and interact along edges in the graph. In this setting, BP algorithms present several opportunities for parallelism: given the messages from the previous iteration, each new message can be computed completely independently and in any order. BP algorithms are certainly natural candidates to leverage the performance and scalability of graph-parallel abstractions. Recent parallel implementations of BP [9] show promising empirical results in this direction and our work makes a significant step towards a better understanding of BP algorithms that could be extended to other optimization problems.

We present our results in the next Section. We first start in Section 2.1 by introducing the two combinatorial optimization problems studied in this paper: matching and vertex cover. We introduce our annealing BP and show its convergence for general graphs. We then introduce a simpler version of BP (corresponding to the standard max-product version) and relate it to our annealing BP. We show that it allows us to compute minimum fractional vertex cover for any graph. In Section 2.2, we show that for bipartite graphs, our algorithm computes a minimum vertex cover. In Section 2.3, we use variational techniques to analyze BP and give exact loop series expansion as developed in [5]. We conclude in Section 3. In the Appendix 4, we provide the detailed proofs.

2 Results

2.1 (Fractional) matching and vertex cover numbers

We consider a graph $G = (V, E)$. We denote by the same symbol ∂v the set of neighbors of node $v \in V$ and the set of edges incident to v. A matching is encoded by a binary vector $\mathbf{B} = (B_e,\, e \in E) \in \{0,1\}^E$ defined by $B_e = 1$ if and only if the edge e belongs to the matching. We have for all $v \in V$, $\sum_{e\in\partial v} B_e \leq 1$. The size of the matching is given by $\sum_e B_e$. For a finite graph G, we define the matching number of G as $\nu(G) = \max\{\sum_e B_e\}$ where the maximum is taken over matchings of G. Similarly a vertex cover is encoded by a binary vector $\mathbf{C} = (C_v,\ v \in V) \in \{0,1\}^V$ defined by $C_v = 1$ if and only if the vertex v belongs to the vertex cover. We have for all $e = (uv) \in E$, $C_u + C_v \geq 1$. The size of the vertex cover is given by $\sum_v C_v$ and the vertex cover number of G is $\tau(G) = \min\{\sum_v C_v\}$ where the minimum is taken over vertex covers of G.

The matching number is the solution of the following binary Integer Linear Program (ILP):

$$\begin{aligned}\nu(G) &= \max \sum_{e \in E} x_e \\ &\quad \text{s.t.} \quad \sum_{e \in \partial v} x_e \leq 1,\ \forall v \in V;\ x_e \in \{0,1\},\end{aligned}$$

and the vertex cover number is the solution of the following ILP:

$$\begin{aligned}\tau(G) &= \min \sum_{v \in V} y_v \\ &\quad \text{s.t.} \quad y_u + y_v \geq 1,\ \forall (uv) \in E;\ y_v \in \{0,1\},\end{aligned}$$

The straightforward Linear Programming (LP) relaxation of these ILP is formed by replacing $x_e \in \{0,1\}$ (resp. $y_v \in \{0,1\}$) by $x_e \in [0,1]$ (resp. $y_v \in [0,1]$).

We define the fractional matching polytope:

$$FM(G) = \left\{ \mathbf{x} \in \mathbb{R}^E,\ x_e \geq 0,\ \sum_{e \in \partial v} x_e \leq 1 \right\}, \tag{1}$$

and the fractional matching number

$$\nu(G) \leq \nu^*(G) = \max_{\mathbf{x} \in FM(G)} \sum_{e \in E} x_e. \tag{2}$$

Similarly, we define the fractional vertex cover polytope:

$$FVC(G) = \left\{ \mathbf{y} \in \mathbb{R}^V,\ 0 \leq y_v \leq 1,\ y_u + y_v \geq 1,\ \forall (uv) \in E \right\}, \tag{3}$$

and the fractional vertex cover number is

$$\tau(G) \geq \tau^*(G) = \min_{\mathbf{y} \in FVC(G)} \sum_v y_v. \tag{4}$$

By linear programming duality, we have $\tau^*(G) = \nu^*(G)$ (see Section 64.6 in [24]). Note however that computing the matching number $\nu(G)$ can be done in polynomial time whereas determining the vertex cover number $\tau(G)$ is NP-complete (Corollary 64.1a in [24]).

We now define our associated BP message passing algorithm: LABP (Loopy Annealing BP). We introduce the set $\overrightarrow{E}$ of directed edges of G comprising two directed edges $u \to v$ and $v \to u$ for each undirected edge $uv \in E$. For $\overrightarrow{e} \in \overrightarrow{E}$, we denote by $-\overrightarrow{e}$ the edge with opposite direction. With a slight abuse of notation, we denote by ∂v the set of incident edges to $v \in V$

How is this different

directed towards v. The update rules of LABP depend on a parameter $z > 0$ and are defined by $m^0_{\vec{e}} = 0$ and for $t \geq 0$ and all u, v neighbors in G:

$$\boxed{m^{t+1}_{u\to v}(z) = \frac{z}{1 + \sum_{w\in\partial u\backslash v} m^t_{w\to u}(z)},} \tag{5}$$

where $\partial u \backslash v$ is the set of neighbors of u in G from which we removed v and with the convention that the sum over the empty set equals zero. We denote by $z\mathcal{R}_G$ the mapping sending $\mathbf{m}^t(z) \in [0,\infty)^{\vec{E}}$ to $\mathbf{m}^{t+1}(z) = z\mathcal{R}_G(\mathbf{m}^t(z))$. We also denote by $z\mathcal{R}_{\vec{e}}$ the local update rule (5): $m^{t+1}_{\vec{e}}(z) = z\mathcal{R}_{\vec{e}}(\mathbf{m}(z))$.

Theorem 1. *For any finite graph G and $z > 0$, LABP converges: $\lim_{t\to\infty} m^t_{\vec{e}}(z) = Y_{\vec{e}}(z)$. For $z > 0$, let $\mathbf{x}(z) \in \mathbb{R}^E$ be defined by*

$$x_e(z) = \frac{Y_{\vec{e}}(z) Y_{-\vec{e}}(z)}{z + Y_{\vec{e}}(z) Y_{-\vec{e}}(z)} \in (0,1). \tag{6}$$

For $z > e^{|E|}$, we have $\mathbf{x}(z) = \mathbf{x}^*$ and $\mathbf{x}^* \in \mathbb{R}^E$ is a maximum fractional matching of G. In particular, we have

$$\sum_{e\in E} x^*_e = \nu^*(G) = \tau^*(G).$$

Example 1. *Consider the cycle with 3 nodes. Then, $Y(z)$ is the same for all edges and has to satisfy $Y(z) = z(1+Y(z))^{-1}$ so that we get $Y(z) = \sqrt{1/4 + z} - 1/2$. Finally, we find that the expression above equals $\nu^*(G) = 3/2$.*

We now define a much simpler message passing algorithm and a simpler expression for $\nu^*(G)$. Given a set of $\{0,1\}$-valued messages $\mathbf{I}$, we define a new set of $\{0,1\}$-valued messages by:

$$J_{u\to v} = \mathbf{1}\left(\sum_{\ell\in\partial u\backslash v} I_{w\to u} = 0\right), \tag{7}$$

with the convention that the sum over the empty set equals zero. We denote by $\mathcal{P}_G$ the mapping sending $\mathbf{I} \in \{0,1\}^{\vec{E}}$ to $\mathbf{J} = \mathcal{P}_G(\mathbf{I})$ and as above, $\mathcal{P}_{\vec{e}}$ denotes the local update rule. Note that $\mathcal{P}_G$ corresponds to the max-product algortihm presented in [22] with all weights equal to one. We define for each $v \in V$ and $\mathbf{I} \in \{0,1\}^{\vec{E}}$,

$$F_v(\mathbf{I}) = 1 \wedge \left(\sum_{u\in\partial v} I_{u\to v}\right) + \left(1 - \sum_{u\in\partial v} I_{v\to u}\right)^+, \tag{8}$$

where $a \wedge b = \min(a,b)$ and $(a)^+ = \max(a,0)$. The second part of the following theorem corresponds to Proposition 3.5 in [14] applied in our setting.

Theorem 2. *For any graph G, if $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$ then the vector $(\frac{F_v(\mathbf{I})}{2},\ v \in V)$ is a fractional vertex cover of G. Moreover, we have*

$$\nu^*(G) = \inf_{\mathbf{I}} \sum_{v\in V} \frac{F_v(\mathbf{I})}{2}, \tag{9}$$

where the infimum is over the solutions of $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$.

By [22], if the LP relaxation (2) has a unique optimum and this optimum is integral, then iterating the map $\mathcal{P}_G$ will allow us to find the unique solution to the fixed point equation $\mathbf{I} = \mathcal{P}_G(\mathbf{I})$. Indeed in this case, [22] shows that the following rule allows us to find the maximum matching from the messages $\mathbf{I}$: put edge e in the matching if and only if $I_{\vec{e}} + I_{-\vec{e}} = 2$. Note that we can then derive a minimum vertex cover from a maximum matching in linear time (Theorem 16.6 in [24]).

We now show that LABP allows us to find $\mathbf{I}$ achieving the minimum in (9) and a minimum fractional vertex cover without any restriction on G.

Proposition 1. *Let $\mathbf{I}^Y$ be the $\{0,1\}$-valued messages defined by $I^Y_{\vec{e}} = 1$ if and only if $\lim_z Y_{\vec{e}}(z) = \infty$. Then $(F_v(\mathbf{I}^Y)/2,\ v \in V)$ is a minimum half-integral vertex cover, i.e. $2\nu^*(G) = \sum_v F_v(\mathbf{I}^Y)$. In particular, $(F_v(\mathbf{I}^Y),\ v \in V)$ is a 2-approximate solution to vertex cover on G.*

Recall that if the unique games conjecture is true, then vertex cover cannot be approximated within any constant factor better than 2 as shown by [12].

Example 2. *Consider the cycle with 3 nodes. Then $I^Y_{\vec{e}} = 1$ for all oriented edges and $F_v(\mathbf{I}^Y) = 1$ for all $v \in V$. Note that $\mathbf{I}^Y$ is not the only fixed point to $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$, the all zeros vector is also a solution. However the map $\mathcal{P}_G$ has no fixed point and max-product BP as defined in [22] does not converge.*

2.2 Bipartite graphs

We now specialize our results to bipartite graphs. If the graph is bipartite then the fractional matching polytope is indeed the matching polytope, i.e. the convex hull of the incidence vectors of matchings (Corollary 18.1(b) in [24]) so that $\nu^*(G) = \nu(G)$. By König's matching theorem (Theorem 16.2 in [24]), we also have in this case $\nu(G) = \tau(G)$. To summarize, a direct application of Theorem 1 gives:

Corollary 1. *If G is bipartite, LABP computes the matching number which is equal to the vertex cover number.*

We now show that for any bipartite graph $G = (V = U \cup W, E)$, LABP allows us to define a minimum vertex cover. For any $\mathbf{I} \in \{0,1\}^{\vec{E}}$, we consider the following subset $V(\mathbf{I})$ of vertices defined differently for vertices in U and W as follows:

$$\text{for } u \in U,\ u \in V(\mathbf{I}) \quad \Leftrightarrow \quad \sum_{v \in \partial u} I_{v \to u} \geq 1, \tag{10}$$

$$\text{for } w \in W,\ w \in V(\mathbf{I}) \quad \Leftrightarrow \quad \sum_{v \in \partial w} \mathcal{P}_{v \to w}(\mathbf{I}) \geq 2. \tag{11}$$

Proposition 2. *For any bipartite graph, the subset of vertices $V(\mathbf{I}^Y)$ is a minimum vertex cover, where $\mathbf{I}^Y$ was defined in Proposition 1.*

Example 3. *Consider the cycle with 4 nodes. Again, we have $Y(z) = \sqrt{1/4 + z} - 1/2$ and the all-one vector is a fixed point of $\mathcal{P}_G \circ \mathcal{P}_G$. We see that if we apply the results of the previous section, we have $F_v(\mathbf{I}^Y) = 1$ for all $v \in V$ and we obtain a minimum fractional vertex cover. The above procedure (10) and (11) gives instead a minimum vertex cover. Note also that $\mathcal{P}_G$ has no fixed point so that the max-product BP of [22] does not converge.*

2.3 Results at positive temperature

In this section, we consider general graphs and LABP for finite z. We introduce the family of probability distributions on the set of matchings parametrised by a parameter $z > 0$:

$$\mu_G^z(\mathbf{B}) = \frac{z^{\sum_e B_e}}{P_G(z)}, \tag{12}$$

where $P_G(z) = \sum_{\mathbf{B}} z^{\sum_e B_e} \prod_{v \in V} \mathbf{1}\left(\sum_{e \in \partial v} B_e \leq 1\right)$. For any finite graph, when z tends to infinity, the distribution μ_G^z converges to the uniform distribution over maximum matchings so that we have

$$\nu(G) = \lim_{z \to \infty} \sum_{e \in E} \mu_G^z(B_e = 1). \tag{13}$$

In statistical physics, this model is known as the monomer-dimer model and its analysis goes back to the work of Heilmann and Lieb [11], see also [26] for a recent contribution in theoretical computer science. The recursion (5) has a natural interpretation in term of the probability distribution (12) when the graph G is a tree (see point (iii) in Proposition 6). The fact that this recursion is still useful for arbitrary graphs and moreover allows us to study not only matching but vertex cover is highly surprising.

In the rest of this section, we introduce the Bethe approximation which is a standard approach to approximate the probability distribution (12). This approach will give results only for the matching problem. The proofs of our results for the vertex cover problem do not rely on this approximation and requires original techniques based on subtle monotonicity arguments for the recursion (5) wich are presented in the Appendix 4.

We define the internal energy $U(z)$ and the canonical entropy $S(z)$ as:

$$\begin{aligned} U_G(z) &= -\sum_{e \in E} \mu_G^z(B_e = 1), \\ S_G(z) &= -\sum_{\mathbf{B}} \mu_G^z(\mathbf{B}) \ln \mu_G^z(\mathbf{B}). \end{aligned}$$

The free entropy $\Phi_G(z)$ is then defined by

$$\Phi_G(z) = -U_G(z) \ln z + S_G(z).$$

A more conventional notation in the statistical physics literature corresponds to an inverse temperature $\beta = \ln z$. A simple computation shows that:

$$\Phi_G(z) = \ln P_G(z).$$

Let $D(G)$ be the set of distribution over matchings, i.e. $\mu \in D(G)$ if and only if $\mu(\mathbf{B}$ is a matching in $G) = 1$. Let $\mu_G \in D(G)$. For any $e \in E$, we define $\mu_{[G,e]}$ the marginal of μ_G restricted to e, i.e.

$$\mu_{[G,e]}(1) = 1 - \mu_{[G,e]}(0) = \mu_G(B_e = 1) = \sum_{\mathbf{B}, B_e = 1} \mu_G(\mathbf{B}).$$

Similarly for any $v \in V$, we define $\mu_{[G,\partial v]}$ the marginal of μ_G restricted to $\partial v \subset E$. For any $\mu_G \in D(G)$, the Bethe internal energy $U^B[\mu_G]$ and the Bethe entropy $S^B[\mu_G]$ are then defined by

$$\begin{aligned}
U^B[\mu_G] &= -\sum_{e\in E} \mu_{[G,e]}(1) \\
S^B[\mu_G] &= -\sum_{v\in V} \sum_{\mathbf{b}_{\partial v}\in\{0,1\}^{|\partial v|}} \mu_{[G,\partial v]}(\mathbf{b}_{\partial v}) \ln\left(\mu_{[G,\partial v]}(\mathbf{b}_{\partial v})\right) \\
&\quad + \sum_{e\in E} \sum_{b_e\in\{0,1\}} \mu_{[G,e]}(b_e) \ln\left(\mu_{[G,e]}(b_e)\right)
\end{aligned}$$

The Bethe free entropy $\Phi^B[\mu_G; z]$ is then defined by

$$\Phi^B[\mu_G; z] = -U^B[\mu_G] \ln z + S^B[\mu_G]$$

It is well known that if G is a tree, i.e. acyclic graph, then we have $\Phi^B[\mu^z_G; z] = \Phi_G(z)$ (see [29]).

We first reformulate the Bethe free entropy function.

Proposition 3. *Let $\mu_G \in D(G)$ be a distribution over matchings. Define $\mathbf{x} \in \mathbb{R}^E$ by $x_e = \mu_{[G,e]}(1)$. Then we have $\mathbf{x} \in FM(G)$ defined by (1) and*

$$\begin{aligned}
U^B[\mu_G] &= -\sum_{e\in E} x_e \\
S^B[\mu_G] &= \frac{1}{2}\sum_{v\in V}\left\{\sum_{e\in\partial v} -x_e \ln x_e + (1-x_e)\ln(1-x_e)\right. \\
&\quad \left. -2\left(1-\sum_{e\in\partial v} x_e\right)\ln\left(1-\sum_{e\in\partial v} x_e\right)\right\},
\end{aligned}$$

with the standard convention $0 \ln 0 = 0$.

We then have

Proposition 4. *The function $S^B(\mathbf{x})$ defined by*

$$\begin{aligned}
S^B(\mathbf{x}) &= \frac{1}{2}\sum_{v\in V}\left\{\sum_{e\in\partial v} -x_e \ln x_e + (1-x_e)\ln(1-x_e)\right. \\
&\quad \left. -2\left(1-\sum_{e\in\partial v} x_e\right)\ln\left(1-\sum_{e\in\partial v} x_e\right)\right\}
\end{aligned}$$

is non-negative and concave on $FM(G)$ defined by (1).

We also define $U^B(\mathbf{x}) = -\sum_{e\in E} x_e$ and $\Phi^B(\mathbf{x}; z) = -U^B(\mathbf{x}) \ln z + S^B(\mathbf{x})$. Note that for any $\mu_G \in D(G)$, we have,

$$\Phi^B(\mu_G; z) = \Phi^B(\mathbf{x}, z),$$

for $\mathbf{x}$ defined by $x_e = \mu_{[G,e]}(1)$.

Proposition 5. *Recall that $\mathbf{x}(z) \in \mathbb{R}^E$ is defined by (6). Then we have:*

$$\sup_{\mathbf{x} \in FM(G)} \Phi^B(\mathbf{x}; z) \quad = \quad \Phi^B(\mathbf{x}(z); z).$$

In words, our BP algorithm is shown to maximize the Bethe free entropy which in our case is concave. Note that in general, a BP fixed point corresponds to a staionnary point of the Bethe free entropy, see [32].

We now give a reparametrization of the Gibbs distribution. For any vector $\mathbf{B} \in \{0,1\}^{\overrightarrow{E}}$, we denote by $\mathbf{B}_{\partial v} \in \{0,1\}^{\partial v}$ its restriction to components in ∂v. We first define the marginal probabilities

$$\mu_{\partial v}(\mathbf{B}_{\partial v}) = \left(1 - \sum_{e \in \partial v} x_e(z)\right)^{1 - \sum_{e \in \partial v} B_e} \prod_{e \in \partial v} x_e(z)^{B_e},$$

and

$$\mu_e(B_e) = x_e(z)^{B_e}(1 - x_e(z))^{1-B_e},$$

where $x_e(z)$ is defined by (6). Given a graph $G = (V, E)$ and some set $F \subset E$, we define $d_F(v)$ as the degree of node v in the subgraph induced by F. A generalized loop is any subset F such that $d_F(v) \neq 1$ for all $v \in V$. We define $V(F)$ as the number of vertices covered by F, i.e. vertices with $d_F(v) \geq 1$.

Theorem 3. *For any graph G, we have for $z > 0$,*

$$\mu_G^z(\mathbf{B}) = \frac{1}{Z} \frac{\prod_{v \in V} \mu_{\partial v}(\mathbf{B}_{\partial v})}{\prod_{e \in E} \mu_e(B_e)}, \tag{14}$$

with

$$Z = 1 + \sum_{\emptyset \neq F \subset E} (-1)^{V(F)} \prod_{v \in V} (d_F(v) - 1) \prod_{e \in F} \frac{x_e(z)}{1 - x_e(z)}, \tag{15}$$

where only generalized loops F lead to non-zero terms in the sum of (15). Moreover, we have

$$\ln Z = \Phi_G(z) - \Phi^B(\mathbf{x}(z); z).$$

Note that if G is a tree, we recover that $Z = 1$ and that our BP algorithm computes exactly the marginals of the Gibbs distribution defined by (12). However for general graphs, BP algorithm is not exact and equation (15) gives the exact correction term as a loop serie expansion [5]. Explicit computation of these loops is in general intractable. Indeed counting the total number of matchings $\exp(\Phi_G(1))$ falls into the class of $\#P$-complete problem. However equation (15) can be used to approximate such quantities by accounting for a small set of significant loop corrections.

3 Conclusion

We introduced an annealing BP algorithm for the vertex cover and matching problems and showed its convergence, its relation to LP relaxation and conditions for correctness. In contrast to previous results of this kind, we do not rely on the a priori uniqueness of the solution to the optimization problem. In view of the recent results [21] and [13], our approach should extend to more complex settings: b-matching, capacited matching. Another direction worth investigating is the question of the convergence time of our algorithm that we left open (techniques used in [25] seems relevant).

References

[1] M. Bayati, C. Borgs, J. Chayes, and R. Zecchina. Belief propagation for weighted b-matchings on arbitrary graphs and its relation to linear programs with integer solutions. *SIAM J. Discrete Math.*, 25(2):989–1011, 2011.

[2] M. Bayati, D. Shah, and M. Sharma. Max-product for maximum weight matching: convergence, correctness, and LP duality. *IEEE Trans. Inform. Theory*, 54(3):1241–1251, 2008.

[3] C. Bordenave, M. Lelarge, and J. Salez. Matchings on infinite graphs. *Probability Theory and Related Fields*, pages 1–26, 2012.

[4] M. Chertkov. Exactness of belief propagation for some graphical models with loops. *CoRR*, abs/0801.0341, 2008.

[5] M. Chertkov and V. Y. Chernyak. Loop calculus in statistical physics and information science. *Phys. Rev. E (3)*, 73(6):065102, 4, 2006.

[6] D. Gamarnik, D. Shah, and Y. Wei. Belief propagation for min-cost network flow: convergence and correctness. *Oper. Res.*, 60(2):410–428, 2012.

[7] A. E. Gelfand, J. Shin, and M. Chertkov. Belief propagation for linear programming. In *ISIT*, pages 2249–2253. IEEE, 2013.

[8] A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing algorithms for map lp-relaxations. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *NIPS*. Curran Associates, Inc., 2007.

[9] J. Gonzalez, Y. Low, and C. Guestrin. Residual splash for optimally parallelizing belief propagation. In *International Conference on Artificial Intelligence and Statistics*, page 177184, 2009.

[10] J. E. Gonzalez, Y. Low, H. Gu, D. Bickson, and C. Guestrin. Powergraph: Distributed graph-parallel computation on natural graphs. In *Proceedings of the 10th USENIX Conference on Operating Systems Design and Implementation*, OSDI'12, pages 17–30, Berkeley, CA, USA, 2012. USENIX Association.

[11] O. J. Heilmann and E. H. Lieb. Theory of monomer-dimer systems. *Comm. Math. Phys.*, 25:190–232, 1972.

[12] S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2- ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.

[13] M. Leconte, M. Lelarge, and L. Massoulié. Convergence of multivariate belief propagation, with applications to cuckoo hashing and load balancing. In S. Khanna, editor, *SODA*, pages 35–46. SIAM, 2013.

[14] M. Lelarge. A new approach to the orientation of random hypergraphs. In Y. Rabani, editor, *SODA*, pages 251–264. SIAM, 2012.

[15] M. Lelarge. Bypassing correlation decay for matchings with an application to xorsat. In *Information Theory Workshop (ITW), 2013 IEEE*, pages 1–5. IEEE, 2013.

[16] Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, and J. M. Hellerstein. Distributed graphlab: A framework for machine learning and data mining in the cloud. *Proc. VLDB Endow.*, 5(8):716–727, Apr. 2012.

[17] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. Pregel: A system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, SIGMOD '10, pages 135–146, New York, NY, USA, 2010. ACM.

[18] T. Meltzer, A. Globerson, and Y. Weiss. Convergent message passing algorithms: a unifying view. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 393–401. AUAI Press, 2009.

[19] J. Pearl. Reverend bayes on inference engines: A distributed hierarchical approach. In *AAAI*, pages 133–136, 1982.

[20] T. J. Richardson and R. L. Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47(2):599–618, 2001.

[21] J. Salez. Weighted enumeration of spanning subgraphs in locally tree-like graphs. *Random Structures & Algorithms*, 2012.

[22] S. Sanghavi, D. Malioutov, and A. Willsky. Belief propagation and LP relaxation for weighted matching in general graphs. *IEEE Trans. Inform. Theory*, 57(4):2203–2212, 2011.

[23] S. Sanghavi, D. Shah, and A. S. Willsky. Message passing for maximum weight independent set. *IEEE Trans. Inform. Theory*, 55(11):4822–4834, 2009.

[24] A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency.*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003.

[25] M. Sharify, S. Gaubert, and L. Grigori. Solution of the optimal assignment problem by diagonal scaling algorithms. *ArXiv e-prints*, Apr. 2011.

[26] A. Sinclair and P. Srivastava. Lee-yang theorems and the complexity of computing averages. In *STOC*, pages 625–634, 2013.

[27] P. O. Vontobel. The Bethe permanent of a nonnegative matrix. *IEEE Trans. Inform. Theory*, 59(3):1866–1901, 2013.

[28] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *Information Theory, IEEE Transactions on*, 51(11):3697–3717, 2005.

[29] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[30] Y. Weiss, C. Yanover, and T. Meltzer. Map estimation, linear programming and belief propagation with convex free energies. *arXiv preprint arXiv:1206.5286*, 2012.

[31] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(7):1165–1179, July 2007.

[32] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Inform. Theory*, 51(7):2282–2312, 2005.

4 Appendix: Proofs

4.1 Convergence of BP

Given a set of messages $\mathbf{X}$, we define a new set of messages $\mathbf{Y}$ by:

$$Y_{u\to v} = \frac{1}{1+\sum_{w\in\partial u\setminus v} X_{w\to u}}, \tag{16}$$

with the convention that the sum over the empty set equals zero. We denote by $\mathcal{R}_G$ the mapping sending $\mathbf{X} \in [0,\infty)^{\overrightarrow{E}}$ to $\mathbf{Y} = \mathcal{R}_G(\mathbf{X})$. We also denote by $\mathcal{R}_{\overrightarrow{e}}$ the local update rule (16): $Y_{\overrightarrow{e}} = \mathcal{R}_{\overrightarrow{e}}(\mathbf{X})$. Note that the mapping $z\mathcal{R}_G$ defined in (5) is simply the mapping multiplying by z each component of the output of the mapping $\mathcal{R}_G$ (making the notation consistent).

Proposition 6. *(i) For any finite graph G and $z > 0$, the fixed point equation:*

$$\mathbf{X} = z\mathcal{R}_G(\mathbf{X}) \tag{17}$$

has a unique attractive solution denoted $\mathbf{Y}(z) \in (0,+\infty)^{\overrightarrow{E}}$.

(ii) The function $z \mapsto \mathbf{Y}(z)$ is non-decreasing and the function $z \mapsto \frac{\mathbf{Y}(z)}{z}$ is non-increasing for $z > 0$.

(iii) If in addition, G is a finite tree, then for all $e \in E$, the law of B_e under μ_G^z is a Bernoulli distribution with

$$\mu_G^z(B_e = 1) = \frac{Y_{\overrightarrow{e}}(z)\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y}(z))}{1+Y_{\overrightarrow{e}}(z)\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y}(z))}. \tag{18}$$

Comparisons between vectors are always componentwise. Note that the right-hand side of (18) does not depend on the choice of orientation of the edge e as $\mathbf{Y}(z)$ satisfies (17). Before proving this proposition, we define for all $v \in V$, the following function of the messages $(Y_{\vec{e}}, \vec{e} \in \partial v)$,

$$\mathcal{D}_v(\mathbf{Y}) = \sum_{\vec{e} \in \partial v} \frac{Y_{\vec{e}} \mathcal{R}_{-\vec{e}}(\mathbf{Y})}{1 + Y_{\vec{e}} \mathcal{R}_{-\vec{e}}(\mathbf{Y})} \tag{19}$$

$$= \frac{\sum_{\vec{e} \in \partial v} Y_{\vec{e}}}{1 + \sum_{\vec{e} \in \partial v} Y_{\vec{e}}}. \tag{20}$$

In view of point (iii) of Proposition 6, we see that if the graph G is a tree, $\mathcal{D}_v(\mathbf{Y}(z))$ is simply the probability for vertex v to be covered by a matching distributed according to μ_G^z. In particular, when G is a tree, we can rewrite (13) as

$$\nu(G) = \lim_{z \to \infty} \frac{1}{2} \sum_{v \in V} \mathcal{D}_v(\mathbf{Y}(z)). \tag{21}$$

Proof. For the first point, we follow the proof of Theorem 3 in [21]. Let $z > 0$ and define the sequence of messages: $\mathbf{X}^0(z) = 0$ and for $t \geq 0$,

$$X_{u \to v}^{t+1}(z) = \frac{z}{1 + \sum_{w \in \partial u \setminus v} X_{w \to u}^t(z)}. \tag{22}$$

The sequence $\mathbf{X}^{2t}(z)$ (resp. $\mathbf{X}^{2t+1}(z)$) is non-decreasing (resp. non-increasing). We define $\lim_{t \to \infty} \uparrow \mathbf{X}^{2t}(z) = \mathbf{X}^-(z)$ and $\lim_{t \to \infty} \downarrow \mathbf{X}^{2t+1}(z) = \mathbf{X}^+(z)$. For any $\mathbf{Y}(z)$ fixed point of (17), a simple induction shows that

$$0 \leq \mathbf{X}^{2t}(z) \leq \mathbf{X}^-(z) \leq \mathbf{Y}(z) \leq \mathbf{X}^+(z) \leq \mathbf{X}^{2t+1}(z) \leq z.$$

We now prove that $\mathbf{X}^-(z) = \mathbf{X}^+(z)$ finishing the proof of the first point. Note that we have $\mathbf{X}^+(z) = z\mathcal{R}_G(\mathbf{X}^-(z))$ and $\mathbf{X}^-(z) = z\mathcal{R}_G(\mathbf{X}^+(z))$. In particular for any $z > 0$, we have $X_{\vec{e}}^+(z)\mathcal{R}_{-\vec{e}}(\mathbf{X}^+(z)) = X_{-\vec{e}}^-(z)\mathcal{R}_{\vec{e}}(\mathbf{X}^-(z))$ so that in view of (19), we have

$$\sum_{v \in V} \mathcal{D}_v(\mathbf{X}^+(z)) = \sum_{v \in V} \mathcal{D}_v(\mathbf{X}^-(z)). \tag{23}$$

We see from (20) that for each $v \in V$, $\mathcal{D}_v$ is an increasing function of the $(X_{\vec{e}}, \vec{e} \in \partial v)$, so that (23) together with $\mathbf{X}^-(z) \leq \mathbf{X}^+(z)$ imply the desired result.

We now prove that $z \mapsto \frac{\mathbf{X}^t(z)}{z}$ and $z \mapsto \mathbf{X}^t(z)$ are respectively non-increasing and non-decreasing, this implies point (ii). We prove it by induction on t: consider $z \leq z'$ if $\mathbf{X}^t(z) \leq \mathbf{X}^t(z')$ then by (22) we have $\frac{\mathbf{X}^{t+1}(z)}{z} \geq \frac{\mathbf{X}^{t+1}(z')}{z'}$ and if $\frac{\mathbf{X}^t(z)}{z} \geq \frac{\mathbf{X}^t(z')}{z'}$ then again by (22), we have $\mathbf{X}^{t+1}(z) \leq \mathbf{X}^{t+1}(z')$.

We consider now the case where G is a tree. For any directed edge $u \to v$, we define $T_{u \to v}$ as the subtree containing u and v and obtained from G by removing all incident edges to v except the edge uv. A simple computation shows that

$$\frac{\mu_{T_{u \to v}}^z(B_{uv=1})}{\mu_{T_{u \to v}}^z(B_{uv=0})} = \frac{z}{1 + \sum_{w \in \partial u \setminus v} \frac{\mu_{T_{w \to u}}^z(B_{wu=1})}{\mu_{T_{w \to u}}^z(B_{wu=0})}}.$$

This directly implies that for a finite tree, $Y_{u\to v}(z) = \frac{\mu_{T^z_{u\to v}}(B_{uv=1})}{\mu_{T^z_{u\to v}}(B_{uv=0})}$. Then a simple computation shows that

$$\begin{aligned}\frac{\mu^z_G(B_{uv=1})}{\mu^z_G(B_{uv=0})} &= \frac{Y_{u\to v}(z)Y_{v\to u}(z)}{z}\\ &= Y_{u\to v}(z)\mathcal{R}_{v\to u}(\mathbf{Y}(z)),\end{aligned}$$

which directly implies (18). □

4.2 Zero temperature limit

In order to compute the matching number, we must let z tend to infinity in $\mathbf{Y}(z) = z\mathcal{R}_G(\mathbf{Y}(z))$. Iterating once this recursion, we get $\mathbf{Y}(z) = z\mathcal{R}_G(z\mathcal{R}_G(\mathbf{Y}(z)))$. Note that we have for any $z > 0$,

$$z\mathcal{R}_{u\to v}(zX) = \frac{1}{z^{-1} + \sum_{w\in\partial u\backslash v} X_{w\to u}}$$

Hence we can define for any $\mathbf{X} \in (0,1]^{\overrightarrow{E}}$, $\mathcal{Q}_G(\mathbf{X}) = \lim_{z\to\infty} \uparrow z\mathcal{R}_G(z\mathbf{X}) \in (0,\infty]^{\overrightarrow{E}}$ by its local update rule:

$$\mathcal{Q}_{u\to v}(\mathbf{X}) = \frac{1}{\sum_{w\in\partial u\backslash v} X_{w\to u}}, \tag{24}$$

with the conventions $1/0 = \infty$ and the sum over the empty set equals zero (in particular, if u is a leaf of the graph G, then $\mathcal{Q}_{u\to v}(\mathbf{X}) = \infty$).

By point (ii) of Proposition 6, we can define $\lim_{z\to\infty} \uparrow \mathbf{Y}(z) = \mathbf{Y} \in [0,\infty]^{\overrightarrow{E}}$ and $\lim_{z\to\infty} \downarrow \frac{\mathbf{Y}(z)}{z} = \mathbf{X} \in [0,1]^{\overrightarrow{E}}$. Then, we have

$$\mathbf{X} = \mathcal{R}_G(\mathbf{Y}) \text{ and, } \mathbf{Y} = \mathcal{Q}_G(\mathbf{X}), \tag{25}$$

provided we can extend the maps $\mathcal{R}_G$ and $\mathcal{Q}_G$ continuously from their respective domains $[0,\infty)^{\overrightarrow{E}}$ and $(0,1]^{\overrightarrow{E}}$ to their compactifications $[0,\infty]^{\overrightarrow{E}}$ and $[0,1]^{\overrightarrow{E}}$ respectively. This can be done easily as follows: if there exists $w \in \partial u\backslash v$ with $Y_{w\to u} = \infty$, then we set $\mathcal{R}_{u\to v}(\mathbf{Y}) = 0$; and if $X_{w\to u} = 0$ for all $w \in \partial u\backslash v$, then we set $\mathcal{Q}_{u\to v}(\mathbf{X}) = \infty$.

Lemma 1. *Let* $\lim_{z\to\infty} \uparrow \mathbf{Y}(z) = \mathbf{Y} \in [0,\infty]^{\overrightarrow{E}}$. *Then* $\mathbf{Y}$ *is the smallest solution to the fixed point equation* $\mathbf{Y} = \mathcal{Q}_G \circ \mathcal{R}_G(\mathbf{Y})$.

Proof. Let $\mathbf{Z} = \mathcal{Q}_G \circ \mathcal{R}_G(\mathbf{Z})$. For any $z > 0$, we have for any $\mathbf{X} \in [0,1]^{\overrightarrow{E}}$, $z\mathcal{R}_G(z\mathbf{X}) \le \mathcal{Q}_G(\mathbf{X})$ so that an easy induction implies that $\mathbf{X}^{2t}(z) \le \mathbf{Z}$ where $\mathbf{X}^{2t}(z)$ is the sequence defined in the proof of Proposition 6. Letting first t and then z tend to infinity, allows us to conclude. □

Note that thanks to (20), we can extend the functions $\mathcal{D}_v(\mathbf{Y})$ continuously on $[0,\infty]^{\overrightarrow{E}}$ by setting $\mathcal{D}_v(\mathbf{Y}) = 1$ as soon as there exists $Y_{\overrightarrow{e}} = \infty$ for $\overrightarrow{e} \in \partial v$. To summarize, we have for each $v \in V$,

$$\lim_{z\to\infty} \mathcal{D}_v(\mathbf{Y}(z)) = \mathcal{D}_v(\mathbf{Y}) \le 1, \tag{26}$$

where $\mathbf{Y}$ is the smallest solution to the fixed point equation $\mathbf{Y} = \mathcal{Q}_G \circ \mathcal{R}_G(\mathbf{Y})$ that can be written as:

$$Y_{u\to v} = \frac{1}{\sum_{w\in\partial u\backslash v} \frac{1}{1+\sum_{w'\in\partial w\backslash u} Y_{w'\to w}}}, \tag{27}$$

with the conventions $1/0 = \infty$ and $1/\infty = 0$ and the sum over the empty set equals zero.

Lemma 2. *We have for any $\mathbf{Y} \in [0,\infty]^{\overrightarrow{E}}$ and $v \in V$,*

$$\begin{aligned}\mathcal{D}_v(\mathbf{Y}) \quad &= \quad \sum_{\overrightarrow{e}\in\partial v} \frac{Y_{\overrightarrow{e}}\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y})}{1+Y_{\overrightarrow{e}}\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y})} \mathbf{1}\left(Y_{\overrightarrow{e}} < \infty\right) \\ &\quad +\mathbf{1}\left(\exists \overrightarrow{e} \in \partial v, Y_{\overrightarrow{e}} = \infty\right),\end{aligned} \tag{28}$$

where the first sum on the right-hand side should be understood as a sum over $\overrightarrow{e} \in \partial v$ with $Y_{\overrightarrow{e}} < \infty$.

Note that since $\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y}) \in [0,1]$, the product $Y_{\overrightarrow{e}}\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y})$ is always well-defined in the expression above.

Proof. We only need to consider the case where there exists $\overrightarrow{e} \in \partial v$ such that $Y_{\overrightarrow{e}} = \infty$. By the discussion before the lemma, we have in this case $\mathcal{D}_v(\mathbf{Y}) = 1$. Hence we need to prove that the first term on the right-hand side of (28) vanishes. This follows form the following fact: let $\overrightarrow{e}' \in \partial v \backslash \overrightarrow{e}$, then $Y_{\overrightarrow{e}} = \infty$ implies that $\mathcal{R}_{-\overrightarrow{e}'}(\mathbf{Y}) = 0$. $\square$

For the messages $\mathbf{Y} \in [0,\infty]^{\overrightarrow{E}}$ (resp. $\mathbf{X} \in [0,1]^{\overrightarrow{E}}$) defined in (25), we define the $\{0,1\}$-valued messages $\mathbf{I}^Y$ (resp. $\mathbf{I}^X$) by $I^Y_{u\to v} = \mathbf{1}(Y_{u\to v} = \infty)$ (resp. $I^X_{u\to v} = \mathbf{1}(X_{u\to v} > 0)$. It follows directly from (25) and the definition of $\mathcal{P}_G$ (7) that

$$\mathbf{I}^Y = \mathcal{P}_G(\mathbf{I}^X), \text{ and, } \mathbf{I}^X = \mathcal{P}_G(\mathbf{I}^Y). \tag{29}$$

We now show that for any finite graph G, the right-hand term in (21) is a function of $\mathbf{I}^X$ and $\mathbf{I}^Y$ only.

For any $Y \in [0,\infty]$, we define $I(Y) = \mathbf{1}(Y = \infty)$ and still denote by I the function acting similarly on vectors componentwise, i.e. if $\mathbf{I} = I(\mathbf{Y})$ then $I_{\overrightarrow{e}} = I(Y_{\overrightarrow{e}})$.

Lemma 3. *For $\mathbf{Y} \in [0,\infty]^{\overrightarrow{E}}$, we define $\mathbf{Y}' = \mathcal{Q}_G \circ \mathcal{R}_G(\mathbf{Y})$. If $\mathbf{Y} \geq$ (resp. $\leq$) $\mathbf{Y}'$, then*

$$\sum_{v\in V} \mathcal{D}_v(\mathbf{Y}) \geq \text{ (resp. } \leq\text{) } \sum_{v\in V} F_v(I(\mathbf{Y}')),$$

where F_v was defined in (8).

Proof. Suppose $\mathbf{Y}' \leq \mathbf{Y}$, then using Lemma 2, we get

$$\begin{aligned}\sum_v \mathcal{D}_v(\mathbf{Y}) \quad &\geq \quad \underbrace{\sum_{\overrightarrow{e}\in\overrightarrow{E}} \frac{Y'_{\overrightarrow{e}}\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y})}{1+Y'_{\overrightarrow{e}}\mathcal{R}_{-\overrightarrow{e}}(\mathbf{Y})} \mathbf{1}\left(Y'_{\overrightarrow{e}} < \infty\right)}_{A} \\ &\quad +\sum_{v\in V} \mathbf{1}\left(\exists \overrightarrow{e} \in \partial v, Y'_{\overrightarrow{e}} = \infty\right).\end{aligned}$$

For the first term A, denote $\mathbf{X} = \mathcal{R}_G(\mathbf{Y})$ so that $\mathbf{Y}' = \mathcal{Q}_G(\mathbf{X})$. Then we have

$$\begin{aligned}
A &= \sum_{\overrightarrow{e} \in \overrightarrow{E}} \frac{\mathcal{Q}_{\overrightarrow{e}}(\mathbf{X}) X_{-\overrightarrow{e}}}{1 + \mathcal{Q}_{\overrightarrow{e}}(\mathbf{X}) X_{-\overrightarrow{e}}} \mathbf{1}\left(\mathcal{Q}_{\overrightarrow{e}}(\mathbf{X}) < \infty\right) \\
&= \sum_{\overrightarrow{e} \in \overrightarrow{E}} \frac{\mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X}) X_{\overrightarrow{e}}}{1 + \mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X}) X_{\overrightarrow{e}}} \mathbf{1}\left(\mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X}) < \infty\right) \\
&= \sum_{v \in V} \underbrace{\sum_{\overrightarrow{e} \in \partial v} \frac{X_{\overrightarrow{e}} \mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X})}{1 + X_{\overrightarrow{e}} \mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X})} \mathbf{1}\left(\mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X}) < \infty\right).}_{B_v}
\end{aligned}$$

We now prove that

$$B_v = \left(1 - \sum_{\overrightarrow{e} \in \partial v} I_{-\overrightarrow{e}}(\mathbf{Y}')\right)^+. \tag{30}$$

First note that if $J(\mathbf{X})$ is defined by $J_{\overrightarrow{e}}(\mathbf{X}) = \mathbf{1}(X_{\overrightarrow{e}} > 0)$, then we have $\mathcal{P}_G(J(\mathbf{X})) = I(\mathbf{Y}')$. Hence if $\sum_{\overrightarrow{e} \in \partial v} I_{-\overrightarrow{e}}(\mathbf{Y}') = 0$, then $\exists w \neq w'$ both in ∂v with $X_{w \to v} X_{w' \to v} > 0$. This in turn implies that $0 < \mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X}) < \infty$ for all $\overrightarrow{e} \in \partial v$, so that in this case we have

$$B_v = \sum_{\overrightarrow{e} \in \partial v} \frac{X_{\overrightarrow{e}}}{\mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X})^{-1} + X_{\overrightarrow{e}}} = 1.$$

Note now that if $B_v > 0$, there must exists $\overrightarrow{e} \in \partial v$ such that $X_{\overrightarrow{e}} > 0$ and $\mathcal{Q}_{-\overrightarrow{e}}(\mathbf{X}) < \infty$ and this last constraint implies that there exists $\overrightarrow{e}' \neq \overrightarrow{e}$ with $\overrightarrow{e}' \in \partial v$ with $X_{\overrightarrow{e}'} > 0$. In particular, we have $B_v = 1$ and $\sum_{\overrightarrow{e} \in \partial v} I_{-\overrightarrow{e}}(\mathbf{Y}') = 0$. This finished the proof of (30). The lemma then follows. $\square$

We are now ready to state our first main result for finite graphs:

Proposition 7. *For any finite graph G, we have*

$$\sum_{v \in V} \mathcal{D}_v(\mathbf{Y}) = \lim_{z \to \infty} \sum_{v \in V} \mathcal{D}_v(\mathbf{Y}(z)) = \inf_{\mathbf{I}} \sum_{v \in V} F_v(\mathbf{I}),$$

where the infimum is over the solutions of $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$.

Proof. Let $\mathbf{Y} = \lim_{z \to \infty} \uparrow \mathbf{Y}(z)$ and recall that we denoted $\mathbf{I}^Y = I(\mathbf{Y})$ so that $\mathbf{I}^Y = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I}^Y)$ by (29). By Lemma 3 and (26), we have

$$\lim_{z \to \infty} \sum_{v \in V} \mathcal{D}_v(\mathbf{Y}(z)) = \sum_{v \in V} \mathcal{D}_v(\mathbf{Y}) = \sum_{v \in V} F_v(\mathbf{I}^Y). \tag{31}$$

We need to prove that if $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$ then we have $\sum_v F_v(\mathbf{I}) \geq \sum_{v \in V} \mathcal{D}_v(\mathbf{Y})$. For any such $\mathbf{I}$, we define $\mathbf{W}^0$ as follows:

$$W^0_{\overrightarrow{e}} = \begin{cases} \infty & \text{if } I_{\overrightarrow{e}} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then let $\mathbf{W}^{k+1} = \mathcal{Q}_G \circ \mathcal{R}_G(\mathbf{W}^k)$ for $k \geq 0$. A simple induction shows that $I(\mathbf{W}^{k+1}) = \mathcal{P}_G \circ \mathcal{P}_G(I(\mathbf{W}^k)) = \mathbf{I}$ for all $k \geq 0$. In particular, $\mathbf{W}^0 \leq \mathbf{W}^1$ and again by induction, we see that the sequence $\{\mathbf{W}^k\}_k$ is non-decreasing and we denote by $\mathbf{W}^{\mathbf{I}}$ its limit. Applying Lemma 3 to $\mathbf{W}^k$, we get

$$\sum_{v \in V} \mathcal{D}_v(\mathbf{W}^k) \leq \sum_{v \in V} F_v(\mathbf{I}).$$

Taking the limit $k \to \infty$, we obtain

$$\sum_{v \in V} \mathcal{D}_v(\mathbf{W}^{\mathbf{I}}) \leq \sum_{v \in V} F_v(\mathbf{I})$$

Moreover $\mathbf{Y}$ being the smallest solution to the fixed point equation $\mathbf{Y} = \mathcal{R}_G \circ \mathcal{R}_G(\mathbf{Y})$, we have $\mathbf{Y} \leq \mathbf{W}^{\mathbf{I}}$ and using the fact that $\mathcal{D}_v$ is increasing, we get

$$\sum_{v \in V} \mathcal{D}_v(\mathbf{Y}) \leq \sum_{v \in V} \mathcal{D}_v(\mathbf{W}^{\mathbf{I}}) \leq \sum_{v \in V} F_v(\mathbf{I}),$$

which concludes the proof. □

We now prove

Lemma 4. *If $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$, then $(F_v(\mathbf{I})/2,\ v \in V)$ is a half-integral vertex cover.*

Proof. We need to prove that for all $(uv) \in E$, $F_u(\mathbf{I}) + F_v(\mathbf{I}) \geq 2$. This follows easily from the fact that if $\sum_{w \in \partial u} I_{w \to u} = 0$ then we have $I_{v \to w} = 0$ for all $w \in \partial v$ and hence $\left(1 - \sum_{w \in \partial v} I_{v \to w}\right)^+ = 1$. Hence we have

$$\begin{aligned} F_u(\mathbf{I}) + F_v(\mathbf{I}) &\geq \sum_{w \in \partial u} I_{w \to u} + \sum_{w \in \partial v} I_{w \to v} \\ &+ \mathbf{1}\left(\sum_{w \in \partial u} I_{w \to u} = 0\right) + \mathbf{1}\left(\sum_{w \in \partial v} I_{w \to v} = 0\right) \\ &\geq 2. \end{aligned}$$

□

In the rest of this subsection, the graph $G = (U \cup W, E)$ is assumed to be bipartite. The following lemma shows that $V(\mathbf{I}^Y)$ is a vertex cover.

Lemma 5. *For any $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$, the set $V(\mathbf{I})$ defined by (10) and (11) is a vertex cover.*

Proof. Consider $u \in U$, $w \in W$ and $(uw) \in E$. We denote $\mathbf{J} = \mathcal{P}_G(\mathbf{I})$ so that $\mathbf{I} = \mathcal{P}_G(\mathbf{J})$. The fact that $u \notin V(\mathbf{I})$ implies that $J_{u \to w} = 1$ and since $I_{w \to u} = 0$, there exists $v \in \partial w \backslash u$ such that $J_{v \to w} = 1$ so that $w \in V(\mathbf{I})$. Similarly if $w \notin V(\mathbf{I})$ then $\sum_{v \in \partial w} J_{v \to w} \leq 1$. Hence if $J_{u \to w} = 1$, then $I_{w \to u} = 1$ and if $J_{u \to w} = 0$ then there exists $v \in \partial u \backslash w$ with $I_{v \to u} = 1$. So in both cases, $u \in V(\mathbf{I})$. □

Lemma 6. *Let $\mathbf{I}$ achieving $\inf_{\mathbf{I}} \sum_{v \in V} F_v(\mathbf{I})$ where the infimum is over the solutions of $\mathbf{I} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{I})$. Then, the size of the set $V(\mathbf{I})$ is $\frac{1}{2} \sum_v F_v(\mathbf{I})$.*

Proof. Again, we denote $\mathbf{J} = \mathcal{P}_G(\mathbf{I})$ so that $\mathbf{I} = \mathcal{P}_G(\mathbf{J})$. First note that $\left(1 - \sum_{\vec{e}\in\partial v} I_{-\vec{e}}\right)^+ = \mathbf{1}\left(\sum_{\vec{e}\in\partial v} J_{\vec{e}} \geq 2\right)$. Hence we have $\sum_v F_v(\mathbf{I}) = A + B$, with

$$\begin{aligned} A &= \sum_{u\in U}\left(1 \wedge \sum_{w\in\partial u} I_{w\to u}\right) + \sum_{w\in W}\mathbf{1}\left(\sum_{u\in\partial w} J_{u\to w} \geq 2\right), \\ B &= \sum_{w\in W}\left(1 \wedge \sum_{u\in\partial w} I_{u\to w}\right) + \sum_{u\in U}\mathbf{1}\left(\sum_{w\in\partial u} J_{w\to u} \geq 2\right). \end{aligned}$$

Clearly A is the size of the set $V(\mathbf{I})$, so we need only to show that $A = B$. Note that A depends only on messages form $\mathbf{I}$ from nodes in W to nodes in U and B depends only on the remaining messages in $\mathbf{I}$. Assume that $A < B$ and consider

$$B' = \sum_{w\in W}\left(1 \wedge \sum_{u\in\partial w} J_{u\to w}\right) + \sum_{u\in U}\mathbf{1}\left(\sum_{w\in\partial u I_{w\to u}}\right).$$

Note that we have

$$\begin{aligned} 1 \wedge \sum_{u\in\partial w} J_{u\to w} &= \left(1 - \sum_{u\in\partial w} I_{w\to u}\right)^+ \\ &= \mathbf{1}\left(\sum_{u\in\partial w} J_{u\to w} \geq 2\right). \end{aligned}$$

In particular we have $B' \leq A < B$. Moreover if $\mathbf{K}$ is such that for any $w \in W$ and $u \in U$, $K_{w\to u} = I_{w\to u}$ and $K_{u\to w} = J_{u\to w}$, then we have $\sum_v F_v(\mathbf{K}) = A + B' < \sum_v F_v(\mathbf{I})$ and $\mathbf{K} = \mathcal{P}_G \circ \mathcal{P}_G(\mathbf{K})$ contradicting the minimality of $\mathbf{I}$. $\square$

4.3 Positive temperature

We first prove Proposition 3. For $\mu_G \in D(G)$, we have $\mu_G(\sum_{e\in\partial v} B_e \leq 1) = 1$ so that by the linearity of expectation, $\sum_{e\in\partial v}\mu_{[G,e]}(1) \leq 1$ and the vector $\mathbf{x}$ with component $x_e = \mu_{[G,e]}(1)$ is in $FM(G)$. Now for each $v \in V$ and $e \in \partial v$, we must have

$$\begin{aligned} \mu_{[G,\partial v]}(b_e = 1) &= \sum_{b_f\in\{0,1\},\, f\in\partial v\backslash e} \mu_{[G,\partial v]}(\mathbf{b}_{\partial v}) \\ &= \mu_{[G,\partial v]}(b_e = 1,\ b_f = 0, f \in \partial v\backslash e) \\ &= x_e. \end{aligned}$$

It then follows that

$$\mu_{[G,\partial v]}(\mathbf{B}_{\partial v}) = \left(1 - \sum_{e\in\partial v} x_e\right)^{1-\sum_{e\in\partial v} B_e} \prod_{e\in\partial v} x_e^{B_e}, \tag{32}$$

and the formula for $S^B[\mu_G]$ follows.

We now give a lemma implying the concavity of the Bethe entropy, Proposition 4. For $k \in \mathbb{N}$, we define $\Delta^k = \{\mathbf{x} \in \mathbb{R}^k,\ x_i \geq 0, \sum_{i=1}^k x_i \leq 1\}$.

Lemma 7. *Let $g : \Delta^k \to \mathbb{R}$ be defined by*

$$
\begin{aligned}
g(\mathbf{x}) &= -\sum_i x_i \ln x_i + \sum_i (1-x_i)\ln(1-x_i) \\
&\quad -2\left(1-\sum_i x_i\right)\ln\left(1-\sum_i x_i\right).
\end{aligned}
$$

For $k \geq 1$, g is concave. Moreover, we have

$$
\frac{\partial g}{\partial x_i} = \ln\left(\frac{\left(1-\sum_j x_j\right)^2}{x_i(1-x_i)}\right).
$$

Proof. From Theorem 20 in [27], we know that the function

$$
\begin{aligned}
h(\mathbf{x}) &= -\sum_i x_i \ln x_i + \sum_i (1-x_i)\ln(1-x_i) \\
&\quad -\left(1-\sum_i x_i\right)\ln\left(1-\sum_i x_i\right) \\
&\quad +\left(\sum_i x_i\right)\ln\left(\sum_i x_i\right)
\end{aligned}
$$

is non-negative and concave on Δ^k. We have

$$
g(\mathbf{x}) = h(\mathbf{x}) + H\left(\sum_i x_i\right),
$$

where $H(p) = -p\ln p - (1-p)\ln(1-p)$ is the entropy of a Bernoulli random variable and is concave in p. $\square$

We now prove Proposition 5. For $e = (uv) \in E$ and $\mathbf{x} \in \mathring{\Delta}^k$ (the interior of Δ^k), we have

$$
\begin{aligned}
\frac{\partial \Phi^B(\mathbf{x}; z)}{\partial x_e} &= -\ln z \\
&\quad +\ln\left(\frac{\left(1-\sum_{f\in\partial v} x_f\right)\left(1-\sum_{f\in\partial u} x_f\right)}{x_e(1-x_e)}\right).
\end{aligned}
$$

Hence, we have $\frac{\partial \Phi^B(\mathbf{x};z)}{\partial x_e} = 0$ if and only if

$$
x_e(1-x_e) = z\left(1-\sum_{f\in\partial v} x_f\right)\left(1-\sum_{f\in\partial u} x_f\right). \tag{33}
$$

Note that $\sum_{f\in\partial v} x_f(z) = \mathcal{D}_v(\mathbf{Y}(z))$, so that we have by (20)

$$\begin{aligned}\left(1-\sum_{f\in\partial v} x_f(z)\right) &= \left(1-\frac{\sum_{\vec{e}\in\partial v} Y_{\vec{e}}(z)}{1+\sum_{\vec{e}\in\partial v} Y_{\vec{e}}(z)}\right)\\ &= \left(1+\sum_{\vec{e}\in\partial v} Y_{\vec{e}}(z)\right)^{-1}\end{aligned}$$

We have for $e=(uv)\in E$,

$$x_e(z) = \frac{Y_{u\to v}(z)}{\frac{z}{Y_{v\to u}(z)}+Y_{u\to v}(z)},$$

and using the fact that $\mathbf{Y}(z)z = z\mathcal{R}_G(\mathbf{Y}(z))$, we get

$$x_e(z) = \frac{Y_{u\to v}(z)}{1+\sum_{w\in\partial v} Y_{w\to v}(z)} = \frac{Y_{v\to u}(z)}{1+\sum_{w\in\partial u} Y_{w\to u}(z)}.$$

Hence, evaluating (33) at $x_e(z)$, we get

$$x_e(z)(1-x_e(z))\frac{Y_{u\to v}(z)Y_{v\to u}(z)}{z} = x_e(z)^2,$$

so that

$$\frac{x_e(z)}{1-x_e(z)} = \frac{Y_{u\to v}(z)Y_{v\to u}(z)}{z},$$

which follows from the definition of $x_e(z)$ in (6). Proposition 5 follows.

We also note that the following equality (which will be used later) is true for $\mathbf{x}(z)$ defined by (6):

$$\frac{x_e(z)(1-x_e(z))}{z} = \left(1-\sum_{e'\in\partial u} x_{e'}(z)\right)\left(1-\sum_{e'\in\partial v} x_{e'}(z)\right) \tag{34}$$

4.4 Proofs of the main results of Sections 2.1 and 2.2

We now prove Theorems 1 and 2. First by Proposition 6 (i), we have $\lim_{t\to\infty} m^t_{\vec{e}}(z) = Y_{\vec{e}}(z)$. To end the proof of Theorem 1, we need to show that $\mathbf{x}(z)$ converges as $z\to\infty$ to a maximum fractional matching. First note that $S^B(\mathbf{x}) \le |E|$ so that for z sufficiently large, Proposition 5 implies that $\mathbf{x}(z)$ is on the optimal face of $FM(G)$, i.e. $\sum_e x_e(z) = \nu^*(G)$ and maximizes the function $S^B(\mathbf{x})$.

The first statement of Theorem 2 is exactly Lemma 4 and the second statement is Proposition 7. Proposition 1 follows from (31) and Lemma 4. Proposition 2 follows from Lemmas 5, 6 and the fact that $\mathbf{I}^Y$ achieves the minimum in this last lemma (see Proposition 1).

4.5 Loop series expansion

In this section, we prove Theorem 3. The fact that μ_G^z can be written as (14) (called tree-based reparameterization in [29]) follows from a direct application of the definitions.

To simplify notation, we write in the proofs x_e instead of $x_e(z)$.

Lemma 8. *For any $v \in V$, $z > 0$, we have*

$$\frac{\mu_{\partial v}(\mathbf{B}_{\partial v})}{\prod_{e\in\partial v}\mu_e(B_e)} = 1 - \sum_{S\subset\partial v}(-1)^{|S|}\left(|S|-1\right)\prod_{e\in S}\frac{B_e - x_e(z)}{1-x_e(z)}.$$

Proof. Note that if $B_f = 1$, the left-hand side is equal to $\prod_{e\neq f}(1-x_e)^{-1}$, while if $\sum_{e\in\partial v} B_e = 0$, it is equal to $\frac{1-\sum_{e\in\partial v}x_e}{\prod_{e\in\partial v}(1-x_e)}$. We need to check that the right-hand side agrees in these two cases. Let consider the case $B_f = 1$, then the right-hand side (denoted R) equals:

$$\begin{aligned}
R &= 1 - \sum_{|S|\geq 1, f\notin S}(-1)^{|S|}\left(|S|-1\right)\prod_{e\in S}\frac{-x_e}{1-x_e} \\
&\quad - \sum_{|S|\geq 1, f\in S}(-1)^{|S|}\left(|S|-1\right)\prod_{e\in S, e\neq f}\frac{-x_e}{1-x_e} \\
&= 1 - \sum_{|S|\geq 1, f\notin S}(-1)^{|S|+1}\prod_{e\in S}\frac{-x_e}{1-x_e} \\
&= 1 + \sum_{|S|\geq 1, f\notin S}\frac{\prod_{e\in S}x_e\prod_{e'\notin S, e'\neq f}(1-x_{e'})}{\prod_{e\neq f}(1-x_e)} \\
&= \frac{1}{\prod_{e\neq f}(1-x_e)}.
\end{aligned}$$

A similar computation shows the second case. $\square$

The following lemma shows (15).

Lemma 9. *We have*

$$Z = 1 - \sum_{\emptyset\neq F\subset E}(-1)^{V(F)}\prod_{v\in V}\left(d_F(v)-1\right)\prod_{e\in F}\frac{x_e(z)}{1-x_e(z)}.$$

Proof. By definition, we have

$$Z = \sum_{\mathbf{B}}\prod_e \mu_e(B_e)\prod_v\frac{\mu_{\partial v}(\mathbf{B}_{\partial v})}{\prod_{e\in\partial v}\mu_e(B_e)}.$$

By Lemma 8, we have

$$\begin{aligned}
P &:= \prod_v\frac{\mu_{\partial v}(\mathbf{B}_{\partial v})}{\prod_{e\in\partial v}\mu_e(B_e)} \\
&= \prod_v\left(1 + \sum_{S\subset\partial v}(-1)^{|S|-1}\left(|S|-1\right)\prod_{e\in S}\frac{B_e - x_e}{1-x_e}\right)
\end{aligned}$$

Z can be seen as an expectation of P where the B_e are independent Bernoulli random variables with parameter x_e. In particular expanding P, we see that only the terms $(B_e - x_e)^2$ will contribute to its expectation so that we get

$$\begin{aligned}
Z &= 1 + \sum_{\emptyset \neq F \subset E} \prod_v \left(\frac{(-1)^{d_F(v)-1}(d_F(v)-1)}{\prod_{e\in\partial v\cap F}(1-x_e)} \right) \prod_{e\in F} x_e(1-x_e) \\
&= 1 + \sum_{\emptyset \neq F \subset E} (-1)^{V(F)} \prod_v (d_F(v)-1) \prod_{e\in F} \frac{x_e}{1-x_e},
\end{aligned}$$

where in the last claim, we used $\prod_v (-1)^{d_F(v)} = 1$. $\square$

The following lemma shows the last statement in Theorem 3.

Lemma 10. *We have*

$$\ln Z = \Phi_G(z) - \Phi^B(\mathbf{x}(z); z).$$

Proof. We first compute

$$\begin{aligned}
e^{\Phi^B(\mathbf{x};z)} &= z^{-U^B(\mathbf{x})} \prod_v \left(1 - \sum_{e\in\partial v} x_e\right)^{-(1-\sum_{e\in\partial v} x_e)} \\
&\quad \prod_v \left(\prod_{e\in\partial v} x_e^{-x_e/2}(1-x_e)^{(1-x_e)/2} \right) \\
&= z^{\sum_e x_e} \prod_e x_e(1-x_e)^{1-x_e} \\
&\quad \prod_v \left(1 - \sum_{e\in\partial v} x_e\right)^{-(1-\sum_{e\in\partial v} x_e)}.
\end{aligned}$$

We now use the relation (34), to get:

$$e^{\Phi^B(\mathbf{x};z)} = \prod_e \frac{z}{x_e} \prod_v \left(1 - \sum_{e\in\partial v} x_e\right)^{|\partial v|-1}. \tag{35}$$

We now compute

$$\begin{aligned}
\frac{\prod_{v\in V} \mu_{\partial v}(\mathbf{B}_{\partial v})}{\prod_{e\in E} \mu_e(B_e)} &= \prod_v \left(1 - \sum_{e\in\partial v} x_e\right)^{1-\sum_{e\in\partial v} B_e} \\
&\quad \prod_e (1-x_e)^{b_e-1},
\end{aligned}$$

again using (34), we obtain

$$\begin{aligned}
\frac{\prod_{v\in V} \mu_{\partial v}(\mathbf{B}_{\partial v})}{\prod_{e\in E} \mu_e(B_e)} &= z^{\sum_e B_e} \prod_e \frac{x_e}{z} \\
&\quad \prod_v \left(1 - \sum_{e\in\partial v} x_e\right)^{-|\partial v|+1}.
\end{aligned}$$

Thanks to (35), we get

$$\frac{\prod_{v \in V} \mu_{\partial v}(\mathbf{B}_{\partial v})}{\prod_{e \in E} \mu_e(B_e)} \quad = \quad \frac{z^{\sum_e B_e}}{e^{\Phi^B(\mathbf{x};z)}},$$

hence summing over all matchings $\mathbf{B}$, we obtain

$$Z = \frac{P_G(z)}{e^{\Phi^B(\mathbf{x};z)}}.$$

$\square$