SIAM J. Discrete Math.
Vol. 25, No. 2, pp. 1012–1034

COUNTING INDEPENDENT SETS USING THE BETHE APPROXIMATION*

VENKAT CHANDRASEKARAN†, MISHA CHERTKOV‡, DAVID GAMARNIK§, DEVAVRAT SHAH†, and JINWOO SHIN¶

Abstract. We consider the #P-complete problem of counting the number of independent sets in a given graph. Our interest is in understanding the effectiveness of the popular belief propagation (BP) heuristic. BP is a simple iterative algorithm that is known to have at least one fixed point, where each fixed point corresponds to a stationary point of the Bethe free energy (introduced by Yedidia, Freeman, and Weiss [*IEEE Trans. Inform. Theory,* 51 (2004), pp. 2282–2312] in recognition of Bethe's earlier work in 1935). The evaluation of the Bethe free energy at such a stationary point (or BP fixed point) leads to the Bethe approximation for the number of independent sets of the given graph. BP is not known to converge in general, nor is an efficient, convergent procedure for finding stationary points of the Bethe free energy known. Furthermore, the effectiveness of the Bethe approximation is not well understood. As the first result of this paper we propose a BP-like algorithm that always converges to a stationary point of the Bethe free energy for *any* graph for the independent set problem. This procedure finds an ε-approximate stationary point in $O(n^2 d^4 2^d \varepsilon^{-4} \log^3(n\varepsilon^{-1}))$ iterations for a graph of n nodes with max-degree d. We study the quality of the resulting Bethe approximation using the recently developed "loop series" framework of Chertkov and Chernyak [*J. Stat. Mech. Theory Exp.*, 6 (2006), P06009]. As this characterization is applicable only for exact stationary points of the Bethe free energy, we provide a slightly modified characterization that holds for ε-approximate stationary points. We establish that for any graph on n nodes with max-degree d and girth larger than $8d \log_2 n$, the multiplicative error between the number of independent sets and the Bethe approximation decays as $1 + O(n^{-\gamma})$ for some $\gamma > 0$. This provides a deterministic counting algorithm that leads to strictly different results compared to a recent result of Weitz [in *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, ACM Press, New York, 2006, pp. 140–149]. Finally, as a consequence of our analysis we prove that the Bethe approximation is exceedingly good for a random 3-regular graph conditioned on the shortest cycle cover conjecture of Alon and Tarsi [*SIAM J. Algebr. Discrete Methods*, 6 (1985), pp. 345–350] being true.

Key words. Bethe free energy, independent set, belief propagation, loop series

AMS subject classifications. 68Q87, 68W25, 68R10

DOI. 10.1137/090767145

1. Introduction.

We consider the problem of counting the number of independent sets in a given graph. This problem has been of great interest as it is a prototypical #P-complete problem. It is worth noting that such counting questions do arise in practice as well, e.g., for performance evaluation of a finite buffered radio network (see Kelly [10]). Recently, the belief propagation (BP) algorithm has become the heuristic of choice in many similar applications where the interest is in computing what physicists refer to as the partition function of a given statistical model or, equivalently, when restricted to our setup, the number of independent sets for a given graph. In this paper we wish

*Received by the editors August 10, 2009; accepted for publication (in revised form) July 16, 2010; published electronically July 1, 2011. This work was supported in part by NSF EMT/MISC collaborative project 0829893.

http://www.siam.org/journals/sidma/25-2/76714.html

†Laboratory for Information and Decision Systems, Department of EECS, MIT, Cambridge, MA 02139 (venkatc@mit.edu, devavrat@mit.edu).

‡Center for Nonlinear Studies and Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 (chertkov@lanl.gov), and New Mexico Consortium, Los Alamos, NM 87544.

§Operations Research Center and Sloan School of Management, MIT, Cambridge, MA 02139 (gamarnik@mit.edu).

¶Algorithms and Randomness Center, Georgia Institute of Technology, Atlanta, GA 30332 (jshin72@cc.gatech.edu).

to understand the effectiveness of such an approximation for counting independent sets in a given graph.

BP is a simple iterative message-passing algorithm. It is well known that this iterative procedure does have fixed points, which correspond to stationary points of the Bethe free energy—for more details on the Bethe approximation and its relation to BP fixed points, see Yedidia, Freeman, and Weiss [21]; also see the book by Georgii [9]. However, there are two key problems. First, BP is not known to converge for general graphs for counting the number of independent sets; indeed, there are known counterexamples for other problems (see [17]). Second, given the BP fixed points, i.e., the Bethe approximation, it is not clear what the quality of the approximation is. The main results in this paper address both of these challenges. Before explaining our results, we provide a brief description of relevant prior work.

1.1. Prior work. Previous work on counting the number of independent sets in a given graph falls into two broad categories. The first and major body of work is based on sampling via Markov chains. In this approach, initiated by the works of Dyer, Frieze, and Kannan [6] and Sinclair and Jerrum [14], one wishes to design a Markov chain that samples independent sets uniformly and has a fast mixing property. Some of the notable results for independent set problems are by [12], [7], [16], [5]. These results show the following: (a) for any graph with max-degree up to 4, there exists a fully polynomial randomized approximation scheme using a fast mixing Markov chain, (b) there is no fast mixing Markov chain (based on local updates) for all graphs with degree larger than or equal to 6, and (c) approximately counting independent sets for all graphs with degree larger than 25 is hard.

The second approach introduced by Weitz [18] provides a deterministic fully polynomial-time approximation scheme for any graph with max-degree up to 5. It is based on establishing a *correlation decay* property for any tree with max-degree up to 5 and an intriguing equivalence relation between an appropriate distribution on a graph and an appropriate distribution on its self-avoiding walk tree. We also note the work by Bandyopadhyay and Gamarnik [2]: it establishes that the Bethe approximation is asymptotically correct for graphs with large girth and degree up to 5 (e.g., random 4-regular graphs). As in [18] it also uses the correlation decay property. On the other hand, it provides an $o(n)$ bound for graphs of size n between the logarithms of the number of independent sets and the Bethe approximation.

In summary all of the above results use some form of a correlation decay property—either dynamic or spatial. Furthermore, the generic conditions based just on max-degree are unlikely to extend beyond what is already known.

1.2. Our results. In order to obtain good approximation results for graphs with larger (> 5) max-degree, but possibly with additional constraints such as large girth, we study the BP/Bethe approximation for counting the number of independent sets. As the main result, we provide a deterministic algorithm based on the Bethe free energy for approximately computing the number of independent sets in a graph of n nodes with max-degree d and girth larger than $8d \log_2 n$ for any d.

As the first step toward establishing this result, we propose a new simple message-passing algorithm that can be viewed as a minor modification of BP. We show that our algorithm always converges to a stationary point of Bethe free energy for any graph for the independent set problem. To obtain an ε-approximate stationary point of the Bethe free energy for a graph on n nodes with max-degree d, the algorithm takes $O(n^2 d^4 2^d \varepsilon^{-4} \log^3(n\varepsilon^{-1}))$ iterations (see Theorem 2).

We analyze the error in the resulting Bethe approximation using the recently developed framework of "loop series" by Chertkov and Chernyak [4], which characterizes this error as a summation of terms with each term associated with a "generalized loop" of the graph. As this characterization is applicable only for exact stationary points of the Bethe free energy, we provide a bound on the error in the loop series expansion for ε-approximate stationary points. Though this approach provides an "explicit" characterization of the error, it involves possibly exponentially many terms and hence is far from trivial to evaluate in general. To tackle this challenge and bound the error, we develop a new combinatorial method to evaluate this summation. We do so by bounding the summation through a product of terms that involves what we call *apples*—an apple is a simple cycle or a cycle plus a connected line. Along with the result of Bermond, Jackson, and Jaeger [3], this leads to the eventual result that the error in the Bethe approximation for the number of independent sets decays as $O(n^{-\gamma})$ for some $\gamma > 0$ for any graph on n nodes with max-degree d and girth larger than $8d \log_2 n$.

By replacing the result of Bermond, Jackson, and Jaeger by its stronger version, also known as the shortest cycle cover conjecture (SCCC) of Alon and Tarsi [1], we obtain a stronger statement for random 3-regular graphs: the difference between the logarithms of the number of independent sets and the Bethe approximation is $O(1)$ with high probability. This is in sharp contrast to the result of Bandyopadhyay and Gamarnik [2] that does not assume the SCCC and suggests that the error is $o(n)$ based on correlation decay arguments (and also as expected by physicists). Thus we have an intriguing situation—either the SCCC is false or the Bethe approximation is terrific for counting the number of independent sets![1] A byproduct of the technique used to establish the result for random 3-regular graphs is the following algorithmic implication: it suggests a systematic way to correct the error in the Bethe approximation, which could be of interest in its own right.

1.3. Organization. Section 2 introduces the Bethe approximation for the problem of computing the number of independent sets in a given graph and the error characterization based on loop series for this approximation. We also briefly discuss the BP algorithm and its relation to the stationary points of the Bethe free energy. In section 4 we describe a new message-passing algorithm for computing a stationary point of the Bethe free energy for the independent set problem. We obtain its rate of convergence in Theorem 2. In section 5 we analyze the error in the resulting Bethe approximation for graphs with large girth. Finally, in section 6 we obtain a sharp bound on the error of the Bethe approximation for random 3-regular graphs assuming the SCCC.

2. Background. Let $G = (V, E)$ be a graph with vertices $V = \{1, \ldots, n\}$, edges $E \subseteq \binom{V}{2}$, and a (vertex labeled) collection of binary variables $\mathbf{X} = \{X_v | v \in V\}$. Let $\mathbf{X}_A = \{X_v | v \in A\}$ for any $A \subset V$. We construct a joint probability distribution over $\mathbf{X}$ as follows:

$$\Pr(\mathbf{X} = x) = \frac{1}{Z} \prod_{(u,v)\in E} (1 - x_u x_v) \tag{1}$$

for $\mathbf{x} = (x_v) \in \{0, 1\}^n$ and where Z is the normalization constant. By construction the distribution of $\mathbf{X}$ is uniform over all independent sets of G, and hence Z is the number of

[1]An experimental study conducted subsequent to our initial submission suggests that the error between the logarithms of the number of independent sets and the Bethe approximation does seem to be $O(1)$ for random 3-regular graphs [20].

independent sets in G. We will use the following notation throughout this paper: $\mathcal{N}(v)$ refers to the set of neighbors of $v \in V$, $d(v) = d_G(v) \triangleq |\mathcal{N}(v)|$ for $v \in V$, and $d \triangleq \max_v d(v)$.

2.1. Bethe approximation. We present the Bethe approximation for Z as a function of the induced node marginals $\{\tau_v\}_{v\in V}$ and pairwise edge marginals $\{\tau_{u,v}\}_{(u,v)\in E}$. The Bethe free energy (see [21]) is optimized over all $\{\tau_v\}$ and $\{\tau_{(u,v)}\}$ subject to the constraints that these are valid distributions and that the edge marginals are consistent with the node marginals. For the problem of interest discussed here, one can check that the following conditions must be satisfied:

$$\begin{gathered}\tau_{u,v}(0,1) = \tau_v(1), \quad \tau_{u,v}(1,0) = \tau_u(1),\\ \tau_{u,v}(1,1) = 0, \quad \tau_{u,v}(0,0) = 1 - \tau_v(1) - \tau_u(1).\end{gathered} \tag{2}$$

Consequently, we have the following simplified expression (cf. page 83 of [17]) for the Bethe free energy $F_B\colon [0,1]^n \to \mathbb{R}$ parameterized only by a vector $\mathbf{y} = (y_v) \in [0,1]^n$ that corresponds to the node marginals $\{\tau_v\}$ via $\tau_v(1) = y_v$:

$$\begin{aligned}F_B(\mathbf{y}) &\triangleq \sum_{v\in V} H(X_v) - \sum_{(u,v)\in E} I(X_u; X_v)\\ &\overset{\text{(a)}}{=} -\sum_{v\in V}(d(v)-1)H(X_v) + \sum_{(u,v)\in E} H(X_u, X_v)\\ &= \sum_{v\in V}(-y_v \ln y_v + (d(v)-1)(1-y_v)\ln(1-y_v))\\ &\quad - \sum_{(u,v)\in E}(1-y_u-y_v)\ln(1-y_u-y_v),\end{aligned}$$

where $H(\cdot)$ is the standard discrete entropy and $I(\cdot)$ is the mutual information. In the above, (a) follows from $I(X_u; X_v) = H(X_u) + H(X_v) - H(X_u, X_v)$.

Definition 1 (Bethe approximation). *Let* $\boldsymbol{\tau} = (\tau_v)_{v\in V}$ *be the node marginals corresponding to a stationary point of the Bethe free energy* F_B*. Then the Bethe approximation denoted by* $\ln Z_B$ *of* $\ln Z$*, the logarithm of the number of independent sets, is defined as*

$$\ln Z_B = \ln Z_B(\boldsymbol{\tau}) \triangleq F_B(\boldsymbol{\tau}(1)),$$

where $\boldsymbol{\tau}(1) = (\tau_v(1))_{v\in V} \in [0,1]^n$.

We note that the gradient of the Bethe free energy $\nabla F_B(\mathbf{y}) = [\frac{\partial F_B}{\partial y_v}]$ is such that

$$\frac{\partial F_B}{\partial y_v} = -(d(v)-1)\ln(1-y_v) + \ln y_v + \sum_{u\in\mathcal{N}(v)} \ln(1-y_u-y_v). \tag{3}$$

Let $\mathbf{y}^*$ be a zero-gradient point (or stationary point) of F_B, i.e., $\nabla F_B(\mathbf{y}^*) = \mathbf{0}$, with $\mathbf{y}^*$ strictly in the interior of $[0,1]^n$. From (3) it follows that

$$\frac{\prod_{u\in\mathcal{N}(v)}(1-y_v^*-y_u^*)}{(1-y_v^*)^{d(v)-1} y_v^*} = 1. \tag{4}$$

We also make the following observation about a collection of node marginals $\boldsymbol{\tau} = (\tau_v)_{v\in V}$ that correspond to a zero-gradient point of F_B, i.e., $\tau_v(1) = y_v^*$ and $\tau_v(0) = 1 - y_v^*$:

$$\tau_v(1) \leq \tau_v(0), \tag{5}$$

which easily follows from (4) where one can check that $y_v^* \leq 1/2$.

Belief propagation (BP) (see [13]) is a widely used heuristic for approximating the partition function Z with the key property that the fixed points of the BP iteration correspond to stationary points of the Bethe free energy [21] (see also [17] for more details). Therefore, if BP converges, one can directly compute the Bethe approximation for the partition function. Unfortunately, BP can fail to converge even for the independent set problem. We remedy this situation by describing a provably convergent algorithm for computing stationary points of the Bethe free energy (see section 4).

2.2. Error in Bethe approximation: Loop series correction. Recently Chertkov and Chernyak [4] showed that the partition function Z can be obtained by "correcting" the Bethe approximation Z_B as follows:

$$Z = Z_B\left(1 + \sum_{\emptyset \neq F \subseteq E} w(F)\right). \tag{6}$$

Here $F \subseteq E$ are (edge) subgraphs of G, and the explicit form of weight $w(F)$ can be obtained as follows (see Proposition 1 in [15]). For F with any node having degree 1, we have that $w(F) = 0$. For all other F, called *generalized loops*,

$$w(F) = (-1)^{|F|} \prod_{v\in V_F} \tau_v(1)\left[1 + (-1)^{d_F(v)}\left(\frac{\tau_v(1)}{\tau_v(0)}\right)^{d_F(v)-1}\right] \tag{7}$$

for the independent set problem. Here, $\boldsymbol{\tau} = (\tau_v)_{v\in V}$ represents the estimate of the node marginals at a stationary point of the Bethe free energy, i.e., $\tau_v(1) = y_v^*$ and $\tau_v(0) = 1 - y_v^*$.

2.3. Near-stationary points of the Bethe free energy. In practice one is typically not able to compute stationary points or, equivalently, zero-gradient points, of the Bethe free energy exactly. Thus, we introduce the concept of an ε-*gradient point*: $\mathbf{y}^*$ is said to be an ε-gradient point of the Bethe free energy if $\|\nabla F_B(\mathbf{y}^*)\|_2 \leq \varepsilon$. Based on the formula (3) for the gradient of the Bethe free energy F_B, we have that[2]

$$\frac{\prod_{u\in\mathcal{N}(v)}(1 - y_v^* - y_u^*)}{(1 - y_v^*)^{d(v)-1} y_v^*} = 1 \pm \varepsilon. \tag{8}$$

Indeed, the convergent algorithm that we describe in section 4 for computing stationary points of the Bethe free energy provides an ε-gradient point, where ε is an input parameter in the algorithm and the number of iterations required to compute an ε-gradient point depends on ε. Further, the $\mathbf{y}^*$ produced by the algorithm satisfies (5); i.e., $\mathbf{y}^* \leq (0, 1/2)^n$. See section 4 for more details.

[2] We say that $\beta = 1 \pm \varepsilon$ if $\beta \in [1 - \varepsilon, 1 + \varepsilon]$.

If we let $\boldsymbol{\tau}$ be the node marginals corresponding to an ε-gradient point, the Bethe approximation for $\ln Z$ denoted by $\ln Z_{B,\varepsilon}$ is defined as

$$\ln Z_{B,\varepsilon} = \ln Z_{B,\varepsilon}(\boldsymbol{\tau}) \triangleq F_B(\boldsymbol{\tau}).$$

In the following section, we provide a loop series correction result for $Z_{B,\varepsilon}$.

3. Loop series correction for $\boldsymbol{Z_{B,\varepsilon}}$. As discussed in section 2.2, the explicit loop series formula (6) relating Z and Z_B is known [4], [15]. However, the proofs in [4], [15] do not naturally extend to the case of $Z_{B,\varepsilon}$. This is essentially because an ε-gradient point may not necessarily be close to a zero-gradient or stationary point of F_B. To resolve this issue we present an "ε-version" of the loop series expansion.

THEOREM 1. *Let Z be the number of independent sets, let $\boldsymbol{\tau}$ be the node marginals corresponding to an ε-gradient point, and let $Z_{B,\varepsilon}$ be the corresponding Bethe approximation. Then,*

$$\frac{Z}{Z_{B,\varepsilon}} = (1 \pm \varepsilon)^{2n}\left(1 + \sum_{\emptyset \neq F \subseteq E} w(F)\right), \tag{9}$$

where $w(F) = (-1)^{|F|} \prod_{v \in V_F} \tau_v(1)[1 + (-1)^{d_F(v)} (\frac{\tau_v(1)}{\tau_v(0)})^{d_F(v)-1}]$.

Proof. We first start by recalling the proof in [15] for (6), which is the case $\varepsilon = 0$. The authors first show that

$$\frac{Z}{Z_B} = \sum_{\mathbf{x} \in \{0,1\}^n} \prod_{v \in V} \tau_v(x_v) \prod_{(u,v) \in E} \frac{\tau_{u,v}(x_u, x_v)}{\tau_u(x_u)\tau_v(x_v)}. \tag{10}$$

Second, they prove (see Proposition 1 of [15]) that

$$\sum_{\mathbf{x} \in \{0,1\}^n} \prod_{v \in V} \tau_v(x_v) \prod_{(u,v) \in E} \frac{\tau_{u,v}(x_u, x_v)}{\tau_u(x_u)\tau_v(x_v)} = 1 + \sum_{\emptyset \neq F \subseteq E} w(F). \tag{11}$$

In the independent set model of interest here, the edge marginals $\tau_{u,v}$ are determined by the node marginals τ_u, τ_v (see (2)). Therefore their proof for (11) still holds for a set of node marginals τ corresponding to an ε-gradient point with $\varepsilon > 0$. This is because their proof for (11) does not depend on the properties of stationary points of F_B (hence, also not on the quality of the ε-gradient points), but only on the fact that the edge marginals are consistent with the node marginals. Therefore to complete the proof of (9), it suffices to show the ε-version of (10), i.e.,

$$\frac{Z}{Z_{B,\varepsilon}} = (1 \pm \varepsilon)^{2n} \sum_{\mathbf{x} \in \{0,1\}^n} \prod_{v \in V} \tau_v(x_v) \prod_{(u,v) \in E} \frac{\tau_{u,v}(x_u, x_v)}{\tau_u(x_u)\tau_v(x_v)}, \tag{12}$$

where the summation is taken over independent sets $\mathbf{x} \in \{0,1\}^n$ and $\boldsymbol{\tau}$ represents the node marginals corresponding to an ε-gradient point $\mathbf{y}$, i.e., $\tau_v(1) = 1 - y_v$ and $\tau_v(0) = 1 - y_v$.

To this end, we obtain the following expression of $Z_{B,\varepsilon}$ in terms of $\mathbf{y}^*$:

$$\begin{aligned}
Z_{B,\varepsilon} &= e^{F_B(\mathbf{y})} \\
&= \prod_{v\in V} y_v^{-y_v} \times \prod_{v\in V} (1-y_v)^{(d(v)-1)(1-y_v)} \times \prod_{(u,v)\in E} (1-y_v-y_u)^{-1+y_v+y_u} \\
&\overset{(a)}{=} (1\pm\varepsilon)^n \prod_{v\in V} y_v^{-y_v} \times \prod_{v\in V} \left(\frac{\prod_{u\in\mathcal{N}(v)}(1-y_v-y_u)}{y_v} \right)^{1-y_v} \\
&\quad \times \prod_{(u,v)\in E} (1-y_v-y_u)^{-1+y_v+y_u} \\
&= (1\pm\varepsilon)^n \prod_{v\in V} y_v^{-1} \times \prod_{v\in V} \prod_{u\in\mathcal{N}(v)} (1-y_v-y_u)^{1-y_v} \\
&\quad \times \prod_{(u,v)\in E} (1-y_v-y_u)^{-1+y_v+y_u} \\
&= (1\pm\varepsilon)^n \prod_{v\in V} y_v^{-1} \times \prod_{(u,v)\in E} (1-y_v-y_u)^{2-y_v-y_u} \\
&\quad \times \prod_{(u,v)\in E} (1-y_v-y_u)^{-1+y_v+y_u} \\
&= (1\pm\varepsilon)^n \prod_{v\in V} y_v^{-1} \times \prod_{(u,v)\in E} 1-y_v-y_u, \qquad (13)
\end{aligned}$$

where (a) is from (8). On the other hand, for an independent set $\mathbf{x}\in\{0,1\}^n$, each term of the summation in (12) can be bounded as

$$\begin{aligned}
&\prod_{v\in V}\tau(x_v) \prod_{(u,v)\in E} \frac{\tau_{u,v}(x_u,x_v)}{\tau_v(x_v)\tau_u(x_u)} = \prod_{v:x_v=1} y_v \times \prod_{v:x_v=0} 1-y_v \times \prod_{v:x_v=1}\prod_{u\in\mathcal{N}(v)} \frac{\tau_{u,v}(0,1)}{\tau_v(1)\tau_u(0)} \\
&\times \prod_{(u,v)\in E:x_u=x_v=0} \frac{\tau_{u,v}(0,0)}{\tau_v(0)\tau_u(0)} = \prod_{v:x_v=1} y_v \times \prod_{v:x_v=0} 1-y_v \times \prod_{v:x_v=1}\prod_{u\in\mathcal{N}(v)} \frac{1}{\tau_u(0)} \\
&\times \prod_{(u,v)\in E:x_u=x_v=0} \frac{\tau_{u,v}(0,0)}{\tau_v(0)\tau_u(0)} = \prod_{v:x_v=1} y_v \times \prod_{v:x_v=0} 1-y_v \times \prod_{v:x_v=0} \left(\frac{1}{\tau_v(0)}\right)^{d(v)} \\
&\times \prod_{(u,v)\in E:x_u=x_v=0} \tau_{u,v}(0,0) = \prod_{v:x_v=1} y_v \times \prod_{v:x_v=0} \left(\frac{1}{1-y_v}\right)^{d(v)-1} \\
&\times \prod_{(u,v)\in E:x_u=x_v=0} 1-y_v-y_u \overset{(a)}{=} (1\pm\varepsilon)^n \prod_{v:x_v=1} y_v \times \prod_{v:x_v=0} \frac{y_v}{\prod_{u\in\mathcal{N}(u)} 1-y_u-y_v} \\
&\times \prod_{(u,v)\in E:x_u=x_v=0} 1-y_v-y_u = (1\pm\varepsilon)^n \prod_{v\in V} y_v \times \prod_{(u,v)\in E} \frac{1}{1-y_u-y_v} \\
&\overset{(b)}{=} (1\pm\varepsilon)^{2n} \frac{1}{e^{F_B(\mathbf{y})}} = (1\pm\varepsilon)^{2n} \frac{1}{Z_{B,\varepsilon}}, \qquad (14)
\end{aligned}$$

where (a) is from (8) and (b) is from (13). Therefore, (12) follows from (14). This completes the proof of Theorem 1. ∎

4. Fast, convergent algorithm for Bethe approximation. As discussed previously, BP is an iterative heuristic procedure that is widely used to compute stationary points of the Bethe free energy. However, BP does not always converge in general (e.g., see [17]); in fact, one can even construct examples for independent set problems in which BP fails to converge. Here, we propose a convergent BP-like alternative to compute stationary points of the Bethe free energy for the independent set problem. This procedure offers several of the advantages of BP in that it is a local iterative method, with the added benefit that it is always guaranteed to converge.

The algorithm computes an ε-gradient point of the Bethe free energy F_B with the number of iterations depending on ε. The Bethe approximation corresponding to such an ε-gradient point is sufficient for our purposes. We note here that computing such an ε-gradient point of F_B is not known to be easy in general since the underlying domain $[0,1]^n$ grows exponentially with respect to n.

4.1. Algorithm description. The algorithm described next computes $\mathbf{y}(t) = (y_v(t))_{v\in V}$ as an ε-gradient point of F_B. It is based on the standard gradient descent algorithm. The nontriviality lies in the choice of the appropriate step size, and subsequent analysis of correctness and rate of convergence.

- Algorithm parameters: number of iterations $T \geq 0$, $\mathbf{y}(t) = (y_v(t))_{v\in V}$. Initially, $t = 0$ and $y_v(0) = 1/4$, $v \in V$.
- $\mathbf{y}(t) = (y_v(t))_{v\in V}$ is updated until $t \leq T$:

$$y_v(t+1) = y_v(t) - \alpha(t) \frac{\partial F_B}{\partial y_v}\bigg|_{\mathbf{y}(t)},$$

  where $\alpha(t) = 1/(2^{d+7}(d^2+6d+2)\sqrt{t+1})$. Recall that

$$\frac{\partial F_B}{\partial y_v}\bigg|_{\mathbf{y}(t)} = \left((d(v)-1)\ln\,(1-y_v(t)) + \ln\, y_v(t) - \sum_{u\in\mathcal{N}(v)} \ln\,(1-y_u(t)-y_v(t))\right).$$

- Choose an $s \leq T$ with probability $\frac{\alpha(s)}{\sum_{t\leq T}\alpha(t)}$; output $\mathbf{y}(s) = (y_v(s))_{v\in V}$.

4.2. Properties of the algorithm: Correctness, convergence. Next we state and prove the correctness and convergence of the algorithm.

THEOREM 2. *Let $\mathbf{y}(t)$ be the sequence of iterates of the algorithm, with $\mathbf{y}(s)$ being the output chosen at random. Then, $\mathbf{y}(s) \in (0,1/2)^n$ and*

$$\mathbb{E}[\|\nabla F_B(y(s))\|_2^2] = O\left(\frac{nd^4 2^d \log\,T}{\sqrt{T}}\right),$$

where $\mathbb{E}[\|\nabla F_B(\mathbf{y}(s))\|_2^2] = \frac{1}{\sum_{t=0}^T \alpha(t)} \sum_{t=0}^T \alpha(t)\|\nabla F_B(y(t))\|_2^2$.

Choice of T. Theorem 2 implies that for $T = \Theta(n^2 d^4 2^d \varepsilon^{-4} \log^3(n/\varepsilon))$, the algorithm will produce an ε-gradient point of F_B for any $\varepsilon > 0$.

Proof of Theorem 2. Recall that $F_B\colon [0,1]^n \to \mathbb{R}$ is such that

$$\begin{aligned} F_B(\mathbf{y}) = &\sum_{v \in V} (y_v \ln y_v - (d(v)-1)(1-y_v) \ln (1-y_v)) \\ &+ \sum_{(u,v)\in E} (1 - y_u - y_v) \ln (1 - y_u - y_v). \end{aligned}$$

Now the updating rule of the algorithm is equal to

$$y_v(t+1) = y_v(t) - \alpha(t) \left.\frac{\partial F_B}{\partial y_v}\right|_{\mathbf{y}(t)}.$$

We start by establishing that under the dynamics of the above algorithm with the chosen initial condition and algorithm parameters, $y_v(t) \in [0, 1/2]$ for all $v \in V$ at all iterations t. For this we need the following three steps: with $\varepsilon_1 = 1/2^{d+2}$, $\varepsilon_2 = 1/2^{d+6}$,

$$\frac{\partial F_B}{\partial y_v} \le 0 \quad \text{if } y_v < 2\varepsilon_1 \quad \text{and} \quad \mathbf{y} \in D \triangleq \left[\varepsilon_1, \frac{1}{2} - \varepsilon_2\right]^n, \tag{15}$$

$$\frac{\partial F_B}{\partial y_v} \ge 0 \quad \text{if } y_v > \frac{1}{2} - 2\varepsilon_2 \quad \text{and} \quad \mathbf{y} \in D, \tag{16}$$

$$\left|\alpha \frac{\partial F_B}{\partial y_v}\right| \le \frac{1}{2} \min\{\varepsilon_1, \varepsilon_2\} \quad \text{if } \mathbf{y} \in D \quad \text{and} \quad \alpha \le \frac{1}{2^{d+7}(d^2 + 6d + 2)}. \tag{17}$$

From (15)–(17) it follows that $\mathbf{y}(t) \in D$; i.e., $y_v(t)$ does not hit 0 or $\frac{1}{2}$. Hence, we have that $y_v(\cdot) \in [0, 1/2]$ for all $v \in V$ under the algorithm's iterations.

Proof of (15). Observe that

$$\begin{aligned} \frac{\partial F_B}{\partial y_v} &= (d(v)-1) \ln (1-y_v) + \ln y_v - \sum_{u \in \mathcal{N}(v)} \ln (1 - y_u - y_v) \\ &\le \ln y_v - d \ln \left(\frac{1}{2} - y_v\right) = \ln \frac{y_v}{(\frac{1}{2} - y_v)^d} = \ln \frac{2^d y_v}{(1-2y_v)^d} \le \ln \frac{2^d y_v}{1 - 2dy_v} \le 0, \end{aligned}$$

where one can easily verify each step using the conditions $y_v \le 2\varepsilon_1 = \frac{1}{2^{d+1}}$ and $y_u \le \frac{1}{2}$ for $u \in \mathcal{N}(v)$.

Proof of (16). Consider the following:

$$\begin{aligned} \frac{\partial F_B}{\partial y_v} &= (d(v)-1) \ln (1-y_v) + \ln y_v - \sum_{u \in \mathcal{N}(v)} \ln (1 - y_u - y_v) \\ &\ge (d(v)-1) \ln (1-y_v) + \ln y_v - \sum_{u \in \mathcal{N}(v)} \ln (1 - \varepsilon_1 - y_v) \\ &= \ln \frac{y_v}{1-y_v} - \sum_{u \in \mathcal{N}(v)} \ln \frac{1 - \varepsilon_1 - y_v}{1 - y_v} \ge \ln \frac{y_v}{1-y_v} - \ln \frac{1 - \varepsilon_1 - y_v}{1-y_v} = \ln \frac{y_v}{1 - \varepsilon_1 - y_v} \\ &> \ln \frac{\frac{1}{2} - 2\varepsilon_2}{\frac{1}{2} + 2\varepsilon_2 - \varepsilon_1} \ge 0, \end{aligned}$$

where each step can be verified using $y_v > \frac{1}{2} - 2\varepsilon_2 = \frac{1}{2} - \frac{1}{2^{d+5}}$ and $y_u \geq \varepsilon_1 = \frac{1}{2^{d+2}}$ for $u \in \mathcal{N}(v)$.

Proof of (17). This follows from our choice of $\boldsymbol{\alpha}$, since for $\mathbf{y} \in D$

$$\begin{aligned}
\left|\frac{\partial F_B}{\partial y_v}\right| &= \left|(d(v)-1)\ln\,(1-y_v) + \ln\,y_v - \sum_{u\in\mathcal{N}(v)} \ln\,(1-y_u-y_v)\right| \\
&\leq \left|\ln\,\frac{y_v}{1-y_v} - \sum_{u\in\mathcal{N}(v)} \ln\,\frac{1-y_u-y_v}{1-y_v}\right| \\
&\leq -\ln\,\frac{y_v}{1-y_v} - \sum_{u\in\mathcal{N}(v)} \ln\,\frac{1-y_u-y_v}{1-y_v} \\
&\leq -\ln\,\frac{\varepsilon_1}{1-\varepsilon_1} - d\,\ln\,\frac{2\varepsilon_2}{\frac{1}{2}+\varepsilon_2} \leq -\ln\,\varepsilon_1 - d\,\ln\,2\varepsilon_2 \\
&= (d+2+d(d+5))\ln\,2 \leq d^2+6d+2, \qquad (18)
\end{aligned}$$

where each step follows from $\mathbf{y} \in [\varepsilon_1, \frac{1}{2} - \varepsilon_2]^n$.

We have established $\mathbf{y}(t) \in D$ as a consequence of the above three steps, which shows that the algorithm is well defined; i.e., $\mathbf{y}(t)$ is always in the valid domain D. Now we consider the dynamics

$$\mathbf{y}(t+1) = \mathbf{y}(t) - \boldsymbol{\alpha}(t)\nabla F_B(y(t)).$$

Using Taylor's expansion,

$$\begin{aligned}
F_B(\mathbf{y}(t+1)) &= F_B(\mathbf{y}(t) - \boldsymbol{\alpha}(t)\nabla F_B(\mathbf{y}(t))) \\
&= F_B(\mathbf{y}(t)) - \nabla F_B(\mathbf{y}(t))' \cdot \boldsymbol{\alpha}(t)\nabla F_B(\mathbf{y}(t)) \\
&\quad + \frac{1}{2}\boldsymbol{\alpha}(t)\nabla F_B(\mathbf{y}(t))' \cdot R \cdot \boldsymbol{\alpha}(t)\nabla F_B(\mathbf{y}(t)), \qquad (19)
\end{aligned}$$

where R is an $n \times n$ matrix such that

$$|R_{vw}| \leq \sup_{\mathbf{y}\in B}\left|\frac{\partial^2 F_B}{\partial y_v \partial y_w}\right|$$

and B is an L_∞-ball in $\mathbb{R}^n$ centered at $\mathbf{y}(t) \in D$ with its radius

$$r = \max_{v\in V}\left|\boldsymbol{\alpha}(t)\frac{\partial F_B}{\partial y_v}(\mathbf{y}(t))\right|.$$

From (17) we know $r \leq \frac{1}{2}\min\,\{\varepsilon_1, \varepsilon_2\}$. Hence, $\mathbf{y} \in [\varepsilon_1/2, \frac{1}{2} - \varepsilon_2/2]^n$ if $\mathbf{y} \in B$. Using this, we can get a bound for $\sup_{\mathbf{y}\in B}|\frac{\partial^2 F_B}{\partial y_v \partial y_w}|$ as follows:

- If $u = w$,

$$\begin{aligned}
\left|\frac{\partial^2 F_B}{\partial y_v^2}\right| &= \left|-\frac{d(v)-1}{1-y_v} + \frac{1}{y_v} + \sum_{u\in\mathcal{N}(v)}\frac{1}{1-y_u-y_v}\right| \\
&< \frac{1}{y_v} + 2\sum_{u\in\mathcal{N}(v)}\frac{1}{1-y_u-y_v} \leq \frac{2}{\varepsilon_1} + \frac{2d}{\varepsilon_2} = O(d2^d).
\end{aligned}$$

- If $w \in \mathcal{N}(v)$,

$$\left|\frac{\partial^2 F_B}{\partial y_v \partial y_w}\right| = \frac{1}{1 - y_w - y_v} \leq \frac{1}{\varepsilon_2} = O(2^d).$$

- Otherwise, $\frac{\partial^2 F_B}{\partial y_v \partial y_w} = 0$.

Therefore, using these bounds with (18), the equality (19) becomes

$$\begin{aligned} F_B(\mathbf{y}(t+1)) &\leq F_B(\mathbf{y}(t)) - \alpha(t)\|\nabla F_B(y(t))\|_2^2 + \alpha^2(t)O(|E|d^5 2^d) \\ &= F_B(\mathbf{y}(t)) - \alpha(t)\|\nabla F_B(\mathbf{y}(t))\|_2^2 + \alpha^2(t)O(nd^6 2^d). \end{aligned} \tag{20}$$

If we sum (20) over t from 0 to $T-1$, we have

$$F_B(\mathbf{y}(T)) \leq F_B(\mathbf{y}(0)) - \sum_{t=0}^{T-1} \alpha(t)\|\nabla F_B(\mathbf{y}(t))\|_2^2 + O(nd^6 2^d)\sum_{t=0}^{T-1}\alpha^2(t). \tag{21}$$

Since $|F_B(\mathbf{y})| = O(nd)$ for $\mathbf{y} \in D$, we obtain

$$\sum_{t=0}^{T-1} \alpha(t)\|\nabla F_B(y(t))\|_2^2 \leq O(nd) + O(nd^6 2^d)\sum_{t=0}^{T-1}\alpha^2(t). \tag{22}$$

Thus, we finally obtain the desired conclusion:

$$\begin{aligned} \mathbb{E}[\|\nabla F_B(\mathbf{y}(s))\|_2^2] &= \frac{1}{\sum_{t=0}^{T}\alpha(t)}\sum_{t=0}^{T}\alpha(t)\|\nabla F_B(\mathbf{y}(t))\|_2^2 \\ &\leq \frac{1}{\sum_{t=0}^{T-1}\alpha(t)}\left(O(nd) + O(nd^6 2^d)\sum_{t=0}^{T-1}\alpha^2(t)\right) \\ &\overset{(a)}{=} O\left(\frac{2^d d^2}{\sqrt{T}}\right)\left(O(nd) + O\left(\frac{nd^2}{2^d}\right)\log\, T\right) \\ &= O\left(\frac{nd^4 2^d \log\, T}{\sqrt{T}}\right), \end{aligned}$$

where (a) follows from our choice of $\alpha(t) = \Theta(\frac{1}{2^d d^2 \sqrt{t}})$. This completes the proof of Theorem 2. ▯

5. Correctness of $Z_{B,\varepsilon}$ for graphs with large girth. The algorithm in the previous section provides an ε-gradient point of the Bethe free energy F_B. This leads to the Bethe approximation, $\ln\, Z_{B,\varepsilon}$ (or $Z_{B,\varepsilon}$), of $\ln\, Z$ (or Z) for the (logarithm of the) number of independent sets of any graph G. Here we establish that the estimation $Z_{B,\varepsilon}$ is asymptotically close to the desired value Z for graphs with large girth. Formally the girth of a graph is the length of the shortest cycle (for trees it is ∞). The formal result is stated below.

THEOREM 3. *Let $g(G)$ be the girth of a graph G. If $g(G) > 8d \times \log_2 n$, then*

$$\frac{Z}{Z_{B,\varepsilon}} = (1 \pm \varepsilon)^{2n}(1 \pm O(n^{-\gamma})),$$

where $\gamma = 4(\frac{g(G)}{8d\, \log_2 n} - 1) > 0$.

We note here that this theorem when combined with the algorithm of the previous section gives a polynomial-time approximation alg orithm for counting independent sets in graphs with girth larger than $8d \log_2 n$ when the maximum degree d is $O(\log n)$.

5.1. Proof of Theorem 3 We start by introducing the notion of *apples*—a special class of connected subgraphs of G.

DEFINITION 2 (apple). *A connected edge subgraph $C \subseteq E$ of G is an apple if* (a) *it is a cycle, or* (b) *it is the union of a cycle and a line; i.e., two vertices $v_1, v_2 \in C$ have $d_C(v_1) = 1$, $d_C(v_2) = 3$, and $d_C(v) = 2$ for $v \in V_C \setminus \{v_1, v_2\}$.*

Given estimates for the node marginals $\tau_v(1)/\tau_v(0)$, $v \in V$, corresponding to an ε-gradient point of F_B and an apple $C \subseteq E$, define the weight of C as

$$\hat{w}(C) \triangleq \left(\prod\nolimits_{\{u,v\}\in C} \sqrt{\frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)}}\right)^{1/2d}. \tag{23}$$

As the first result, we will establish the following bound on the summation of weights over all apples. The proof is presented in section 5.2.

LEMMA 4. *Let $g = g(G)$ be the girth of G with $g(G) > 8d \log_2 n$. Then*

$$\sum_{C\subset E} \hat{w}(C) = O(n^{-\gamma})$$

over all apples C and where $\gamma = 4(\frac{g(G)}{8d \log_2 n} - 1) > 0$.

To establish Theorem 3 from the ε-version of loop series in Theorem 1, it is sufficient to show that

$$\sum_{\varnothing\neq F\subseteq E} |w(F)| = O(n^{-\gamma}). \tag{24}$$

We first bound the term $\sum_{\varnothing\neq F\subseteq E}|w(F)|$ by the summation $\sum_{C\subset E} \hat{w}(C)$ as follows. The proof is presented in section 5.3.

LEMMA 5. *For any graph G,*

$$1 + \sum_{\varnothing\neq F\subseteq E} |w(F)| \le e^{\sum_{C\subset E} \hat{w}(C)}.$$

Now from Lemmas 4 and 5, as well as the fact that $e^x = 1 + O(x)$ for $x = O(n^{-\gamma})$ with $\gamma > 0$, the desired bound (24) follows immediately. This completes the proof of Theorem 3.

5.2. Proof of Lemma 4. The key to the proof of Lemma 4 is to (*) bound the number of apples of a given size (i.e., the number of edges), and (**) bound the weight of an apple of a given size. As we shall show, under the large girth condition of Theorem 3, the product of (*) and (**) will decay exponentially in the size of the apple. This will prove the claim of Lemma 4.

To this end we first bound (**), i.e., the weight of an apple of a given size, say k. We state the following proposition.

PROPOSITION 6. *For any apple C of size k, $\hat{w}(C) < 2^{-(k/2d)}$.*

Proof. From the definition of $\hat{w}$ in (23) it is enough to show that

$$\sqrt{\frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)}} \leq \frac{1}{2}$$

for $(u, v) \in C$. Note that

$$\tau_v(1) + \tau_u(1) = \tau_{u,v}(0,1) + \tau_{u,v}(1,0) \leq 2\tau_{u,v}(0,0) = 2(1 - \tau_v(1) - \tau_u(1)),$$

where each inequality (or equality) follows from the properties[3] noted in (2) and (5). Thus we have

$$\tau_v(1) + \tau_u(1) \leq \frac{2}{3} \tag{25}$$

for $(u, v) \in C$. Also $\tau_v(1) \leq 1/2$ and $\tau_v(1) + \tau_v(0) = 1$. Using these, we obtain the desired bound:

$$\frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)} = \frac{\tau_u(1)}{1-\tau_u(1)}\frac{\tau_v(1)}{1-\tau_v(1)} \overset{(a)}{\leq} \left(\frac{\frac{\tau_u(1)+\tau_v(1)}{2}}{1-\frac{\tau_u(1)+\tau_v(1)}{2}}\right)^2 \overset{(b)}{\leq} \left(\frac{\frac{1}{3}}{1-\frac{1}{3}}\right)^2 = \frac{1}{4}.$$

Here (a) follows from Jensen's inequality and the convexity of $\log \frac{x}{1-x}$ when $0 \leq \tau_v(1)$, $\tau_u(1) \leq 1/2$. For (b) we use (25) and the monotonicity of $f(x) = \frac{x}{1-x}$. ∎

Next, we bound (*), i.e., the number of apples of a given size k.

PROPOSITION 7. *Given girth $g = g(G) > 8d \log_2 n$ for graph G, the number of apples of size k is at most $n^2(e^{2/c_1})^k$, where $c_1 = g/\ln n$.*

Proof. Let C be a given apple. If C has a degree 1 vertex, say v, then define it as its starting vertex; otherwise if C is a cycle, let the starting vertex be arbitrary. Now consider $T_v(G)$, the self-avoiding walk tree (cf. [11]) of G rooted at $v \in V$. It is easy to see that there is an injective map from the apples of size k with starting vertex v to the paths of length k (i.e., having a leaf at level k) starting at v in $T_v(G)$. Given this injection, it follows that the number of apples of size k with starting vertex v is at most the number of leaves at level k of $T_v(G)$. Now the number of nodes up to level $g/2$ (where g is the girth, $g = g(G)$) in $T_v(G)$ must be at most n, or else there will be two nodes in $T_v(G)$ at level up to $g/2$ that are copies of the same vertex, leading to the existence of a cycle of length less than g in G. For the very same reason, it also follows that any subtree of $T_v(G)$ must have at most n nodes up to (its) level $g/2$. Using these properties, it can be shown that the number of vertices (and hence leaves) up to level k of $T_v(G)$ is at most

$$n^{\lceil k/(g/2)\rceil} < n^{(2k/g)+1} = n(e^{2\ln n/g})^k.$$

Now since there are n possible starting vertices, the number of apples of size k is at most

$$n^2(e^{2\ln n/g})^k = n^2(e^{2/c_1})^k.$$

This completes the proof of Proposition 7. ∎

To complete the proof of Lemma 4, consider the following. From Propositions 6 and 7,

[3] As we discuss in section 2.3, the node marginals from an ε-gradient point produced by our algorithm satisfy (5).

$$\sum_{C\subset E}\hat{w}(C) \le \sum_{k\ge g} n^2(e^{2/c_1})^k 2^{-(k/2d)} = n^2\left(\sum_{k\ge g}\delta^k\right)$$
$$< n^2\left(\frac{\delta^g}{1-\delta}\right) = O(n^2 n^{c_1 \ln \delta}) = O(n^{-\gamma}).$$

Here we have used $c_1 = g/\ln n > \frac{8d}{\ln 2}$ and the definition

$$\delta \triangleq 2^{-(1/2d)}e^{2/c_1} = e^{-(1/2d)\ln 2 + 2/c_1} < 1.$$

5.3. Proof of Lemma 5. In this section we are going to use the following result by Bermond, Jackson, and Jaeger [3].

THEOREM 8. *Given a connected graph $G = (V, E)$ without a bridge (i.e., there is no edge $e \in E$ such that $G' = (V, E \setminus \{e\})$ is not connected), there exists a list of cycles so that every edge is contained in exactly four cycles of the list.*

Inspired by such a result, we say that a list of apples $\{C_i\}$ is a good decomposition of a given generalized loop F if it satisfies the following conditions:

$$F = \bigcup C_i \quad \text{and} \quad |w(F)| \le \prod_i \hat{w}(C_i).$$

Observe that the existence of a good decomposition for any generalized loop F is sufficient to complete the proof of Lemma 5. This is because

$$1 + \sum_{\varnothing \ne F \subseteq E} |w(F)| \le \prod_{C\subset E}(1+\hat{w}(C)) < \prod_{C\subset E} e^{\hat{w}(C)} = e^{\sum_{C\subset E}\hat{w}(C)}, \tag{26}$$

where the first inequality is due to existence of a good decomposition for any generalized loop F. Now we are left with proving the existence of a good decomposition for any generalized loop in order to complete the proof of Lemma 5. This is what we do next.

First some notation: given a list of apples $\{C_i\}$ and $F = \bigcup C_i$, for $(u, v) \in F$ let $N_{(u,v)}$ be the number of C_i that include (u, v); let $N_{\max} = \max_{(u,v)\in F} N_{(u,v)}$. We have the following result that uses Theorem 8.

PROPOSITION 9. *For any generalized loop $F = (V_F, E_F)$, there exists a list $\{C_i\}$ of apples with $N_{\max} \le 4$.*

Proof. Assume F is connected, or else apply the argument to each connected component separately. Now we will prove Proposition 9 by induction on $|V_F|$. The base case when $|V_F| = 3$ is trivial. Further, if F has no bridge, Proposition 9 follows from Theorem 8 since there exists a list $\{C_i\}$ of cycles (and hence apples) which cover every edge exactly 4 times. Hence $N_{\max} = 4$. Now suppose F has a bridge. Then, we first claim that

(*) F has a bridge e such that F_1 is a bridgeless graph of size > 1, where F_1 is a connected component of $F \setminus \{e\}$; i.e., $F \setminus \{e\} = F_1 \bigcup F_2$.

The claim (*) follows from the following recursive argument. Suppose a bridge $e(0) = e$ does not satisfy the claim; this means that both $F_1(0) = F_1$ and $F_2(0) = F_2$ have bridges $e_1(0)$ and $e_2(0)$, respectively. Both $e_1(0)$ and $e_2(0)$ become bridges of F as well. Consider $e(1) = e_1(0)$ with $F \setminus \{e(1)\} = F_1(1) \bigcup F_2(1)$, and suppose $e(0) \in F_2(1)$ without loss of generality. Then, either $F_1(1)$ is bridgeless or $F_1(1)$ has a bridge. In case $F_1(1)$ is bridgeless, $e(1)$ is the desired bridge of F. Otherwise if $F_1(1)$ has a bridge, then we can recursively find a new bridge $e(2)$ in $F_1(1)$. However, the size of $F_1(1)$ is strictly smaller than that of the previous component $F_1(0)$ since $e \in F_2(1)$. We can recursively reduce the size of $F_1(\cdot)$ until we find the desired bridge $e(\cdot) \in F_1(\cdot)$. Since the size of $F_1(\cdot)$ is always greater than 1 (otherwise one of the vertices in the bridge $e(\cdot)$ has a degree 1, hence contradicting the fact that F is a generalized loop), this recursive procedure eventually finds the desired $e(\cdot) \in F_1(\cdot)$.

Let $e = (u, v)$ be the bridge in the claim (*), where $u \in F_1$ and $v \in F_2$. There are two cases: (a) $d_F(v) = 2$, and (b) $d_F(v) \geq 3$. First consider the case (a), and let w be another neighbor of v other than u. If we remove v and add a new edge (u, w) in F, the new graph F' is also a generalized loop. Since $|V_{F'}| = |V_F| - 1$, we can apply the induction hypothesis and find a list $\{C'_i\}$ of apples with $N_{\max} \leq 4$ on $V_{F'}$. The desired list $\{C_i\}$ to cover F is naturally constructible from $\{C'_i\}$ by adding the vertex v to C'_i, which includes (u, w). Now consider the case (b). In this case, F_2 is a generalized loop. Hence from the induction hypothesis, we can find the desired list $\{C^2_i\}$ of apples to cover F_2. On the other hand, since F_1 is a bridgeless graph, F_1 is covered by a list $\{C^1_i\}$ of cycles with $N_{\max} \leq 4$ from Theorem 8. Without loss of generality, let C^1_1 be the cycle which covers the vertex u. Then, the desired list of apples is $\{C^1_1 \bigcup (u, v)\} \bigcup \{C^1_i : i \geq 2\} \bigcup \{C^2_i\}$. This completes the induction. □

Finally, to complete the proof of Lemma 5, consider the list of apples produced by Proposition 9 and observe that

$$\prod_i \hat{w}(C_i) = \left(\prod_i \prod_{(u,v)\in C_i} \sqrt{\frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)}}\right)^{1/2d} \geq \left(\prod_i \prod_{(u,v)\in C_i} \sqrt{\frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)}}\right)^{2/(N_{\max}\cdot d)}$$
$$= \left(\prod_{(u,v)\in F} \left(\sqrt{\frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)}}\right)^{N_{(u,v)}}\right)^{2/(N_{\max}\cdot d)} \geq \left(\prod_{(u,v)\in F} \frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)}\right)^{1/d},$$

where we use $\frac{\tau_v(1)}{\tau_v(0)} \leq 1$ and $N_{\max} \leq 4$ for the inequalities. Hence, we obtain the desired bound:

$$\prod_i \hat{w}(C_i) \geq \left(\prod_{(u,v)\in F} \frac{\tau_u(1)}{\tau_u(0)}\frac{\tau_v(1)}{\tau_v(0)}\right)^{1/d} = \left(\prod_{v\in V_F} \left(\frac{\tau_v(1)}{\tau_v(0)}\right)^{d_F(v)}\right)^{1/d} \geq \prod_{v\in V_F} \frac{\tau_v(1)}{\tau_v(0)}$$
$$= \prod_{v\in V_F} \tau_v(1)\left[1 + \frac{\tau_v(1)}{\tau_v(0)}\right] \geq \prod_{v\in V_F} \tau_v(1)\left[1 + (-1)^{d_F(v)}\left(\frac{\tau_v(1)}{\tau_v(0)}\right)^{d_F(v)-1}\right]$$
$$= |w(F)|,$$

where the inequalities follow from $\frac{\tau_v(1)}{\tau_v(0)} \leq 1$.

6. Correctness of Z_B for random 3-regular graphs.

In this section we consider the error in the Bethe approximation for a random 3-regular graph. To obtain sharp results, we will utilize the SCCC of Alon and Tarsi [1].

CONJECTURE 10 (SCCC). *Given a bridgeless graph G with m edges, all of its edges can be covered by a collection of cycles with the sum of their lengths being at most $7m/5 = 1.4m$.*

We have the following result that implies that the difference between the Bethe approximation $\ln Z_B$ and $\ln Z$ is uniformly bounded, independent of n, with probability 1.

THEOREM 11. *Let G be chosen uniformly at random among all 3-regular graphs with n vertices. Assuming that the SCCC is true, there exists a function $f : (0, 1) \to \mathbb{R}^+$ so that*

$$|\ln Z - \ln Z_B| \le f(\varepsilon) \quad \text{with probability } 1-\varepsilon,$$

where $\frac{1}{n} \ln Z_B \approx \ln 1.545$.

6.1. Proof of Theorem 11. From (6) it is equivalent to show that

$$\left| \ln \left(1 + \sum_{\varnothing \neq F \subseteq E} w(F) \right) \right| \le f(\varepsilon) \quad \text{with probability } 1-\varepsilon.$$

Similar to the case of large-girth graphs, we consider $\sum_{\varnothing \neq F \subseteq E} |w(F)|$. First, we show that it is less than $h(\varepsilon)$ with probability $1-\varepsilon$ for some function $h\colon (0,1) \to \mathbb{R}^+$. This gives us an upper bound, i.e.,

$$\ln \left(1 + \sum_{\varnothing \neq F \subseteq E} w(F) \right) \le \ln \left(1 + \sum_{\varnothing \neq F \subseteq E} |w(F)| \right) \le \ln\,(1 + h(\varepsilon)). \tag{27}$$

The details are explained in section 6.2.

If we have $h(\varepsilon)$ uniformly bounded below 1, say always at most $1/2$, for example, then (27) would be sufficient to establish the claim of Theorem 11. In order to show this, we need additional proof techniques to obtain an appropriate lower bound on the quantity of interest. This lower bounding technique needs a longer explanation and is presented in section 6.3. Note that our lower bounding technique is essentially an algorithm that tries to "correct" the error in the Bethe approximation in a systematic manner by means of the loop series characterization.

6.2. Upper bound. As discussed in section 6.1, we show that $\sum_{\varnothing \neq F \subseteq E} |w(F)|$ is less than $h(\varepsilon)$ with probability $1-\varepsilon$. To this end it is enough to prove that

$$\mathbb{E}\left[\ln \left(1 + \sum_{\varnothing \neq F \subseteq E} |w(F)| \right) \right] = O(1). \tag{28}$$

If (28) holds, we can choose $h(\varepsilon) = e^{O(1/\varepsilon)} - 1$ by Markov's inequality. If G is a 3-regular graph, we can find the explicit homogeneous stationary point of F_B. From 4 and setting $y_v^* = z$ for all $v \in V$, we obtain

$$\frac{(1-2z)^3}{(1-z)^2 z} = 1,$$

where such a z can be found numerically to be $z \approx 0.241$. Furthermore, the corresponding Z_B can be calculated as $\ln Z_B \approx n \ln 1.545$.

LEMMA 12. *If G is a 3-regular graph and the SCCC is true, then*

$$\ln \left(1 + \sum_{\varnothing \neq F \subseteq E} |w(F)| \right) \le \sum_{C \subset E} \tilde{w}(C)$$

over all apples C and where $\tilde{w}(C) = \alpha^{|C|}$ and $\alpha \triangleq (z(1-z))^{2/(3\times 1.4)} \approx 0.48$.

Proof. The proof of this lemma uses arguments similar to those used to establish Lemma 5. Specifically, it suffices to find a good decomposition (list of apples) $\{C_i\}$ for any generalized loop F such that

$$F = \bigcup C_i \quad \text{and} \quad |w(F)| \le \prod_i \tilde{w}(C_i).$$

Using arguments similar to those used to establish Proposition 9, but with SCCC replacing Theorem 8, it can be guaranteed that there exists a list of apples, $\{C_i\}$, such that

$$F = \bigcup C_i \quad \text{and} \quad \sum_i |C_i| \leq 1.4 \times |F|. \tag{29}$$

Then

$$\begin{aligned} \prod_i \tilde{w}(C_i) &= \prod_i \alpha^{|C_i|} \\ &= \alpha^{\sum_i |C_i|}. \end{aligned} \tag{30}$$

On the other hand, $|w(F)|$ can be bounded in terms of α as follows:

$$\begin{aligned} |w(F)| &= \prod_{v \in V_F} \tau_v(1) \left[1 + (-1)^{d_F(v)} \left(\frac{\tau_v(1)}{\tau_v(0)}\right)^{d_F(v)-1}\right] \\ &\overset{(a)}{\leq} \prod_{v \in V_F} (z(1-z))^{(d_F(v))/3} \\ &= \prod_{v \in V_F} \alpha^{((3\times 1.4)/2)\times((d_F(v))/3)} \\ &= \alpha^{\sum_{v\in V_F}(1.4\times d_F(v))/2} = \alpha^{1.4\times|F|}, \end{aligned} \tag{31}$$

where the inequality (a) can be established in each of the possible cases $d_F(v) = 0, 1, 2, 3$ using the explicit values of $\tau_v(0) = 1 - z \approx 0.759$ and $\tau_v(1) = z \approx 0.241$. Further, the inequality (a) is tight only when $d_F(v) = 3$. Therefore, from (29)–(31) (and the fact that $\alpha < 1$) we have

$$|w(F)| \leq \prod_i \tilde{w}(C_i). \qquad \square$$

It follows from Lemma 12 that to establish (28) we need

$$\mathbb{E}\left[\sum_{C \subset E} \tilde{w}(C)\right] = O(1). \tag{32}$$

Let R_k, A_k be the number of cycles and apples, respectively, of size k in a 3-regular graph. Then

$$A_k \leq \sum_{i \leq k} R_i \times i \times 2^{k-i}, \tag{33}$$

since apples can be made only by attaching a line to a cycle. It is well known [19], [8] that the expected value of R_k for random 3-regular graphs is at most $2^{k-1}/k$. Using this fact and (33), it follows that the expected value of A_k for random 3-regular graphs is at most $k2^{k-1}$. Therefore, the desired bound (32) can be obtained as

$$\mathbb{E}\left[\sum_{C \subset E} \tilde{w}(C)\right] = \mathbb{E}\left[\sum_k A_k \alpha^k\right] \leq \sum_k k 2^{k-1} \alpha^k = O(1),$$

where the last inequality follows from $\alpha \approx 0.48$ in Lemma 12.

6.3. Lower bound. Using (33), it follows that

$$\begin{aligned}\sum_{C\subset E}\tilde{w}(C) &= \sum_k A_k\alpha^k \le \sum_k\sum_{i\le k} R_i\times i\times 2^{k-i}\alpha^k\\ &= \sum_i\sum_{k\ge i} R_i\times i\times 2^{k-i}\alpha^k\\ &= \sum_i R_i\frac{i}{2^i}\sum_{k\ge i}(2\alpha)^k \le \sum_i R_i\frac{i}{2^i}\frac{(2\alpha)^i}{1-2\alpha}\\ &= \frac{1}{1-2\alpha}\sum_i R_i i(\alpha)^i \le c_\alpha\times\sum_i R_i(0.49)^i \triangleq \rho(G),\end{aligned}\tag{34}$$

where c_α is a constant that depends on α and the last inequality is due to $\alpha\approx 0.48$. One naive way to define c_α is as follows:

$$c_\alpha = \frac{1}{1-2\alpha}\max_i\frac{i(\alpha)^i}{(0.49)^i},$$

where c_α is a finite constant since $\alpha\approx 0.48 < 0.49$. Once c_α is defined in this manner, the last inequality in (34) holds trivially since each term $\frac{1}{1-2\alpha}R_i i(\alpha)^i$ is dominated by the corresponding term $c_\alpha\times R_i(0.49)^i$ for all i. We state the following lemma, which is key to the proof of the lower bound.

LEMMA 13. *Given a random 3-regular graph G on n vertices, there exists another 3-regular graph $G'=(V',E')$ with $V\subset V'$ such that with probability (over the random choice of G) $1-\varepsilon$, we have*

1. $|\ln Z_B(G') - \ln Z_B(G)| < \Gamma(\varepsilon)$,
2. $|\ln Z(G') - \ln Z(G)| < \Gamma(\varepsilon)$, *and*
3. $\rho(G') < 0.5$.

Here $\Gamma(\varepsilon)$ is some ε dependent constant, independent of n.

The proof of this lemma is deferred to section 6.3.1. We show how it implies the desired lower bound. Since $\rho(G') < 0.5$,

$$\begin{aligned}\frac{Z(G')}{Z_B(G')} &= 1 + \sum_{\emptyset\ne F\subseteq E'} w(F)\\ &\ge 1 - \sum_{\emptyset\ne F\subseteq E'}|w(F)|\\ &\overset{(a)}{\ge} 1 - \left(e^{\sum_{C\subset E'}\tilde{w}(C)} - 1\right)\\ &\overset{(b)}{>} 2 - e^{0.5} > 0.3,\end{aligned}$$

where (a) is from Lemma 12[4] under the SCCC assumption, and (b) follows from (34) and $\rho(G') < 0.5$. Using properties 1 and 2 of Lemma 13, it follows that $\ln Z(G) - \ln Z_B(G) > -2\Gamma(\varepsilon) - O(1)$, which completes the proof of the lower bound.

[4]Recall that Lemma 12 uses only the fact that the graph under consideration is 3-regular, but does not require it to be "random."

6.3.1. Proof of Lemma 13. We start by defining the operator $\odot$ on 3-regular graphs. Figure 1 illustrates the definition of $\odot$.

DEFINITION 3. *Given connected* 3*-regular graphs* $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ *and an edge* $e = (v_1, v_2) \in E_1$, *create a new* 3*-regular* $(G_1, e) \odot G_2$ *as follows:*

1. *Construct the union of* G_1 *and* G_2, $G = G_1 \bigcup G_2$*—a disconnected graph with connected components as* G_1 *and* G_2.
2. *Add the two vertices* v_1, v_2 *that compose the edge* e, *and remove an edge, say* (v_3, v_4), *from* G_2 *arbitrarily.*
3. *Remove edge* (v_1, v_2) *from* G_1 *and add edges* $e_1 = (v_1, v_3)$, $e_2 = (v_2, v_4)$.
4. *Finally, contract* e_1 *and* e_2.

We study the effect of operator $\odot$ on the function ρ defined in (34). Let $G_3 = (G_1, e) \odot G_2$; we are interested in bounding $\rho(G_3)$ in terms of $\rho(G_1)$ and $\rho(G_2)$. By the definition (34) we have that $\rho(G_3)$ is a summation of terms over simple cycles of G_3. Simple cycles in G_3 can be classified into three types: (a) cycles in $G_1 \setminus \{e\}$, (b) cycles in G_2, and (c) cycles which intersect both G_1 and G_2. For cycles of types (a) and (b), the contribution to $\rho(G_3)$ is at most $\rho(G_1 \setminus \{e\})$ and $\rho(G_2)$, respectively. On the other hand, consider simple cycles of type (c). Specifically, let R_3 be one such simple cycle. Then it can be thought of as the union of $R_1 \setminus \{e\}$ and $R_2 \setminus \{e_2\}$ for some $e_2 \in R_2$, where R_1 and R_2 are cycles in G_1 and G_2, respectively. For this reason $|R_3| = |R_1| + |R_2|$, and it follows that the contribution of R_3 to $\rho(G_3)$ is at most

$$(0.49)^{|R_3|} = (0.49)^{|R_1|} \times (0.49)^{|R_2|}.$$

Using this, the contribution of the cycles of type (c) to $\rho(G_3)$ can be bounded as

$$c_\alpha \times \frac{(\rho(G_1) - \rho(G_1 \setminus \{e\}))}{c_\alpha} \times \frac{\rho(G_2)}{c_\alpha} = \frac{(\rho(G_1) - \rho(G_1 \setminus \{e\})) \times \rho(G_2)}{c_\alpha},$$

where $\rho(G_1) - \rho(G_1 \setminus \{e\})$ describes the contribution of cycles containing $\{e\}$ to $\rho(G_1)$. Thus

$$\rho(G_3) \le \rho(G_1 \setminus \{e\}) + \rho(G_2) + (\rho(G_1) - \rho(G_1 \setminus \{e\})) \times \rho(G_2) \times \frac{1}{c_\alpha}. \tag{35}$$

Therefore, if $\rho(G_2) < \min\{\frac{\rho(G_1) - \rho(G_1 \setminus \{e\})}{B}, \frac{c_\alpha}{B}\}$ with $B \ge 2$, $\rho(G_3)$ can be bounded as follows:

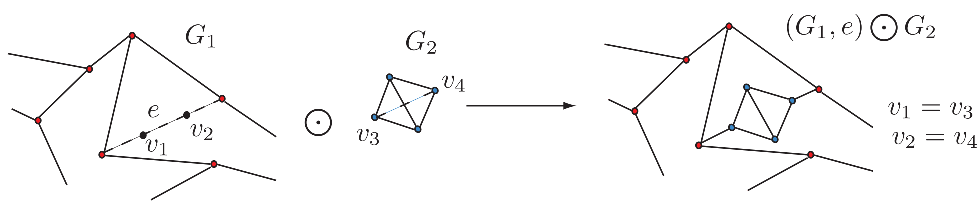

FIG. 1. 3*-regular* $(G_1, e) \odot G_2$ *is created from* 3*-regular graphs* G_1 *and* G_2 *as per Definition* 3.

$$\rho(G_3) \le \rho(G_1 \setminus \{e\}) + \frac{\rho(G_1) - \rho(G_1 \setminus \{e\})}{B} + \frac{\rho(G_1) - \rho(G_1 \setminus \{e\})}{B}$$
$$= \rho(G_1 \setminus \{e\}) + \frac{2}{B}(\rho(G_1) - \rho(G_1 \setminus \{e\}))$$
$$= \rho(G_1) - \left(1 - \frac{2}{B}\right)(\rho(G_1) - \rho(G_1 \setminus \{e\})). \tag{36}$$

Equipped with our understanding of the effect of $\bigodot$ on ρ and (36), we describe the following procedure for constructing the graph G' desired in Lemma 13. Given a random 3-regular graph G, generate G' iteratively as follows:

- Initially, let $t = 0$, and let $G'(0) = G$.
- Let g be the smallest number such that $c_\alpha \sum_{i \ge g} R_i (0.49)^i < 0.25$, where R_i is the number of cycles of length i in G.
- Repeat the following until $G'(t)$ is left with no cycle of length less than g:
  1. Let R be the smallest cycle in $G'(t)$. Choose an edge $e_t \in R$ arbitrarily.
  2. Set $G'(t+1) = (G'(t), e_t) \bigodot G_2$, where G_2 has a 3-regular graph that will be chosen later.
  3. Increment t by 1.
- Output $G' = G'(t)$.

First observe the following properties (*) and (**):

(*) $\ln Z_B(G'(t+1)) = \ln Z_B(G'(t)) + \ln Z_B(G_2)$

since $G'(t+1)$, $G'(t)$, G_2 are all 3-regular and $\ln Z_B$ is just a linear function in the number of their vertices.

(**) $\ln Z(G'(t+1)) \le \ln Z(G'(t)) + \ln Z(G_2)$

since any independent set in $G'(t+1)$ can be decomposed into two independent sets in $G'(t)$ and G_2, respectively. In other words (*) and (**) imply that $\ln Z_B(G'(t))$ and $\ln Z(G'(t))$ increase by at most a constant additive factor per round if the size of G_2 is a constant. Equipped with these observations, for establishing that G' thus produced has properties 1–3 of Lemma 13, it is sufficient to show that with probability $1-\varepsilon$ the repeat-loop in the above procedure terminates in $\Gamma_1(\varepsilon)$ steps, $\rho(G') < 0.5$, and G_2 is of size $\Gamma_2(\varepsilon)$. Here and in what follows $\Gamma_1(\varepsilon)$, $\Gamma_2(\varepsilon)$, $\ldots$, are constants dependent on ε and independent of n. Recall the definition of ρ in (34):

$$\rho(G) = c_\alpha \sum_i R_i (0.49)^i.$$

For a random 3-regular graph, we have

$$\mathbb{E}[R_i] \le 2^{i-1}/i.$$

Therefore, if we define appropriately large constants $g = \Gamma_3(\varepsilon)$ and $\Gamma_1(\varepsilon)$ so that

$$c_\alpha \sum_{i \ge g} \mathbb{E}[R_i](0.49)^i \le c_\alpha \sum_{i \ge g} \frac{2^{i-1}}{i}(0.49)^i < 0.25 \times \frac{\varepsilon}{2} \quad \text{and} \quad \Gamma_1(\varepsilon) = \frac{2}{\varepsilon} \sum_{i<g} \mathbb{E}[R_i],$$

the following two events happen simultaneously with probability $1-\varepsilon$ from Markov's inequality and the union bound:

$$\sum_{i<g} R_i \le \Gamma_1(\varepsilon) \quad \text{and} \quad c_\alpha \sum_{i \ge g} R_i (0.49)^i < 0.25.$$

Clearly, under these events the repeat-loop of the procedure to generate G' will terminate in $\Gamma_1(\varepsilon)$ steps as long as the graph G_2 is such that it has girth larger than g. Therefore, the only remaining step toward completing the proof of Lemma 13 is to establish existence of graph G_2 such that (a) it has size $\Gamma_2(\varepsilon)$, (b) it has girth larger than $g = \Gamma_3(\varepsilon)$, and (c) the resulting G' has $\rho(G') < 0.5$. Suppose G_2 can be chosen so that for all rounds $t \le \Gamma_1(\varepsilon)$ with $B \ge 2$,

$$\rho(G_2) \le \min\left\{\frac{\rho(G'(t)) - \rho(G'(t)\setminus\{e_t\})}{B}, \frac{c_\alpha}{B}\right\}. \tag{37}$$

Under this assumption, we obtain the following bound on $\rho(G')$ using (36) recursively:

$$\begin{aligned}
\rho(G') &\le \rho(G) - \sum_t \left(1 - \frac{2}{B}\right)(\rho(G'(t)) - \rho(G'(t)\setminus\{e_t\})) \\
&\le \rho(G) - \left(1 - \frac{2}{B}\right)\left(\sum_t \rho(G'(t)) - \rho(G'(t)\setminus\{e_t\})\right) \\
&\overset{(*)}{\le} \rho(G) - \left(1 - \frac{2}{B}\right) c_\alpha \left(\sum_{i<g} R_i (0.49)^i\right) \\
&\overset{(**)}{<} \rho(G) - \left(1 - \frac{2}{B}\right)(\rho(G) - 0.25) \\
&\le \frac{2}{B}\rho(G) + 0.25.
\end{aligned}$$

Here, (*) is due to the fact that each cycle of length up to g is "broken" in one of the steps $t \le \Gamma_1(\varepsilon)$. Each term $\rho(G'(t)) - \rho(G'(t)\setminus\{e_t\})$ accounts for all broken cycles in round t. Therefore, the bound used in (*) follows. For (**), by definition of g we have $c_\alpha \sum_{i\ge g} R_i (0.49)^i < 0.25$. Therefore, if we choose $B = 8\rho(G)$, the desired bound $\rho(G') < 0.5$ follows.

In summary, we are now left with showing the existence of G_2 which has properties (a) size $\Gamma_2(\varepsilon)$, (b) girth larger than $g = \Gamma_3(\varepsilon)$, and (c) the condition 37 with $B = 8\rho(G)$. The choice of B suggests that $B = \Gamma_4(\varepsilon)$ (due to selection of an event that has probability at least $1 - \varepsilon$). Consider

$$\rho(G'(t)) - \rho(G'(t)\setminus\{e\}) \ge c_\alpha (0.49)^g = \Gamma_5(\varepsilon), \tag{38}$$

where we have used the fact that for $t \le \Gamma_1(\varepsilon)$, a cycle of length at most g is broken and its corresponding contribution to $\rho(\cdot)$ is accounted for in the above difference. Therefore, we have

$$\frac{\rho(G'(t)) - \rho(G'(t)\setminus\{e\})}{B} \ge \Gamma_5(\varepsilon)/\Gamma_4(\varepsilon). \tag{39}$$

Hence to satisfy (c), it is sufficient to show that there exists G_2 with arbitrarily small $\rho(G_2)$ and girth value with size dependent on the "smallness" of $\rho(G_2)$. But if we establish existence of such a G_2, then the condition (a) about its size follows immediately, and the girth condition (b) will follow from the definition of ρ. This is established precisely in the following proposition.

PROPOSITION 14. *For any $\delta > 0$, there exists a 3-regular graph G_2 such that $\rho(G_2) < \delta$. Further, its girth is at least $\log_{1/0.49}(\frac{c_\alpha}{\delta})$.*

Proof. Recall that R_i is the number of cycles of length i in the graph G_2. For a random 3-regular graph, it is well known [19] that for $3 \le r \le \frac{1}{3}\log n$, R_r become asymptotically independent Poisson random variables with mean $\mu_r = 2^{r-1}/r$. Thus, for $g < \frac{1}{3}\log n$,

$$\Pr\,[R_3 = R_4 = \cdots = R_g = 0] \approx e^{-e^{\Theta(g)}}. \tag{40}$$

We divide the summation $c_\alpha \sum_i R_i(0.49)^i = \sum_i a_i$ with $a_i \triangleq c_\alpha R_i(0.49)^i$ into the following three terms: $A_1 = \sum_{r<g} a_r$, $A_2 = \sum_{g \le r < \frac{1}{3}\log n} a_r$, and $A_3 = \sum_{g \ge \frac{1}{3}\log n} a_r$. Define the events E_1, E_2, and E_3 such that E_1 is the event $A_1 = 0$, E_2 is the event $A_2 \le 2\mathbb{E}[A_2]$, and E_3 is the event $A_3 \le (3\log n)\mathbb{E}[A_3]$. From Markov's inequality, $\Pr\,[E_2] \ge \frac{1}{2}$ and $\Pr\,[E_3] \ge 1 - \frac{1}{3\log n}$. For E_1 one can choose $g = \Theta(\log\ \log\ \log n)$ from 40 such that $\Pr\,[E_1] = \frac{1}{\log n}$. Therefore, we have

$$\Pr\,[E1 \cap E_2 \cap E_3] \ge \Pr\,[E1 \cap E_2] + \Pr\,[E_3] - 1 \ge \frac{1}{2\log n} + 1 - \frac{1}{3\log n} - 1 > 0,$$

where we use the union bound and the independence between E_1 and E_2.[5] In other words, all events E_1, E_2, and E_3 happen with strictly positive probability. Under the events E_1, E_2, and E_3,

$$\rho(G_2) \le 2\mathbb{E}[A_2] + (3\log n)\mathbb{E}[A_3] \le O(1) \times (0.98)^g + O(3\log n) \times (0.98)^{1/3\log n} \to 0$$

as n goes to ∞ since $g = \Theta(\log\ \log\ \log n)$ also goes to ∞. Here, we have used the fact that $\mathbb{E}[R_r] \le 2^{r-1}/r$. In conclusion, there exists a 3-regular graph G_2 such that $\rho(G_2)$ is arbitrarily small. Finally, the bound on the girth follows immediately from the definition of ρ. ∎

7. Conclusion. In this paper we considered the Bethe approximation for counting independent sets in an arbitrary graph. We presented a simple message-passing algorithm that converges to a near stationary point of the Bethe free energy for the independent set problem for *any* graph. Our algorithm finds an ε-gradient point in $O(n^2\varepsilon^{-4}\log^3(n\varepsilon^{-1}))$ iterations for bounded degree graphs on n nodes. The algorithm can be viewed as a ``time-varying'' modification of the usual BP algorithm. Therefore, our algorithm (and its adaptation to other problems) provides a fast, convergent message-passing alternative to BP.

Next, to quantify the error in the Bethe approximation based on an ε-gradient point produced by our algorithm, we provide an ε-version of the loop series expansion approach of Chertkov and Chernyak. This does not naturally follow from the proofs in [4], [15] since they crucially depend on the exactness of the stationary point of F_B.

Finally using this ε-version of the loop calculus, we establish that for *any* graph with sufficiently large girth the error in the Bethe approximation for the number of independent sets is essentially $O(n^{-\gamma})$ for some $\gamma > 0$. In addition we find that for random 3-regular graphs, the Bethe approximation of the log-partition function (i.e., the logarithm of the number of independent sets) is within $O(1)$ (with high probability) of the correct log-partition function assuming the SCCC of Alon and Tarsi; thus, either the SCCC is false or the Bethe approximation is extremely good and stronger than the prediction of physicists.

[5]For the independence between E_1 and E_2, we use the fact that events of cycles of length $< \frac{1}{3}\log n$ are asymptotically independent from [19] as $n \to \infty$.

REFERENCES

[1] N. Alon and M. Tarsi, *Covering multigraphs by simple circuits*, SIAM J. Algebr. Discrete Methods, 6 (1985), pp. 345–350.

[2] A. Bandyopadhyay and D. Gamarnik, *Counting without sampling. New algorithms for enumeration problems using statistical physics*, in Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, SIAM, Philadelphia, 2006, pp. 890–899.

[3] J. C. Bermond, B. Jackson, and F. Jaeger, *Shortest coverings of graphs with cycles*, J. Combin. Theory Ser. B, 35 (1983), pp. 297–308.

[4] M. Chertkov and V. Y. Chernyak, *Loop series for discrete statistical models on graphs*, J. Stat. Mech. Theory Exp., (2006), P06009.

[5] M. Dyer, A. Frieze, and M. Jerrum, *On counting independent sets in sparse graphs*, SIAM J. Comput., 31 (2002), pp. 1527–1541.

[6] M. Dyer, A. Frieze, and R. Kannan, *A random polynomial-time algorithm for approximating the volume of convex bodies*, J. ACM, 38 (1991), pp. 1–17.

[7] M. Dyer and C. Greenhill, *On Markov chains for independent sets*, J. Algorithms, 35 (2000), pp. 17–49.

[8] H. Garmo, *The asymptotic distribution of long cycles in random regular graphs*, Random Structures Algorithms, 15 (1999), pp. 43–92.

[9] H.-O. Georgii, *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin, 1988.

[10] F. P. Kelly, *Stochastic models of computer communication systems*, J. R. Stat. Soc. Ser. B, 47 (1985), pp. 379–395.

[11] N. Madras and G. Slade, *The Self-Avoiding Walk. Probability and Its Applications*, Birkhäuser Boston, Boston, 1993.

[12] F. Martinelli, A. Sinclair, and D. Weitz, *Fast mixing for independent sets, colorings and other models on trees*, in Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, SIAM, Philadelphia, 2004, pp. 449–458.

[13] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA, 1988.

[14] A. Sinclair and M. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Inform. and Comput., 82 (1989), pp. 93–133.

[15] E. B. Sudderth, M. J. Wainwright, and A. S. Willsky, *Loop series and Bethe variational bounds in attractive graphical models*, in Advances in Neural Information Processing Systems 20, MIT Press, Cambridge, MA, 2008, pp. 1425–1432.

[16] E. Vigoda, *A note on the Glauber dynamics for sampling independent sets*, Electron. J. Combin., 8 (2001), research paper 8.

[17] M. Wainwright and M. Jordan, *Graphical models, exponential families, and variational inference*, Found. Trends Mach. Learn., 1 (2008), pp. 1–305.

[18] D. Weitz, *Counting independent sets up to the tree threshold*, in Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing, ACM Press, New York, 2006, pp. 140–149.

[19] N. C. Wormald, *Asymptotic distribution of short cycles in random regular graphs*, J. Combin. Theory Ser. B, 31 (1981), pp. 168–182.

[20] A. B. Yedidia, *Counting Independent Sets and Kernels of Regular Graphs*, manuscript, 2009; available online from http://arxiv.org/abs/0910.4664.

[21] J. Yedidia, W. Freeman, and Y. Weiss, *Constructing free energy approximations and generalized belief propagation algorithms*, IEEE Trans. Inform. Theory, 51 (2004), pp. 2282–2312.