

# Counting Independent Sets Using the Bethe Approximation

V. Chandrasekaran <sup>\*</sup>    M. Chertkov <sup>†</sup>    D. Gamarnik <sup>‡</sup>    D. Shah <sup>\*</sup>    J. Shin <sup>§</sup>

## 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 and iterative algorithm that is known to have at least one fixed point. Each fixed point corresponds to a stationary point of the Bethe free energy (introduced by Yedidia, Freeman and Weiss (2004) in recognition of Hans Bethe’s earlier work (1935)). The evaluation of the Bethe Free Energy at such a stationary point (or BP fixed point) leads to the Bethe approximation to the number of independent sets of the given graph. In general BP is not known to converge nor is an efficient, convergent procedure for finding stationary points of the Bethe free energy known. Further, effectiveness of Bethe approximation is not well understood.

As the first result of this paper, we propose a BP-like algorithm that always converges to a BP fixed point for *any* graph. Further, it finds an  $\varepsilon$  approximate fixed 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  $d$ . As the next step, we study the quality of this approximation. Using the recently developed ‘loop series’ approach by Chertkov and Chernyak, we establish that for any graph of  $n$  nodes with max-degree  $d$  and girth larger than  $8d \log_2 n$ , the multiplicative error 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 (2006).

Finally as a consequence of our results, we prove that the Bethe approximation is exceedingly good for a random 3-regular graph conditioned on the Shotest Cycle Cover Conjecture of Alon and Tarsi (1985) being true.

---

<sup>\*</sup>Laboratory for Information and Decision Systems, Dept. of EECS, MIT, Cambridge, MA, 02139, e-mail: {venkatc, devavrat}@mit.edu

<sup>†</sup>Center for Nonlinear Studies and Theoretical Division at Los Alamos National Laboratory and New Mexico Consortium, e-mail: chertkov@lanl.gov

<sup>‡</sup>Operations Research Center and Sloan School of Management, MIT, Cambridge, MA, 02139, e-mail: gamarnik@mit.edu

<sup>§</sup>Laboratory for Information and Decision Systems, Dept. of Mathematics, MIT, Cambridge, MA, 02139, e-mail: jinwoos@mit.edu

# 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 questions do arise in practice as well, e.g. for performance evaluation of a finite buffered radio network (cf. see Kelly [11]). Recently, the Belief Propagation (BP) algorithm has become the heuristic of choice in many similar applications where the interest is in computing, what physicists call, 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 to understand the effectiveness of the BP based approximation (also known as the Bethe approximation) for counting independent set in a given graph.

BP is a simple, iterative or message-passing algorithm. It is well understood that this iterative algorithm does have fixed points. Further they correspond to stationary points of what is known as the Bethe free energy – for more details on the Bethe approximation and its relation to BP fixed points, see Yedidia et al. [21]; also see book by Georgii [9]. However, there are two key problems. First, the BP algorithm is not known to converge for general graphs for counting the number of independent sets; there are known counter-examples for other problems (cf. see [18]). Second, given the BP fixed points, i.e. the Bethe approximation, it is not clear what the approximation error is. In this paper, we will address both of these challenges as our main result. 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 using sampling through Markov chain. In this approach, initiated by the works of Dyer, Frieze and Kannan [7] and Jerrum and Sinclair [15], one wishes to design a Markov chain that samples independent sets uniformly and has fast mixing property. Some of the notable results for independent set problems are by [13, 6, 17, 5]. In summary, these results show the following: (a) for any graphs with max-degree upto 4, there is exists an FPRAS using a fast mixing Markov chain, (b) there is no fast mixing Markov chain (based on local updates) for all graphs with degree larger or equal to 6, and (c) counting independent sets for all graphs with degree larger than 25 is hard.

The second approach introduced by Weitz [19] provides a deterministic FPTAS for any graph with max-degree upto 5. It is based on establishing *correlation decay* property for any tree with max-degree upto 5 and an intriguing equivalence relation between (an appropriate) distribution on graph with (an appropriate) distribution on its self-avoiding walk tree. We take note of work by Bandyopadhyay and Gamarnik [2]: it establishes that the BP based approximation is asymptotically correct for graphs with large girth and degree upto 5 (e.g. random 4-regular graph). Like [19], it also uses correlation decay property. On the flip side, it provides approximation upto logarithm and error in logarithm is  $o(n)$  for graph of size  $n$ .

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

## 1.2 Our Results

Motivated to obtain good approximation results for graphs with larger ( $> 5$ ) max-degree, but possibly with additional constraints such as large girth, we consider utilizing Belief Propagation/Bethe

approximation for counting the number of independent sets. As the main result, we provide a deterministic algorithm based using Bethe approximation 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  $d$ .

To establish this result, as the first step we propose a new, simple message-passing algorithm that can be thought of as a minor modification of the BP. We show that our algorithm *always* converges to a fixed-point of BP or equivalently a stationary point of Bethe free energy for any graph. The algorithm takes  $O(n^2 d^4 2^d \varepsilon^{-4} \log^3(n \varepsilon^{-1}))$  iterations to obtain an  $\varepsilon$  approximate fixed point for a graph of  $n$  nodes and max-degree  $d$  (see Theorem 2).

Thus computed BP fixed points readily leads to the Bethe approximation for the number of independent sets. We analyze the error in this approximation, using the recently developed framework of ‘loop series’ by Chertkov and Chernyak [4], which character this error as a summation of terms with each term associated to a ‘generalized loop’ of the graph. Though this approach provides an ‘explicit’ characterization of the error, it involves possibly super-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 explicitly obtain a handle over this summation. We do so by bounding this 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. This, along with the result of Bermond, Jackson and Jaeger [3], leads to the eventual result that the error in Bethe Approximation decays as  $O(n^{-\gamma})$  for some  $\gamma > 0$  for any graph of  $n$  nodes with max-degree  $d$  and girth larger than  $8d \log_2 n$ .

As an implication of this result for random 3-regular graphs with the result of Bermond et al. replaced by its stronger version, also known as the Shortest Cycle Cover Conjecture (SCCC) by Alon and Tarsi [1] suggests the following result: the difference between the logarithms of the number of independent sets and the Bethe Free Energy is  $O(1)$  with high probability. This is in sharp contrast with the result of Bandyopadhyay and Gamarnik [2] that suggest error to be  $o(n)$ , based on correlation decay and expected by Physicists. Thus, we have an intriguing situation – either SCCC is false or the Bethe approximation is terrific! A byproduct of the technique used to establish the result for random 3-regular graph is the following algorithmic implication: it suggests a systematic way to correct the error in Bethe Approximation and this could be of interest in its own right.

### 1.3 Organization

Section 2 introduces the BP algorithm, the relation of its fixed points with the Bethe Free Energy, the Bethe Approximation for the problem of computing the number of independent sets in a given graph and the loop series based error characterization in this approximation. In Section 3, we describe a new message-passing algorithm for computing a fixed point of BP for this problem. We obtain its rate of convergence in Theorem 2. In Section 4, we analyze the error in the Bethe approximation based on thus computed BP fixed point for graphs with large girth. Finally, in Section 5 we obtain an unexpectedly sharp correctness of Bethe Approximation for random 3-regular graph assuming the Shortest Cycle Cover Conjecture.

## 2 Background

Let  $G = (V, E)$  be a graph with the vertices  $V = \{1, \dots, n\}$ , edges  $E \subseteq \binom{V}{2}$  and a (vertex labeled) collection of binary random variables  $\mathbf{X} = \{X_v | v \in V\}$ . Let  $\mathbf{X}_A = \{X_v | v \in A\}$  for any  $A \subset V$ .

The joint distribution of  $X$  is defined as follows: for  $\boldsymbol{\sigma} = (\sigma_v) \in \{0, 1\}^n$ ,

$$\Pr(\mathbf{X} = \boldsymbol{\sigma}) = \frac{1}{Z} \prod_{(u,v) \in E} (1 - \sigma_u \sigma_v), \quad (1)$$

where  $Z$  is the normalization constant. By definition, the distribution of  $\mathbf{X}$  is uniform over all independent sets of  $G$  and hence  $Z$  is the number of independent sets. We will use the following notations:  $\mathcal{N}(v)$  be 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 Belief Propagation (BP)

The BP algorithm (cf. [14]) is a heuristic to evaluate  $Z$  for any graphical model based distribution, such as (1). Next, we describe this algorithm for the problem of our interest. BP is iterative with parameters as messages –  $m_{u \rightarrow v}^t(\sigma), m_{v \rightarrow u}^t(\sigma)$  for all  $(u, v) \in E, \sigma \in \{0, 1\}$  for iteration time  $t$ . Initial,  $t = 0$  and  $m_{u \rightarrow v}^0(\sigma) = m_{v \rightarrow u}^0(\sigma) = 1$  for all  $(u, v) \in E, \sigma \in \{0, 1\}$ . For  $t \geq 0$ , the messages are updated as follows:

$$\frac{m_{u \rightarrow v}^{t+1}(1)}{m_{u \rightarrow v}^{t+1}(0)} = \frac{1}{1 + \prod_{w \in \mathcal{N}(u) \setminus v} \frac{m_{w \rightarrow u}^t(1)}{m_{w \rightarrow u}^t(0)}}. \quad (2)$$

The initial conditions and (2) suggest that the updating function corresponding to the ratios of messages (i.e.  $\frac{m_{w \rightarrow u}^t(1)}{m_{w \rightarrow u}^t(0)}$ ) is continuous. Further, it maps elements from the closed, convex set  $[0, 1]^{2|E|}$  to itself. Therefore, by Brouwer's fixed point theorem [10] at least one fixed point does exist. Any such fixed point, say  $(m_{u \rightarrow v}(\cdot))_{(u,v) \in E}$ , by definition must satisfy

$$\frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} = \frac{1}{1 + \prod_{w \in \mathcal{N}(u) \setminus v} \frac{m_{w \rightarrow u}(1)}{m_{w \rightarrow u}(0)}}. \quad (3)$$

Given a fixed point, the BP estimate of marginal distributions for  $X_v, v \in V$ , denoted by  $\tau_v(\sigma), \sigma \in \{0, 1\}$ , are defined as

$$\frac{\tau_v(1)}{\tau_v(0)} = \prod_{u \in \mathcal{N}(v)} \frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)}, \quad \text{and} \quad \tau_v(1) = 1 - \tau_v(0). \quad (4)$$

It also follows from (1), that for any  $u, v \in V$  with  $(u, v) \in E$ ,

$$\tau_{u,v}(0, 1) = \tau_v(1), \quad \tau_{u,v}(1, 0) = \tau_u(1), \quad \tau_{u,v}(1, 1) = 0, \quad \tau_{u,v}(0, 0) = 1 - \tau_v(1) - \tau_u(1), \quad (5)$$

$$\tau_v(1) \leq \tau_v(0), \quad \max\{\tau_{u,v}(1, 0), \tau_{u,v}(0, 1)\} \leq \tau_{u,v}(0, 0). \quad (6)$$

In above,  $\tau_{u,v}(\sigma, \sigma')$  is a BP based estimate for pair-wise joint distribution of  $(X_u, X_v)$ . For details on relation between BP estimates and node/pair-wise marginals, please refer to the monograph by Wainwright and Jordan [18].

## 2.2 Bethe Approximation

Now we present the Bethe approximation for  $Z$  based on the BP fixed point or equivalently the induced node/pair-wise marginals as per (4) and (5). To this end, the Bethe free energy (cf. see [21]) for problem of interest is a function  $F : [0, 1]^{|V|} \rightarrow \mathbb{R}$ , defined as follows: for  $\mathbf{x} = (x_v) \in [0, 1]^{|V|}$ ,

$$F(\mathbf{x}) = \sum_{v \in V} (x_v \ln x_v - (d(v) - 1)(1 - x_v) \ln(1 - x_v)) + \sum_{(u,v) \in E} (1 - x_u - x_v) \ln(1 - x_u - x_v) \quad (7)$$

**Definition 1.** (*Bethe Approximation*) Let  $\boldsymbol{\tau} = (\tau_v(1))_{v \in V} \in [0, 1]^{|V|}$  be BP fixed point based node marginals (cf. (4)). Then, the Bethe approximation for  $\ln Z$ , the logarithm of number of independent sets, denoted by  $\ln Z_B$  is defined as

$$\ln Z_B = \ln Z_B(\boldsymbol{\tau}) \triangleq F(\boldsymbol{\tau}).$$

### 2.3 BP Fixed Pt. = Stationary Pt. of $F$

Now we state the relation between BP fixed point and the stationary point of  $F$ . Given this, it follows that an alternative way to obtain Bethe approximation for  $\ln Z$  (or  $Z$ ) is to first find the stationary point of  $F$  and then evaluate  $F$  at that stationary point (instead of using BP to find the fixed point and evaluate  $F$  at that fixed point). To this end, note that the gradient of  $F$ ,  $\nabla F(\mathbf{x}) = \left[ \frac{\partial F}{\partial x_v} \right]$  is such that

$$\frac{\partial F}{\partial x_v} = (d(v) - 1) \ln(1 - x_v(t)) + \ln x_v(t) - \sum_{u \in \mathcal{N}(v)} \ln(1 - x_u(t) - x_v(t)). \quad (8)$$

Let  $\mathbf{x}^*$  be a zero gradient point (or stationary point) of  $F$ , i.e.  $\nabla F(\mathbf{x}^*) = \mathbf{0}$ . From (8), it follows that

$$\frac{\prod_{u \in \mathcal{N}(v)} (1 - x_v^* - x_u^*)}{(1 - x_v^*)^{d(v)-1} x_v^*} = 1. \quad (9)$$

We state the following result that relates a zero gradient point with the fixed point of BP messages as well as corresponding marginal estimates. For completeness, we include its proof in the Appendix A (also see Yedidia et al. [21]).

**Lemma 1.** *Given a zero gradient point  $\mathbf{x}^* = (x_v^*)_{v \in V}$  of  $F$  as defined above, for any  $(u, v) \in E$  define*

$$\frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} = \frac{1 - x_v^* - x_u^*}{1 - x_v^*}.$$

*Then such messages are BP fixed point, i.e. they satisfy (3). Further, the corresponding node marginals*

$$\frac{\tau_v(1)}{\tau_v(0)} = \prod_{u \in \mathcal{N}(v)} \frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} = \frac{x_v^*}{1 - x_v^*}.$$

### 2.4 Error in Bethe Approximation: Loop Series Correction

Recently, Chertkov and Chernyak [4] showed that the  $Z$  can be obtained by “correcting” the Bethe approximation  $Z_B$  as follows:

$$Z = Z_B \left( 1 + \sum_{\emptyset \neq F \subset E} w(F) \right). \quad (10)$$

Here  $F \subset E$  are (edge) subgraphs of  $G$  with weight  $w(F)$ , defined next (cf. see [4, 16]). For  $F$  with any node having degree 1,  $w(F) = 0$ . For all other  $F$ , called *generalized loops*,

$$w(F) = (-1)^{|F|} \prod_{v \in V_F} \tau_v(1) \left[ 1 + (-1)^{d_F(v)} \left( \frac{\tau_v(1)}{\tau_v(0)} \right)^{d_F(v)-1} \right]. \quad (11)$$

### 3 Fast, Convergent Algorithm for Bethe Approximation

In order to compute the Bethe approximation to the number of independent sets, one could obtain a stationary point of the Bethe free energy using BP. However, since BP does not always converge in general (e.g. see [18]), we propose a convergent BP-like alternative in this section to compute the Bethe approximation.

#### 3.1 Algorithm Description

The algorithm described next computes  $\mathbf{x}(t) = (x_v(t))_{v \in V}$  as an approximation of a zero gradient point of  $F$  (cf. (9)) (and hence BP fixed point as per Lemma 1). It is based on the standard gradient algorithm. The non-triviality lies in the choice of the appropriate ‘step-size,’ and subsequent analysis of correctness and rate of convergence.

- Algorithm parameters: iteration  $t \geq 0$ ,  $\mathbf{x}(t) = (x_v(t))_{v \in V}$ . Initially,  $t = 0$  and  $x_v(0) = 1/4$ ,  $v \in V$ .
- $\mathbf{x}(t) = (x_v(t))_{v \in V}$  is updated till  $t \leq T$ :

$$x_v(t+1) = x_v(t) - \alpha(t) \left. \frac{\partial F}{\partial x_v} \right|_{\mathbf{x}(t)},$$

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

$$\left. \frac{\partial F}{\partial x_v} \right|_{\mathbf{x}(t)} = \left( (d(v) - 1) \ln(1 - x_v(t)) + \ln x_v(t) - \sum_{u \in \mathcal{N}(v)} \ln(1 - x_u(t) - x_v(t)) \right).$$

- Choose a  $s \leq T$  with probability  $\frac{\alpha(s)}{\sum_{t \leq T} \alpha(t)}$ , output  $\mathbf{x}(s) = (x_v(s))_{v \in V}$  as an approximate zero gradient of  $F$  and approximate BP fixed point messages, for any  $(u, v) \in E$  as  $\frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} = \frac{1 - x_v(s) - x_u(s)}{1 - x_v(s)}$ .

#### 3.2 Properties of Algorithm: Correctness, Convergence

Consider the following definition of  $\varepsilon$ -approximation of BP fixed point.

**Definition 2.** Given  $\varepsilon \in (0, 1)$ , messages  $\{m_{u \rightarrow v}(1), m_{u \rightarrow v}(0)\}_{(u,v) \in E}$  is called the BP  $\varepsilon$ -fixed (or  $\varepsilon$ -BP fixed) point messages if for all  $(u, v) \in E$ ,

$$\left| \frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} \left( 1 + \prod_{w \in \mathcal{N}(u) \setminus v} \frac{m_{w \rightarrow u}(1)}{m_{w \rightarrow u}(0)} \right) - 1 \right| \leq \varepsilon.$$

Next we state the result about correctness, convergence of the algorithm. Its proof is in Appendix B.

**Theorem 2.** Let  $\hat{\mathbf{x}} = (\hat{x}_v)_{v \in V}$ ,  $\hat{m}_{u \rightarrow v}(1)/\hat{m}_{u \rightarrow v}(0)$ ,  $(u, v) \in E$  be output of the algorithm. Then,

$$\mathbb{E} \left[ \max_{(u,v) \in E} \left| \frac{\hat{m}_{u \rightarrow v}(1)}{\hat{m}_{u \rightarrow v}(0)} \left( 1 + \prod_{w \in \mathcal{N}(u) \setminus v} \frac{\hat{m}_{w \rightarrow u}(1)}{\hat{m}_{w \rightarrow u}(0)} \right) - 1 \right|^2 \right] = O \left( \frac{n d^4 2^d \log T}{\sqrt{T}} \right). \quad (12)$$

*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  $\varepsilon$ -BP fixed point for any  $\varepsilon > 0$ .

## 4 Correctness of $Z_B$ for Graphs With Large Girth

The algorithm in previous section provides BP fixed point or stationary point of Bethe Free Energy. This leads to the Bethe Approximation,  $\ln Z_B$  (or  $Z_B$ ), of  $\ln Z$  (or  $Z$ ) for any graph  $G$ . Here, we establish that the error in  $Z_B$  in estimating  $Z$  asymptotically decays to 0 for graphs with large girth. Formally, girth of the graph is the length of the shortest cycle (i.e. for tree, it is  $\infty$ ). The formal result is stated below.

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

$$\left| \frac{Z}{Z_B} - 1 \right| = O(n^{-\gamma}),$$

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

### 4.1 Proof Sketch: Theorem 3

We start by introducing a useful notion of *apples* – a special class of connected subgraphs of  $G$ .

**Definition 3.** (*Apple*) A connected edge subgraph  $C \subset E$  of  $G$  is an apple if (a) it is a cycle, or (b) it is 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 an apple  $C \subset E$  and BP fixed point based node marginals  $\tau_v(1)/\tau_v(0), v \in V$ , define its weight as

$$\hat{w}(C) \triangleq \left( \prod_{\{u,v\} \in C} \sqrt{\frac{\tau_u(1) \tau_v(1)}{\tau_u(0) \tau_v(0)}} \right)^{\frac{1}{2d}}. \quad (13)$$

As the first result, we will establish the following bound on summation of weights over all apples. The proof is presented in Section 4.2.

**Lemma 4.** Let  $g = g(G)$  be girth of  $G$  so that  $g(G) > 8d \log_2 n$ . Then

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

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

To establish Theorem 3, it is sufficient to show that

$$\sum_{\emptyset \neq F \subset E} |w(F)| = O(n^{-\gamma}). \quad (14)$$

For this, we first bound the term  $\sum_{\emptyset \neq F \subset E} |w(F)|$  by summation  $\sum_{C \subset E} \hat{w}(C)$  as follows. The proof is presented in Section 4.3.

**Lemma 5.** For any graph  $G$ ,

$$1 + \sum_{\emptyset \neq F \subset E} |w(F)| \leq e^{\sum_{C \subset E} \widehat{w}(C)}.$$

Now from Lemma 5 and 4 as well as  $e^x = 1 + O(x)$  for  $x = O(n^{-\gamma})$  with  $\gamma > 0$ , the desired bound (14) follows immediately. This completes the proof of Theorem 3.

## 4.2 Proof of Lemma 4

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

To this end, first we bound (a), i.e. the number of apples of a given size, say  $k$ . We state the following proposition (see Appendix C for its proof).

**Proposition 6.** For any apple  $C$  of size  $k$ ,  $\widehat{w}(C) < 2^{-\frac{k}{2d}}$ .

Next, we bound (b), 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.* Given an apple  $C$ , if  $C$  has a degree 1 vertex, say  $v$ , then define it as its starting vertex or else (i.e.  $C$  a cycle) let it be arbitrary. Now consider  $T_v(G)$ , the self avoiding walk tree (cf. [12]) 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 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 upto 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 upto  $g/2$  that are copies of the same vertex and hence leading to existence of a cycle of length  $< g$  in  $G$ . For the very same reason, it also follows that any sub-tree of  $T_v(G)$  must have at most  $n$  nodes upto (its) level  $g/2$ . Using these properties, it can be shown that the number of vertices (and hence leaves) upto level  $k$  of  $T_v(G)$  is at most

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

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

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

This completes the proof of Proposition 7. □

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

$$\begin{aligned} \sum_{C \subset E} \widehat{w}(C) &\leq \sum_{k \geq g} n^2 \left( e^{2/c_1} \right)^k 2^{-\frac{k}{2d}} = n^2 \left( \sum_{k \geq g} \delta^k \right) \\ &< n^2 \left( \frac{\delta^g}{1 - \delta} \right) = O \left( n^2 n^{c_1 \ln \delta} \right) = O(n^{-\gamma}). \end{aligned}$$

In above, we have used  $c_1 = g/\ln n > \frac{8d}{\ln 2}$  and definition  $\delta \triangleq 2^{-\frac{1}{2d}} e^{2/c_1} = e^{-\frac{1}{2d} \ln 2 + 2/c_1} < 1$ .

### 4.3 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 define a list of apples  $\{C_i\}$  as a *good decomposition* of a given generalized loop  $F$  if it satisfies the following conditions:

$$F = \cup C_i \quad \text{and} \quad |w(F)| \leq \prod_i \widehat{w}(C_i).$$

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

$$1 + \sum_{\emptyset \neq F \subseteq E} |w(F)| \leq \prod_{C \subseteq E} (1 + \widehat{w}(C)) < \prod_{C \subseteq E} e^{\widehat{w}(C)} = e^{\sum_{C \subseteq E} \widehat{w}(C)}, \quad (15)$$

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

Now some notations. Given a list of apples  $\{C_i\}$  and  $F = \cup C_i$ , for  $(u, v) \in F$  let  $N_{(u,v)}$  be the number of  $C_i$  that include the  $(u, v)$ ; let  $N_{\max} = \max_{(u,v) \in F} N_{(u,v)}$ . Then, we state the following that uses Theorem 8.

**Proposition 9.** *For any generalized loop  $F$ , there exists a list  $\{C_i\}$  of apples with  $N_{\max} \leq 4$ .*

*Proof.* Assume  $F$  is connected, or else apply the argument to each connected component separately. Now if  $F$  has no bridge, by the Theorem 8, there exists a list  $\{C_i\}$  of cycle (and hence apples) which cover every edge exactly 4 times, hence  $N_{\max} = 4$ . Now suppose  $F$  has a bridge. Then, break  $F$  into two components of size  $> 1$  which are connected by a bridge. Recursively, break thus generated components via a bridge till each component is left without a bridge. Since each of these eventually resulting bridgeless components are of size  $> 1$ , they are essentially vertices of a meta-tree, say  $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ , whose edges are essentially line subgraphs of  $F$ . Now choose a component of this meta tree arbitrarily as its root. By another application of Theorem 8, it follows that for each component  $v \in \mathcal{V}$  (i.e. a meta-vertex of  $\mathcal{T}$ ), there is a list  $L_v$  of cycles that covers all of  $v$ 's edges exactly four times. Now for each line subgraph or meta-edge  $e \in \mathcal{E}$  of  $\mathcal{T}$ , let  $v(e)$  be the ‘child’ node as per the rooting of  $\mathcal{T}$ . Then, attach this line subgraph or meta-edge  $e$  to an appropriate cycle in the list  $L_{v(e)}$ . With abuse of notation, call thus modified list of cycles, now some of them apples, as  $L_{v(e)}$ . Note that this will lead to a valid list of apples,  $\cup_{v \in \mathcal{V}} L_v$  that will cover every edge of  $G$  at least once and at most 4 times. This completes the proof of Proposition 9.  $\square$

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

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

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:

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

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

## 5 Correctness of $Z_B$ : Random 3-Regular Graphs

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

**Conjecture 10** (Shortest Cycle Cover Conjecture). *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  $7m/5 = 1.4m$ .*

Now we state our result that implies 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 SCCC is true, there exists a function  $f : (0, 1) \rightarrow \mathbb{R}^+$ , so that*

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

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

### 5.1 Proof Sketch: Theorem 11

From (10), it is equivalent to show that

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

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

$$\ln \left( 1 + \sum_{\emptyset \neq F \subseteq E} w(F) \right) \leq \ln \left( 1 + \sum_{\emptyset \neq F \subseteq E} |w(F)| \right) \leq \ln(1 + g(\varepsilon)). \quad (16)$$

The details are explained in Section 5.2.

Now, if we have  $g(\varepsilon)$  uniformly bounded below 1, say always at most 1/2 for example, then the (16) would be sufficient to establish the claim of Theorem 11. However, that may not be true.

For this reason, we need additional proof-techniques to obtain an appropriate lower bound on the quantity of interest. This lower bounding technique needs longer explanation and hence presented in Appendix E. A careful reader may notice that our lower bounding technique is essentially an algorithm that tries to ‘correct’ the error in Bethe approximation by means of the loop series characterization in a systematic manner.

## 5.2 Upper Bound

As we explain in Section 5.1, we show that  $\sum_{\emptyset \neq F \subseteq E} |w(F)|$  is less than  $g(\varepsilon)$  with probability  $1 - \varepsilon$ . To this end, it is enough to prove that

$$\mathbb{E} \left[ \ln \left( 1 + \sum_{\emptyset \neq F \subseteq E} |w(F)| \right) \right] = O(1). \quad (17)$$

If (17) holds, we can choose  $g(\varepsilon) = e^{O(1/\varepsilon)} - 1$  by Markov inequality. If  $G$  is a 3-regular graph, we can find the explicit homogeneous (and unique due to correlation decay cf. Weitz [19]) fixed point of BP. From (3) and setting  $\frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} = z$ ,  $\forall (u, v) \in E$ , we obtain

$$z = \frac{1}{1 + z^2},$$

where such a  $z$  can be found numerically to be  $z \approx 0.682$ . Thus,  $\frac{\tau_v(1)}{\tau_v(0)} = z^3 \approx 0.317$  from (3), hence  $\tau_v(0) \approx 0.759$  and  $\tau_v(1) \approx 0.241$ . Furthermore, the corresponding  $Z_B$  can be calculated as  $\ln Z_B = \frac{1}{2}n \ln \left( \frac{1}{z^3(2-z)} \right) \approx n \ln 1.545$ .

**Lemma 12.** *If  $G$  is a 3-regular graph and SCCC is true,*

$$\ln \left( 1 + \sum_{\emptyset \neq F \subseteq E} |w(F)| \right) \leq \sum_{C \subseteq E} \tilde{w}(C),$$

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

From Lemma 12 (see Appendix D for its proof), to establish (17), we need

$$\mathbb{E} \left[ \sum_{C \subseteq E} \tilde{w}(C) \right] = O(1). \quad (18)$$

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

$$A_k \leq \sum_{i \leq k} C_i \times i \times 2^{k-i}, \quad (19)$$

since apples can be made only by attaching a line to a cycle. It is well-known [20, 8] that the expected value of  $C_k$  for random 3-regular graphs is at most  $2^{k-1}/k$ . Using this fact and (19), it follows that the expected value of  $A_k$  for random 3-regular graphs is at most  $k 2^{k-1}$ . Therefore, the desired bound (18) can be obtained as:

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

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

## References

- [1] N. Alon and M. Tarsi. Covering multigraphs by simple circuits. *SIAM Journal on Algebraic and Discrete Methods*, 6:345, 1985.
- [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 algorithm*, pages 890–899. ACM New York, NY, USA, 2006.
- [3] JC Bermond, B. Jackson, and F. Jaeger. Shortest coverings of graphs with cycles. *Journal of combinatorial theory. Series B*, 35(3):297–308, 1983.
- [4] M. Chertkov and V.Y. Chernyak. Loop series for discrete statistical models on graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 6:P06009, 2006.
- [5] M. Dyer, A. Frieze, and M. Jerrum. On counting independent sets in sparse graphs. In *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pages 210–217, 1999.
- [6] M. Dyer and C. Greenhill. On Markov chains for independent sets. *Journal of Algorithms*, 35(1):17–49, 2000.
- [7] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, 1991.
- [8] H. Garmo. The asymptotic distribution of long cycles in random regular graphs. *Random Struct. Algorithms*, 15(1):43–92, 1999.
- [9] Hans-Otto Georgii. *Gibbs measures and phase transitions*. Walter de Gruyter, 1988.
- [10] S. Kakutani. A generalization of Brouwers Fixed Point Theorem . *Duke Math*, 8:457 – 459, 1941.
- [11] F. P. Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society (Series B)*, 47:379–395, 1985.
- [12] N. Madras and G. Slade. *The self-avoiding walk*. Birkhauser, 1996.
- [13] 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*, pages 456–465. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2004.
- [14] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.
- [15] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, 1989.
- [16] E.B. Sudderth, M.J. Wainwright, and A.S. Willsky. Loop series and Bethe variational bounds in attractive graphical models. *Advances in neural information processing systems*, 20:1425–1432, 2008.
- [17] E. Vigoda. A note on the Glauber dynamics for sampling independent sets. *Electronic Journal of Combinatorics*, 8(1), 2001.

- [18] M. Wainwright and M. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1:1–305, 2008.
- [19] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 140–149, New York, NY, USA, 2006. ACM Press.
- [20] NC Wormald. Asymptotic Distribution of Short Cycles in Random Regular Graphs. *J. COMBINAT. THEORY, SER. B.*, 31(2):168–182, 1981.
- [21] J. Yedidia, W. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51:2282–2312, 2004.

## A Proof of Lemma 1

From (9), the zero gradient point  $\mathbf{x}^* = (x_v^*)_{v \in V}$  of  $F$  satisfies

$$(1 - x_v^*)^{d(v)-1} x_v^* = \prod_{u \in \mathcal{N}(v)} (1 - x_u^* - x_v^*), \quad \text{for all } v \in V. \quad (20)$$

Therefore,

$$\begin{aligned} \frac{1 - x_v^* - x_u^*}{1 - x_v^*} &= \frac{1}{1 + \frac{x_u^*}{1 - x_v^* - x_u^*}} \\ &= \frac{1}{1 + \prod_{w \in \mathcal{N}(u) \setminus v} \frac{1 - x_w^* - x_u^*}{1 - x_u^*}}. \end{aligned} \quad (21)$$

Now as per definition in Lemma 1,

$$\frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} = \frac{1 - x_v^* - x_u^*}{1 - x_u^*}.$$

Then the (21) is equivalent to the fixed point equation (3). The consequence  $\tau_v(1)/\tau_v(0) = x_v^*/(1 - x_v^*)$  follows immediately. This completes the proof of Lemma 1.

## B Proof of Theorem 2

Recall that  $F : [0, \frac{1}{2}]^n \rightarrow \mathbb{R}$  is such that

$$F(\mathbf{x}) = \sum_{v \in V} (x_v \ln x_v - (d(v) - 1)(1 - x_v) \ln(1 - x_v)) + \sum_{(u,v) \in E} (1 - x_u - x_v) \ln(1 - x_u - x_v).$$

Now the updating rule of the algorithm is equal to

$$x_v(t+1) = x_v(t) - \alpha(t) \left. \frac{\partial F}{\partial x_v} \right|_{\mathbf{x}(t)}.$$

If  $\|\nabla F(\mathbf{x}(s))\|_2 \leq \varepsilon < 1$ , it can be checked (similar to Lemma 1) that the corresponding output set by  $\frac{m_{u \rightarrow v}(1)}{m_{u \rightarrow v}(0)} = \frac{1 - x_v(s) - x_u(s)}{1 - x_v(s)}$  is a BP  $O(\varepsilon)$ -fixed point. Therefore, it suffices to show that

$$\mathbb{E}[\|\nabla F(\mathbf{x}(s))\|_2^2] = O\left(\frac{n d^4 2^d \log T}{\sqrt{T}}\right). \quad (22)$$

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

$$\frac{\partial \ln F}{\partial x_v} \leq 0 \quad \text{if } x_v < 2\varepsilon_1 \quad \text{and } \mathbf{x} \in D \triangleq \left[ \varepsilon_1, \frac{1}{2} - \varepsilon_2 \right]^n, \quad (23)$$

$$\frac{\partial \ln F}{\partial x_v} \geq 0 \quad \text{if } x_v > \frac{1}{2} - 2\varepsilon_2 \quad \text{and } \mathbf{x} \in D, \quad (24)$$

$$\left| \alpha \frac{\partial F}{\partial x_v} \right| \leq \frac{1}{2} \min\{\varepsilon_1, \varepsilon_2\} \quad \text{if } \mathbf{x} \in D \quad \text{and } \alpha \leq \frac{1}{2^{d+7}(d^2 + 6d + 2)}. \quad (25)$$

From (23), (24) and (25) it follows that  $\mathbf{x}(t) \in D$  i.e.  $x_v(t)$  does not hit 0 or  $\frac{1}{2}$  and hence the claim that  $x_v(\cdot) \in [0, 1/2]$  under the algorithm's iterations.

*Proof of (23).* Observe that

$$\begin{aligned} \frac{\partial \ln F}{\partial x_v} &= (d(v) - 1) \ln(1 - x_v) + \ln x_v - \sum_{u \in \mathcal{N}(v)} \ln(1 - x_u - x_v) \\ &\leq \ln x_v - d \ln \left( \frac{1}{2} - x_v \right) = \ln \frac{x_v}{\left( \frac{1}{2} - x_v \right)^d} \\ &= \ln \frac{2^d x_v}{(1 - 2x_v)^d} \leq \ln \frac{2^d x_v}{1 - 2d x_v} \leq 0, \end{aligned}$$

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

*Proof of (24).* Consider the following:

$$\begin{aligned} \frac{\partial \ln F}{\partial x_v} &= (d(v) - 1) \ln(1 - x_v) + \ln x_v - \sum_{u \in \mathcal{N}(v)} \ln(1 - x_u - x_v) \\ &\geq (d(v) - 1) \ln(1 - x_v) + \ln x_v - \sum_{u \in \mathcal{N}(v)} \ln(1 - \varepsilon_1 - x_v) \\ &= \ln \frac{x_v}{1 - x_v} - \sum_{u \in \mathcal{N}(v)} \ln \frac{1 - \varepsilon_1 - x_v}{1 - x_v} \\ &\geq \ln \frac{x_v}{1 - x_v} - \ln \frac{1 - \varepsilon_1 - x_v}{1 - x_v} = \ln \frac{x_v}{1 - \varepsilon_1 - x_v} \\ &> \ln \frac{\frac{1}{2} - 2\varepsilon_2}{\frac{1}{2} + 2\varepsilon_2 - \varepsilon_1} \geq 0, \end{aligned}$$

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

*Proof of (25).* This follows from our choice of  $\alpha$ , since for  $\mathbf{x} \in D$

$$\begin{aligned} \left| \frac{\partial \ln F}{\partial x_v} \right| &= \left| (d(v) - 1) \ln(1 - x_v) + \ln x_v - \sum_{u \in \mathcal{N}(v)} \ln(1 - x_u - x_v) \right| \\ &\leq \left| \ln \frac{x_v}{1 - x_v} - \sum_{u \in \mathcal{N}(v)} \ln \frac{1 - x_u - x_v}{1 - x_v} \right| \\ &\leq -\ln \frac{x_v}{1 - x_v} - \sum_{u \in \mathcal{N}(v)} \ln \frac{1 - x_u - x_v}{1 - x_v} \\ &\leq -\ln \frac{\varepsilon_1}{1 - \varepsilon_1} - d \ln \frac{2\varepsilon_2}{\frac{1}{2} + \varepsilon_2} \\ &\leq -\ln \varepsilon_1 - d \ln 2\varepsilon_2 \\ &= \ln 2(d + 2 + d(d + 5)) \leq d^2 + 6d + 2, \end{aligned} \tag{26}$$

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

Now that we have established  $\mathbf{x}(t) \in D$  as a consequence of the above three steps, we consider the dynamics

$$\mathbf{x}(t + 1) = \mathbf{x}(t) - \alpha(t) \nabla F(\mathbf{x}(t)).$$

From the Taylor's expansion

$$\begin{aligned} F(\mathbf{x}(t+1)) &= F(\mathbf{x}(t) - \alpha(t) \nabla F(\mathbf{x}(t))) \\ &= F(\mathbf{x}(t)) - \nabla F(\mathbf{x}(t)) \cdot \alpha(t) \nabla F(\mathbf{x}(t)) + \frac{1}{2} \alpha(t) \nabla F(\mathbf{x}(t)) \cdot R \cdot \alpha(t) \nabla F(\mathbf{x}(t)), \end{aligned} \quad (27)$$

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

$$|R_{vw}| \leq \sup_{\mathbf{x} \in B} \left| \frac{\partial^2 F}{\partial x_v \partial x_w} \right|,$$

and  $B$  is a  $L_\infty$ -ball in  $\mathbb{R}^n$  centered at  $\mathbf{x}(t) \in D$  with its radius  $r = \max_{v \in V} \left| \alpha(t) \frac{\partial \ln F}{\partial x_v}(\mathbf{x}(t)) \right|$ . From (25), we know  $r \leq \frac{1}{2} \min\{\varepsilon_1, \varepsilon_2\}$ . Hence, if  $\mathbf{x} \in B$ ,  $\mathbf{x} \in [\varepsilon_1/2, \frac{1}{2} - \varepsilon_2/2]^n$ . Using this, we can get a bound for  $\sup_{\mathbf{x} \in B} \left| \frac{\partial^2 F}{\partial x_v \partial x_w} \right|$  as follows:

◦ If  $u = w$ ,

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

◦ If  $w \in \mathcal{N}(v)$ ,

$$\left| \frac{\partial^2 F}{\partial x_v \partial x_w} \right| = \frac{1}{1-x_w-x_v} \leq \frac{1}{\varepsilon_2} = O(2^d).$$

◦ Otherwise,  $\frac{\partial^2 F}{\partial x_v \partial x_w} = 0$ .

Therefore, using these bounds with (26), the equality (27) becomes

$$\begin{aligned} F(\mathbf{x}(t+1)) &\leq F(\mathbf{x}(t)) - \alpha(t) \|\nabla F(\mathbf{x}(t))\|_2^2 + \alpha^2(t) O(|E| d^5 2^d) \\ &= F(\mathbf{x}(t)) - \alpha(t) \|\nabla F(\mathbf{x}(t))\|_2^2 + \alpha^2(t) O(n d^6 2^d). \end{aligned} \quad (28)$$

If we sum (28) over  $t$  from 0 to  $T-1$ , we have

$$F(\mathbf{x}(T)) \leq F(\mathbf{x}(0)) - \sum_{t=0}^{T-1} \alpha(t) \|\nabla F(\mathbf{x}(t))\|_2^2 + O(n d^6 2^d) \sum_{t=0}^{T-1} \alpha^2(t). \quad (29)$$

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

$$\sum_{t=0}^{T-1} \alpha(t) \|\nabla F(\mathbf{x}(t))\|_2^2 \leq O(nd) + O(n d^6 2^d) \sum_{t=0}^{T-1} \alpha^2(t). \quad (30)$$

Thus, we finally obtain the desired conclusion for (22):

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

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

## C Proof of Proposition 6

From definition of  $\hat{w}$  in (13), it is enough to show that  $\sqrt{\frac{\tau_u(1) \tau_v(1)}{\tau_u(0) \tau_v(0)}} \leq \frac{1}{2}$  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 properties noted in (4)-(6). Thus, we have

$$\tau_v(1) + \tau_u(1) \leq \frac{2}{3}, \tag{31}$$

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:

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

In above, (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 (31) and the monotonicity of  $f(x) = \frac{x}{1-x}$ .

## D Proof of Lemma 12

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

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

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

$$F = \cup C_i \quad \text{and} \quad \sum_i |C_i| \leq 1.4 \times |F|. \quad (32)$$

Then

$$\begin{aligned} \prod_i \tilde{w}(C_i) &= \prod_i \alpha^{|C_i|} \\ &= \alpha^{\sum_i |C_i|}. \end{aligned} \quad (33)$$

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

$$\begin{aligned} |w(F)| &= \prod_{v \in V_F} \tau_v(1) \left[ 1 + (-1)^{d_F(v)} \left( \frac{\tau_v(1)}{\tau_v(0)} \right)^{d_F(v)-1} \right] \\ &\stackrel{(a)}{\leq} \prod_{v \in V_F} (z^3(1-z^3))^{\frac{d_F(v)}{3}} \\ &= \prod_{v \in V_F} \alpha^{\frac{3 \times 1.4}{2} \times \frac{d_F(v)}{3}} \\ &= \alpha^{\sum_{v \in V_F} \frac{1.4 \times d_F(v)}{2}} \\ &= \alpha^{1.4 \times |F|}, \end{aligned} \quad (34)$$

where the inequality (a) can be established in each of the possible cases  $d_F(v) = 0, 1, 2, 3$  using the explicit value of  $z = 0.682\dots$ ,  $\tau_v(0) \approx 0.759$  and  $\tau_v(1) \approx 0.241$ . Further, the inequality (a) is tight only when  $d_F(v) = 3$ . Therefore, from (32), (33) and (34) (and  $\alpha < 1$ ) we have

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

## E Proof of Theorem 11: Lower Bound

Using (19) it follows that

$$\begin{aligned} \sum_{CCE} \tilde{w}(C) &= \sum_k A_k \alpha^k \\ &\leq \sum_k \sum_{i \leq k} C_i \times i \times 2^{k-i} \alpha^k \\ &= \sum_i \sum_{k \geq i} C_i \times i \times 2^{k-i} \alpha^k \\ &= \sum_i C_i \frac{i}{2^i} \sum_{k \geq i} (2\alpha)^k \\ &\leq \sum_i C_i \frac{i}{2^i} \frac{(2\alpha)^i}{1-2\alpha} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{1-2\alpha} \sum_i C_i i(\alpha)^i \\
&\leq c_\alpha \times \sum_i C_i (0.49)^i \\
&\triangleq \rho(G), \tag{35}
\end{aligned}$$

where  $c_\alpha$  is a constant which depends on  $\alpha$  and the last inequality is due to  $\alpha \approx 0.48$ . We state the following lemma which is a key for the proof of the lower bound.

**Lemma 13.** *Given a random 3-regular graph  $G$  of  $n$  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  $\varepsilon$  dependent constant, independent of  $n$ .

The proof of this Lemma is deferred to Section E.1. Now we show how it implies the desired lower bound. Since  $\rho(G') < 0.5$ ,

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

where (a) is from Lemma 12<sup>1</sup> under assuming SCCC and (b) follows from (35) 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.

## E.1 Proof of Lemma 13

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

**Definition 4.** *Given connected 3-regular graphs  $G_1 = (V_1, E_1)$ ,  $G_2 = (V_2, E_2)$  and the edge  $e \in E_1$ , create a new 3-regular  $(G_1, e) \odot G_2$  as follows:*

1. Produce union of  $G_1$  and  $G_2$ ,  $G = G_1 \cup G_2$  - a disconnected graph with connected components as  $G_1$  and  $G_2$ .
2. Add two vertices  $v_1, v_2$  inside 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  $e_2$ .

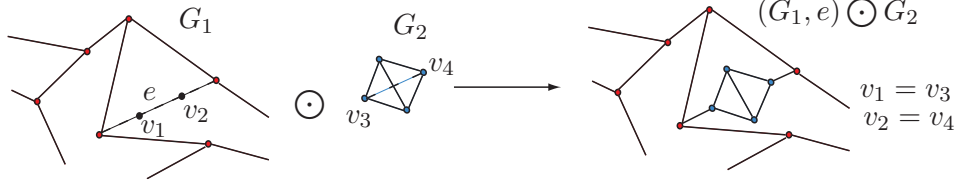


Figure 1: 3-regular  $(G_1, e) \odot G_2$  is created from 3-regular graphs  $G_1$  and  $G_2$  as per Definition 4.

Now we study the effect of operator  $\odot$  on function  $\rho$  defined in (35). Let  $G_3 = (G_1, e) \odot G_2$  and our interest is in bounding  $\rho(G_3)$  in terms of  $\rho(G_1)$  and  $\rho(G_2)$ . Now by definition (35), the  $\rho(G_3)$  is summation of terms over simple cycles of  $G_3$ . Now 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 the type (a) and (b), their 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 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

$$\rho(G_3) \leq \rho(G_1 \setminus \{e\}) + \rho(G_2) + (\rho(G_1) - \rho(G_1 \setminus \{e\})) \times \rho(G_2) \times \frac{1}{c_\alpha}. \quad (36)$$

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

$$\begin{aligned} \rho(G_3) &\leq \rho(G_1 \setminus \{e\}) + \frac{\rho(G_1) - \rho(G_1 \setminus \{e\})}{B} + \frac{\rho(G_1) - \rho(G_1 \setminus \{e\})}{B} \\ &= \rho(G_1 \setminus \{e\}) + \frac{2}{B} (\rho(G_1) - \rho(G_1 \setminus \{e\})) \\ &= \rho(G_1) - \left(1 - \frac{2}{B}\right) (\rho(G_1) - \rho(G_1 \setminus \{e\})). \end{aligned} \quad (37)$$

Equipped with our understanding of effect of  $\odot$  on  $\rho$  and (37), we describe the following procedure to construct graph  $G'$  desired in Lemma 13. To this end, given a random 3-regular graph  $G$ , generate  $G'$  iteratively as follows:

- Initially, iteration  $t = 0$  and  $G'(0) = G$ .
- Let  $g$  be the smallest number such that  $c_\alpha \sum_{i \geq g} C_i (0.49)^i < 0.25$ , where  $C_i$  is the number of cycles of length  $i$  in  $G$ .

---

<sup>1</sup>Recall that Lemma 12 only uses the fact that the graph under consideration is 3-regular, but does not require it being 'random'.

- Repeat the following till  $G'(t)$  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  $G'(t+1) = (G'(t), e_t) \odot G_2$ , where  $G_2$  has a 3-regular graph that will be chosen later.
  3. Increment  $t$  by 1.
- Output  $G' = G'(t)$ .

To establish that  $G'$  thus produced has properties 1, 2 and 3 of Lemma 13, it is sufficient to show that, with probability  $1 - \varepsilon$ , the repeat-loop in the above described procedure terminates in  $\Gamma_1(\varepsilon)$  steps,  $\rho(G') < 0.5$  and  $G_2$  of size  $\Gamma_2(\varepsilon)$ . Here and in what follows  $\Gamma_1(\varepsilon), \Gamma_2(\varepsilon), \dots$  are constants dependent on  $\varepsilon$  and independent of  $n$ . To this end, recall the definition of  $\rho$  in (35):

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

For random 3-regular graph, we have

$$\mathbb{E}[C_i] \leq 2^{i-1}/i.$$

Therefore, using Markov's inequality it follows that with probability  $1 - \varepsilon$  for an appropriate choice of  $g = \Gamma_3(\varepsilon)$ ,

$$\sum_{i < g} C_i = \Gamma_1(\varepsilon), \quad \text{and} \quad c_\alpha \sum_{i \geq g} C_i (0.49)^i < 0.25.$$

Clearly, under this event 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 towards completing the proof of Lemma 13 is to establish existence of graph  $G_2$  such that it has (a) size  $\Gamma_2(\varepsilon)$ , (b) girth larger than  $g = \Gamma_3(\varepsilon)$ , and (c) the resulting  $G'$  has  $\rho(G') < 0.5$ . Towards this, suppose  $G_2$  can be chosen so that for all rounds  $t \leq \Gamma_1(\varepsilon)$  with  $B \geq 2$ ,

$$\rho(G_2) \leq \min \left\{ \frac{\rho(G'(t)) - \rho(G'(t) \setminus \{e_t\})}{B}, \frac{c_\alpha}{B} \right\} \quad (38)$$

Under this assumption, we obtain the following bound on  $\rho(G')$  using (37) recursively:

$$\begin{aligned} \rho(G') &\leq \rho(G) - \sum_t \left(1 - \frac{2}{B}\right) (\rho(G'(t)) - \rho(G'(t) \setminus \{e_t\})) \\ &\leq \rho(G) - \left(1 - \frac{2}{B}\right) \left(\sum_t \rho(G'(t)) - \rho(G'(t) \setminus \{e_t\})\right) \\ &\stackrel{(a)}{\leq} \rho(G) - \left(1 - \frac{2}{B}\right) c_\alpha \left(\sum_{i < g} C_i (0.49)^i\right) \\ &\stackrel{(b)}{<} \rho(G) - \left(1 - \frac{2}{B}\right) (\rho(G) - 0.25) \\ &\leq \frac{2}{B} \rho(G) + 0.25. \end{aligned}$$

In above, (a) is because the each cycle of length upto  $g$  is 'broken' in one of the steps  $t \leq \Gamma_1(\varepsilon)$ . And, each of the term  $\rho(G'(t)) - \rho(G'(t) \setminus \{e_t\})$  accounts for all broken cycles in round  $t$ . Therefore, the

bound used in (a) follows. For (b), by definition of  $g$ , we have  $c_\alpha \sum_{i \geq g} C_i (0.49)^i < 0.25$ . Therefore, if we choose  $B = 8\rho(G)$ , the desired bound  $\rho(G') < 0.5$  follows.

In summary, now we are left with showing existence of  $G_2$  that has properties (a) size  $\Gamma_2(\varepsilon)$ , (b) girth larger than  $g = \Gamma_3(\varepsilon)$ , and (c) the condition (38) with  $B = 8\rho(G)$ . Now the choice of  $B$  suggests that  $B = \Gamma_4(\varepsilon)$  (due to selection of event that has probability at least  $1 - \varepsilon$ ). Now consider

$$\rho(G'(t)) - \rho(G'(t) \setminus \{e\}) \geq c_\alpha (0.49)^g = \Gamma_5(\varepsilon), \quad (39)$$

where we have used the fact that for  $t \leq \Gamma_1(\varepsilon)$ , 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

$$\frac{\rho(G'(t)) - \rho(G'(t) \setminus \{e\})}{B} \geq \Gamma_5(\varepsilon)/\Gamma_4(\varepsilon). \quad (40)$$

Hence to satisfy (c), it is sufficient to show that there exists  $G_2$  with arbitrarily small girth and  $\rho(G_2)$  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 definition of  $\rho$ . Precisely this is established 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  $\log_{1/0.49} \left(\frac{c_\alpha}{\delta}\right)$ .*

*Proof.* Recall that  $C_i$  is the number of cycles in the graph  $G_2$ . For a random 3-regular graph, it is well-known that [20] for  $3 \leq r \leq \frac{1}{3} \log n$ ,  $C_r$  become asymptotically independent Poisson random variables with mean  $\mu_r = 2^{r-1}/r$ . Thus, for  $g < \frac{1}{3} \log n$ ,

$$\Pr[C_3 = C_4 = \dots C_g = 0] \approx e^{-e^{\Theta(g)}}. \quad (41)$$

Now, we divide the summation  $c_\alpha \sum_i C_i (0.49)^i = \sum_i a_i$  with  $a_i \triangleq c_\alpha C_i (0.49)^i$  into the following 3 terms:  $A_1 = \sum_{r < g} a_r$ ,  $A_2 = \sum_{g \leq r < \frac{1}{3} \log n} a_r$  and  $A_3 = \sum_{g \geq \frac{1}{3} \log n} a_r$ . Define the events  $E_1$ ,  $E_2$  and  $E_3$  such that  $E_1$  be the event of  $A_1 = 0$ ,  $E_2$  be the event of  $A_2 \leq 2\mathbb{E}[A_2]$  and  $E_3$  be the event of  $A_3 \leq (3 \log n)\mathbb{E}[A_3]$ . From Markov inequality,  $\Pr[E_2] \geq \frac{1}{2}$  and  $\Pr[E_3] \geq 1 - \frac{1}{3 \log n}$ . For  $E_1$ , one can choose  $g = \Theta(\log \log \log n)$  from (41) such that  $\Pr[E_1] = \frac{1}{\log n}$ . Therefore, we have

$$\Pr[E_1 \cap E_2 \cap E_3] \geq \Pr[E_1 \cap E_2] + \Pr[E_3] - 1 \geq \frac{1}{2 \log n} + 1 - \frac{1}{3 \log n} - 1 > 0,$$

where we use the union bound and the independence between  $E_1$  and  $E_2$ . 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  $E_3$ ,

$$\rho(G_2) \leq 2\mathbb{E}[A_2] + (3 \log n)\mathbb{E}[A_3] \leq O(1) \times (0.98)^g + O(3 \log n) \times (0.98)^{\frac{1}{3} \log n} \rightarrow 0,$$

as  $n$  goes to  $\infty$ . In above, we have used the fact that  $\mathbb{E}[C_r] \leq 2^{r-1}/r$ . 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  $\rho$ .  $\square$