



Dimension reduction in vertex-weighted exponential random graphs

Ryan DeMuse, Mei Yin*

Department of Mathematics, University of Denver, Denver, CO 80208, USA

ARTICLE INFO

Article history:

Received 13 February 2020

Received in revised form 26 August 2020

Available online 15 September 2020

Keywords:

Exponential random graphs

Stochastic block model

Vertex-weighted

ABSTRACT

We investigate the behavior of vertex-weighted exponential random graphs. We show that vertex-weighted exponential random graphs with edge weights induced by products of independent vertex weights are approximate mixtures of graphs whose vertex weight vector is a near fixed point of a certain vector equation. For graphs with Hamiltonians counting cliques, it is demonstrated that, under appropriate conditions, every solution to this equation is close to a block vector with a small number of communities. We prove that for the cases of positive weights and small weights in the Hamiltonian in particular, the vector equation has a unique solution. Lastly, the behavior of vertex-weighted exponential random graphs counting triangles is studied in detail and the solution to the vector equation is shown to approach the zero vector as the weight diverges to negative infinity for sufficiently large networks.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Complex networks have become omnipresent structures, especially with the popularity of technological and social networks, but also in economics, healthcare, and infrastructure [1–3]. Their ubiquity has hastened the need to develop new models that aid in the study of these elaborate networks and techniques to study their local and global structures, as well as their development over time. One of the features of many modern networks that makes them difficult to study is the fact that one cannot assume, and is generally untrue, that edges between different vertices are independent of each other. The random graph model $G(n, p)$ is a well-studied probabilistic graph model where edges are placed between the n vertices independently with identical probability p [4]. This independence assumption is one of the characteristics of $G(n, p)$ that makes it so amenable to study. Unfortunately, this simultaneously makes $G(n, p)$ inadequate for modeling many modern networks. Think of the example of social networks, whose popularity has skyrocketed in recent years. Consider Persons A , B , and C . If A is a friend of B and B is a friend of C , it is more likely for A to be a friend of C than it would be if there were no friendships between A and B , and B and C . This transitivity of friendship leads us to assume that edges in social networks are unlikely to appear independently. Similarly, the mechanism that suggests new friends for users in social networks generically works upon the presumption that if Person X is a friend to a large number of your friends, there is a greater chance that you would like to connect with Person X than with some random user.

The exponential random graph model (ERGM) is a popular model for random graphs, which assigns exponential families of probability distributions over spaces of graphs. Let \mathcal{G}_n be the space of all finite simple graphs on n vertices and $G \in \mathcal{G}_n$. The probability of G in an exponential random graph model is given by

$$\mathbb{P}(\alpha \mid G) = \exp \left(n^2 \left(f(\alpha, G) - \psi_n^f \right) \right), \quad (1.1)$$

* Corresponding author.

E-mail addresses: ryan.demuse@du.edu (R. DeMuse), mei.yin@du.edu (M. Yin).

where $f : \mathcal{G}_n \rightarrow \mathbb{R}$ is some function on the graph space, and ψ_n^f is a normalization constant given by

$$\psi_n^f = \frac{1}{n^2} \log \sum_{G \in \mathcal{G}_n} e^{n^2 f(\alpha, G)}.$$

The function f is referred to as the *Hamiltonian* and typically consists of a weighted sum of subgraph counting functions. We remark that it is in general not necessary for f to be a sum of subgraph counting functions. The Hamiltonian could be some other function on the graph space such as the degree distribution; however, in the present setting we will mostly concentrate on f counting subgraphs. Some popular, well-studied Hamiltonians include the edge-two-star and edge-triangle models. For example, the Hamiltonian of the edge-triangle model is given by

$$f(\alpha, G) = \alpha_1 t(K_2, G) + \alpha_2 t(K_3, G),$$

where $\alpha_1, \alpha_2 \in \mathbb{R}$ are weights, K_k is the complete graph on k vertices, and $t(K_k, G)$ is a homomorphism density function. For any finite simple graphs H and G , define

$$t(H, G) = \frac{\text{hom}(H, G)}{|V(G)|^{|V(H)|}},$$

where $\text{hom}(H, G)$ is the number of edge-preserving homomorphisms of H in G , and $V(H)$, $V(G)$ are the vertex sets of H and G , respectively. This quantity can be interpreted as the probability that a random map from the set $V(G)^{V(H)}$ is edge-preserving.

Even though exponential random graph models on simple graphs already display a wide variety of behaviors and modeling capabilities, the assumption of non-weighted simple graphs still renders them somewhat insufficient when it comes to the modeling of many intricate complex networks [5,6]. Social networks exhibit clustering of vertices which corresponds to the formation of cliques as an important subgraph structure of a social network. Groups of friends or colleagues in a social network tend to form clusters, where the number of friendship relations in the cluster is close to the maximum possible. In the present work, we place weights on the vertices to accommodate this phenomenon. To start, we restrict our attention to vertex-weighted exponential random graph models with vertex weights having values 0 or 1. The case of vertex weights being 0-1 random variables encourages the formation of cliques if the weight is one, and discourages their formation if the weight is zero. We assume that the vertices have weight 1 with probability p and 0 with probability $1-p$. Thus we obtain a distribution on the space of 0-1 vertex-weighted graphs such that any vertex-weighted graph G has probability

$$\mathbb{P}(\alpha, G) = \mathbb{P}(\alpha \mid G) \mathbb{P}(G), \quad (1.2)$$

where the normalization constant

$$\psi_n^f = \frac{1}{n} \log \mathbb{E} \left(e^{n f(\alpha, G)} \right), \quad (1.3)$$

and the underlying measure on graphs induced by the distribution of vertex weights is given by

$$\mathbb{P}(G) = p^{w(G)} (1-p)^{n-w(G)},$$

where $w(G)$ is the number of vertices with weight 1 and \mathbb{E} is the associated expectation.

There has been previous work done in the area of vertex-weighted ERGM including [7], where expanding upon [8] the efficiency of the Glauber dynamics on vertex-weighted ERGM was considered, and [9], where under suitable assumptions on the vertex weights continuity of functionals in the specification of the model was addressed. Unlike the traditional ERGM, which in certain regimes is indistinguishable from an Erdős-Rényi random graph, the vertex-weighted ERGM is amenable to consistent estimation using pseudo likelihood at least in some scenarios. The first such work is [10]. There are other works as well for example [11,12]. This work builds upon the framework of [13]; therein, Eldan and Gross showed that any exponential random graph model where the Hamiltonian counts subgraphs is close in 1-norm distance to a mixture of stochastic block models with a small number of blocks (Theorem 14 in [13]). Furthermore, the authors showed that if the weights are positive or sufficiently small (either positive or negative), there is a constant matrix X_c , such that every mixture element is close in 1-norm distance to the constant matrix (Theorems 18 and 19 in [13]). This implies that mixture elements behave like a $G(n, p)$ random graph since constant stochastic block matrices are equivalent to the random graph model $G(n, p)$.

We extend the results given in Theorems 14, 18, and 19 of [13] to the vertex-weighted setting. The probability distribution defined in Eq. (1.2) may be explicitly thought of as a distribution on the vector of vertex weights. That is, for $X \in \{0, 1\}^n$, let

$$\mathbb{P}(\alpha, X) = \mathbb{P}(\alpha \mid X) \mathbb{P}(X), \quad (1.4)$$

with

$$\mathbb{P}(\alpha \mid X) = \exp \left(f(\alpha, X) - n \psi_n^f \right),$$

where for ease of computation we incorporate the number of vertices n into the Hamiltonian and ψ_n^f is similarly defined as in Eq. (1.3). In the context of 0-1 vertex weights, the underlying distribution on the vertex weight configurations becomes

$$\mathbb{P}(X) = p^{\|X\|_1} (1-p)^{n-\|X\|_1},$$

where $\|\cdot\|_1$ is the 1-norm of the vector X . The results of [13] and, by extension, this paper relies on the set of near fixed-points of a certain function on the vertex weight configurations, specifically the set of X such that

$$\|X - (\mathbf{1}_n + \tanh(\nabla f(\alpha, X))) / 2\|_1 \leq cn^{7/8} \quad (1.5)$$

for some constant c , where \tanh is applied element-wise, $\mathbf{1}_n$ is the n -dimensional all ones vector, and $\nabla f(\alpha, X)$ is the discrete gradient of the subgraph counting function f as defined in Eq. (2.5). Theorem 4.1 shows that for any Hamiltonian counting cliques, there exists a mixture measure such that the support of the measure is, asymptotic in n , given by the set of configurations satisfying Eq. (1.5). Theorem 4.4 further shows that any vertex weight configuration on n vertices is close to a configuration with the number of vertex communities independent of n in 1-norm distance. Theorems 4.7 and 4.9 prove the existence and uniqueness of a constant vector such that every mixture element is close to this constant vector when the weights are positive or sufficiently small in absolute value. In other words, we can partition the vertex set into communities where two vertices in the same community have the same vertex weight. This allows one to reduce the dimension by considering the fixed number of communities and the weight that vertices common to the same community share. We remark that the concentration phenomenon of vertex weights on a small number of communities has also been observed in the prototype Ising model, and the number of such communities change above a certain threshold in the high temperature regime [11]. This is not a coincidence; in later sections we will make explicit the hidden connections between the Ising model and the vertex-weighted ERGM.

2. Background

Let $n > 0$ be the number of vertices. Let the set $\mathcal{C}_n = \{0, 1\}^n$ represent the n -dimensional, discrete, Boolean hypercube and $\overline{\mathcal{C}}_n = [0, 1]^n$ be the continuous hypercube. We also use the convention $[n] = \{1, 2, \dots, n\}$. A vector $X \in \mathcal{C}_n$ represents a vertex-weight configuration for a graph on n vertices, where X_i is the weight at vertex i , and similarly for $X \in \overline{\mathcal{C}}_n$. Let $f : \mathcal{C}_n \rightarrow \mathbb{R}$ be a subgraph counting function of the form

$$f(\alpha, X) = n \left(\sum_{q=1}^l \frac{\alpha_q}{n^{m_q}} \sum_{i_1 \neq \dots \neq i_{m_q}} X_{i_1} \cdots X_{i_{m_q}} \right) + \log \left(\frac{p}{1-p} \right) \|X\|_1, \quad (2.1)$$

where $m_q \geq 2$ are distinct integers, $\alpha_q \in \mathbb{R}$, and $p \in (0, 1)$. Note that the probability distribution induced by the vertex weights has been subsumed into the Hamiltonian. That is, we are considering a subgraph counting function that counts the number of certain cliques that appear in the graph with vertex weights given by $X \in \{0, 1\}^n$. Often we will take $m_1 = 2$ counting the number of edges, whose appearance alone gives the Ising model with an external magnetic field. Consider an edge $e = \{i, j\}$ present if the product $X_i X_j = 1$. This only occurs if both X_i and X_j are vertex weight 1. As previously stated, we let $\|\cdot\|_1$ be the 1-norm of a vector in \mathcal{C}_n and $\|\cdot\|_2$ be the 2-norm so that for $X \in \mathcal{C}_n$,

$$\|X\|_1 = \sum_{i=1}^n X_i \quad \text{and} \quad \|X\|_2 = \sqrt{\sum_{i=1}^n X_i^2}.$$

2.1. Graph limits and the variational problem

In their seminal paper [14], Chatterjee and Diaconis utilized the framework of graph limit theory to show that certain exponential random graph models behave like Erdős-Rényi random graphs. We present some of the basics of graph limit theory here so that we may state a few of their results. Let \mathcal{W} be the space of all symmetric, measurable functions $h : [0, 1]^2 \rightarrow [0, 1]$, called the *graphon* or *graph limit space*. From any simple graph G on n vertices, one can construct an associated graphon, h^G , such that

$$h^G(x, y) = \begin{cases} 1 & \text{if } \{\lceil nx \rceil, \lceil ny \rceil\} \text{ is an edge in } G, \\ 0 & \text{otherwise.} \end{cases}$$

Define the *cut norm* on the graphon space \mathcal{W} as

$$\|h\|_{\square} = \sup_{A, B \subseteq [0, 1]} \left| \int_{A \times B} h(x, y) dx dy \right|.$$

This gives a pseudometric such that for $h, g \in \mathcal{W}$,

$$d_{\square}(h, g) = \|h - g\|_{\square}.$$

The induced metric is given by identifying equivalent graphons h, g such that $h \sim g$ if they agree on a set of full measure or there exists a measure preserving bijection $\sigma : [0, 1] \rightarrow [0, 1]$ such that $h(x, y) = g(\sigma x, \sigma y) = g_\sigma(x, y)$. Denote by \tilde{h} the closure in (\mathcal{W}, d_\square) of the orbit $\{h_\sigma\}$. A quotient space $\tilde{\mathcal{W}}$ is obtained, referred to as the *reduced* or *unlabeled graphon space*, and d_\square becomes a metric on this space, δ_\square , by defining

$$\delta_\square(\tilde{h}, \tilde{g}) = \inf_{\sigma} d_\square(h, g_\sigma).$$

It follows that $(\tilde{\mathcal{W}}, \delta_\square)$ is a compact metric space and the homomorphism density functions are continuous with respect to the cut distance [15,16].

Let f be a Hamiltonian on the reduced graphon space and consider $\psi_\infty(f) := \lim_{n \rightarrow \infty} \psi_n(f)$, known as the *limiting normalization constant*. Define $I : [0, 1] \rightarrow \mathbb{R}$ as

$$I(u) = \frac{1}{2}u \log u + \frac{1}{2}(1-u) \log(1-u),$$

and extend I to $\tilde{\mathcal{W}}$ as

$$I(\tilde{h}) = \int_0^1 \int_0^1 I(h(x, y)) dx dy,$$

where h is any representative element of the equivalence class \tilde{h} .

Theorem 2.1 (Theorem 3.1 in [14]). *If f is a bounded continuous function on the reduced graphon space, then*

$$\psi_\infty(f) = \sup_{\tilde{g} \in \tilde{\mathcal{W}}} \{f(\tilde{g}) - I(\tilde{g})\}. \quad (2.2)$$

Using this formulation of the limiting normalization constant, Chatterjee and Diaconis showed that a graph drawn from the model defined in Eq. (1.1) lies close in cut distance to the set of maximizers of the limiting normalization constant.

Theorem 2.2 (Theorem 3.2 in [14]). *Let G be drawn from \mathbb{P} defined in Eq. (1.1) and \tilde{F}^* be the set of maximizers of Eq. (2.2). Then for all $\varepsilon > 0$ there exists a $C > 0$ and a $\delta > 0$ such that for all $n \in \mathbb{N}$,*

$$\mathbb{P}(\delta_\square(\tilde{g}^G, \tilde{F}^*) > \varepsilon) \leq Ce^{-n^{2\delta}}. \quad (2.3)$$

Specializing to subgraph counting function f of the form

$$f(\alpha, G) = \sum_{q=1}^l \alpha_q t(H_q, G),$$

where $\alpha_q \in \mathbb{R}$ and H_q are finite, simple graphs, Chatterjee and Diaconis further derived the following.

Theorem 2.3 (Theorem 4.2 in [14]). *If $H_1 = K_2$ and $\alpha_2, \dots, \alpha_m$ are nonnegative, then the set of maximizers \tilde{F}^* may be identified as a finite subset of $[0, 1]$ and*

$$\min_{u \in \tilde{F}^*} \delta_\square(\tilde{g}^G, u) \rightarrow 0$$

in probability as $n \rightarrow \infty$. That is, the sampled graph G behaves like an Erdős–Rényi random graph $G(n, u)$ where u is picked randomly from some distribution on \tilde{F}^ .*

These results for standard exponential random graphs were extended to the realm of edge-weighted exponential random graphs by Yin in [17]. However, by utilizing graphons and the cut distance, the approach is inherently limited to the dense regime. In this paper, we will prove corresponding results for vertex-weighted exponential random graphs in both the dense and the sparse regime. Moreover, the metric used in the current investigation is the 1-norm distance, which induces a finer topology than the cut distance [16]. This different toolset and perspective yields concrete information about the nature of dependence between the vertex weights and other aspects of the random graph that are not captured by the cut metric. Our results are technical and yet challenging, in particular because the approximation error bound depends sublinearly on the dimension of the vector space. See Theorem 4.1 and the associated Eq. (1.5), and Theorems 4.4, 4.7 and 4.9 and the following remarks, where n denotes the number of vertices. On the contrary, the approximation error bound concerning the cut distance for standard and edge-weighted exponential random graphs are linear in terms of the dimension. See Eq. (2.3), where n^2 in the exponent corresponds with the number of edges and is the dimension of the vector space in this counterpart setting.

2.2. Definitions

Several results in this paper are first proven for a single subgraph counting function, then generalized to sums using triangle inequalities and subadditivity. For a Hamiltonian counting a single type of subgraph, we have

$$f(\alpha, X) = \frac{\alpha}{n^{m-1}} \left(\sum_{i_1 \neq \dots \neq i_m} X_{i_1} \cdots X_{i_m} \right) + \log \left(\frac{p}{1-p} \right) \|X\|_1 \quad (2.4)$$

for some positive integer $m \geq 2$, $\alpha \in \mathbb{R}$, and $p \in (0, 1)$. A single subgraph counting function in this work therefore counts the number of m -cliques in the graph obtained from the vertex configuration X . For a subgraph counting function f and $j \in [n]$, define the *discrete derivative* of f at vertex j as

$$\partial_j f(\alpha, X) = \frac{1}{2} \left(f(\alpha, X_1, \dots, X_{j-1}, 1, X_{j+1}, \dots, X_n) - f(\alpha, X_1, \dots, X_{j-1}, 0, X_{j+1}, \dots, X_n) \right).$$

This further gives a definition of the *discrete gradient* of the function f , $\nabla f(\alpha, X)$, as the vector of discrete derivatives where

$$\nabla f(\alpha, X) = (\partial_1 f(\alpha, X), \dots, \partial_n f(\alpha, X)), \quad (2.5)$$

and the Lipschitz constant of f is given by

$$\text{Lip}(f) = \max_{i \in [n], X \in \mathcal{C}_n} |\partial_i f(\alpha, X)|.$$

It will be convenient to define C_j^m for $j \in [n]$ and $m \geq 2$ as

$$C_j^m(X) = \sum_{j \neq i_1 \neq \dots \neq i_{m-1}} X_{i_1} \cdots X_{i_{m-1}}, \quad (2.6)$$

with the dependence on X suppressed when it is clear from the context. For $m = 1$ let $C_j^1 = 1$, which corresponds to the sum of the empty product. With this notation and f a single subgraph counting function, for $j \in [n]$, we obtain a more compact form of the discrete derivative of f at coordinate j as

$$\partial_j f(\alpha, X) = \frac{m\alpha}{2n^{m-1}} C_j^m + \frac{1}{2} \log \left(\frac{p}{1-p} \right). \quad (2.7)$$

Let $\mathbf{C}^m \in \mathbb{R}^n$ be the vector $\mathbf{C}^m = (C_1^m, C_2^m, \dots, C_n^m)$ for $m \in \mathbb{N}$. Since a generic subgraph counting function is a weighted sum of single subgraph counting functions, the discrete gradient of a generic subgraph counting function may then be written as a weighted sum of vectors \mathbf{C}^m . The following [Lemma 2.4](#) shows that the vector \mathbf{C}^m can be defined inductively as a matrix product.

Lemma 2.4. *Let J be the $n \times n$ matrix consisting of all ones and I the $n \times n$ identity matrix. If \mathbf{C}^m is defined as above, then*

$$\mathbf{C}^m = \mathbf{C}^{m-1} \mathbf{X} (J - (m-1)I), \quad (2.8)$$

where \mathbf{X} is the matrix with the vector X on the main diagonal and zeros off diagonal.

Proof. Firstly, note that we can write C_j^m as

$$C_j^m = \left(\sum_{i=1}^n X_i C_i^{m-1} \right) - (m-1) X_j C_j^{m-1}. \quad (2.9)$$

This can be argued as follows. We will start with the right hand side, which can be expanded as

$$\left(\sum_{i=1}^n X_i \sum_{i \neq i_1 \neq \dots \neq i_{m-2}} X_{i_1} \cdots X_{i_{m-2}} \right) - (m-1) X_j \sum_{j \neq i_1 \neq \dots \neq i_{m-2}} X_{i_1} \cdots X_{i_{m-2}}. \quad (2.10)$$

The expression in parentheses in Expression (2.10) is equivalent to

$$\sum_{i_1 \neq \dots \neq i_{m-1}} X_{i_1} \cdots X_{i_{m-1}},$$

while the term on the right represents all terms involving the variable X_j . Note that there are $m-1$ choices to place the variable X_j in the product. Once X_j is factored out, we simply multiply by all possible nomials of distinct terms not involving X_j . Eq. (2.9) thus holds. Using this equation, we find

$$\mathbf{C}^{m-1} \mathbf{X} (J - (m-1)I) = (X_1 C_1^{m-1}, \dots, X_n C_n^{m-1}) (J - (m-1)I)$$

$$\begin{aligned}
&= \left(\sum_{j=1}^n X_j C_j^{m-1} \right) \mathbf{1}_n - (m-1) (X_1 C_1^{m-1}, \dots, X_n C_n^{m-1}) \\
&= \mathbf{C}^m,
\end{aligned}$$

and the desired expression readily follows. ■

Note that we can also write \mathbf{C}^m as an elementary matrix product of the form

$$\mathbf{C}^m = \mathbf{1}_n \prod_{i=1}^{m-1} \mathbf{X}(J - iI),$$

with J , I , and \mathbf{X} as defined in [Lemma 2.4](#), [Theorem 3.1](#), which appears as a main structural theorem in [\[18\]](#), will constitute a backbone of one of the main results of this paper, showing that vertex-weighted exponential random graphs are close to mixtures of independent graph models. The formulation of this theorem requires the concept of the *Gaussian-width* of a subset of \mathbb{R}^n .

Definition 2.5. The **Gaussian-width** of a set $K \subseteq \mathbb{R}^n$ is defined as

$$\text{GW}(K) = \mathbb{E} \left(\sup_{X \in K} \langle X, \Gamma \rangle \right),$$

where $\Gamma \sim N(0, I)$ is a standard normal Gaussian vector in \mathbb{R}^n . For a function $f : \mathcal{C}_n \rightarrow \mathbb{R}$, the *gradient complexity* of f is defined as

$$\mathcal{D}(f) = \text{GW}(\{\nabla f(\alpha, X) : X \in \mathcal{C}_n\} \cup \{0\}).$$

Definition 2.6. For $\mathbf{p} \in [0, 1]^n$, denote by $G(n, \mathbf{p})$ the random vertex-weighted graph such that vertex i has a Bernoulli 0-1 weight with success probability \mathbf{p}_i . Let ρ be a probability measure on $[0, 1]^n$. Define $G(n, \rho)$ as

$$\mathbb{P}(G(n, \rho) = G) = \int \mathbb{P}(G(n, \mathbf{p}) = G) d\rho(\mathbf{p}),$$

and we say $G(n, \rho)$ is a ρ -*mixture*.

Definition 2.7. Let $\delta > 0$ and let ρ be a probability measure on $[0, 1]^n$. A random vertex-weighted graph G is called a (ρ, δ) -*mixture* if there exists a coupling between $G(n, \rho)$ and G such that

$$\mathbb{E} \|G(n, \rho) - G\|_1 \leq \delta n.$$

Definition 2.8. A k -**block vector** or a **block vector with k communities** is a vector $v \in \mathbb{R}^n$ such that there exists a set $U \subseteq \mathbb{R}$ with $|U| = k$ and $v_i \in U$ for all $1 \leq i \leq n$.

3. Setup

We wish to use Theorem 9 of [\[18\]](#) to show that vertex-weighted exponential random graphs behave like $G(n, \rho)$ mixtures for some measure ρ . The theorem is stated as follows.

Theorem 3.1 (Theorem 9 in [\[18\]](#)). Let $n > 0$, $f : \mathcal{C}_n \rightarrow \mathbb{R}$ a Hamiltonian, and X_n^f a random vector given by

$$\mathbb{P}(X_n^f = X) = \exp(f(\alpha, X)) / Z,$$

where Z is a normalizing constant. Denote

$$D = \mathcal{D}(f),$$

$$L_1 = \max\{1, \text{Lip}(f)\},$$

$$L_2 = \max \left\{ 1, \max_{X \neq Y \in \mathcal{C}_n} \frac{\|\nabla f(\alpha, X) - \nabla f(\alpha, Y)\|_1}{\|X - Y\|_1} \right\}.$$

Furthermore, define \mathcal{X}_f to be the set

$$\mathcal{X}_f = \left\{ X \in \overline{\mathcal{C}_n} : \left\| X - \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} \right\|_1 \leq 5000 L_1 L_2^{3/4} D^{1/4} n^{3/4} \right\},$$

where $\mathbf{1}_n$ is the n -dimensional all ones vector, $\nabla f(\alpha, X)$ is extrapolated to \bar{C}_n by Eq. (2.7) and with tanh applied entry-wise. Then X_n^f is a $(\rho, 80 \frac{D^{1/4}}{n^{1/4}})$ -mixture such that

$$\rho(X_f) \geq 1 - 80 \frac{D^{1/4}}{n^{1/4}}.$$

In particular, if $D = o(n)$, then X_n^f is a $(\rho, o(1))$ -mixture with $\rho(X_f) = 1 - o(1)$.

In order to use Theorem 3.1, we must establish bounds on D , L_1 , and L_2 . Throughout Section 3, we assume that f is a subgraph counting function of the form in Eq. (2.1). Consider first the determination of D . Using Lemma 2.4 and the definition of Gaussian-width we derive a bound on $\mathcal{D}(f)$ for any f of the form (2.1).

Lemma 3.2. *Let f be a subgraph counting function of the form in Eq. (2.1). Then*

$$\mathcal{D}(f) \leq \frac{\sqrt{n}}{2} \left(\sum_{q=1}^l m_q |\alpha_q| \sqrt{1 + \frac{(2 - m_q)^2 - 1}{n}} \right).$$

Proof. For a single subgraph counting function, by Lemma 2.4, the gradient of f has the form

$$\nabla f(\alpha, X) = \frac{1}{2} \frac{m\alpha}{n^{m-1}} \mathbf{C}^{m-1} \mathbf{X} (J - (m-1)I) + \frac{1}{2} \log \left(\frac{p}{1-p} \right) \mathbf{1}_n.$$

Recall Definition 2.5 for the gradient complexity $\mathcal{D}(f)$. Taking the expectation over Γ yields

$$\begin{aligned} \mathcal{D}(f) &\leq \frac{1}{2} \mathbb{E} \sup_{X \in C_n} \left| \left\langle \frac{m\alpha}{n^{m-1}} \mathbf{C}^{m-1} \mathbf{X} (J - (m-1)I), \Gamma \right\rangle \right| \\ &\quad + \frac{1}{2} \mathbb{E} \left| \log \left(\frac{p}{1-p} \right) \mathbf{1}_n, \Gamma \right| \\ &\leq \frac{m|\alpha|}{2} \mathbb{E} \sup_{X \in C_n} \left| \left\langle n^{-(m-2)} \mathbf{C}^{m-1} \mathbf{X}, n^{-1}(J - (m-1)I)\Gamma \right\rangle \right| \\ &\leq \frac{m|\alpha|}{2} \sqrt{\sup_{X \in C_n} \left\| n^{-(m-2)} \mathbf{C}^{m-1} \mathbf{X} \right\|_2^2 \mathbb{E} \left\| n^{-1}(J - (m-1)I)\Gamma \right\|_2^2} \\ &\leq \frac{m|\alpha|}{2} \sqrt{n \operatorname{tr}((n^{-1}(J - (m-1)I))^2)} \\ &= \frac{m|\alpha|}{2} \sqrt{n \left(1 + \frac{(2-m)^2 - 1}{n} \right)}. \end{aligned} \quad (3.1)$$

Cauchy-Schwarz is used from Line 2 to 3 in the above Eq. (3.1). From Line 3 to 4, we are using that $\mathbb{E} \|A\Gamma\|_2^2 = \operatorname{tr}(A\Sigma A^T)$, where Σ is the covariance matrix of the Gaussian vector Γ , and $A = n^{-1}(J - (m-1)I)$. This general fact may be justified as follows. Set $\Delta = A\Gamma$, then the transformed Gaussian vector $\Delta \sim \mathcal{N}(0, A\Sigma A^T)$. Therefore

$$\mathbb{E} \|A\Gamma\|_2^2 = \mathbb{E} [\Delta \Delta^T] = \sum_{i=1}^n \mathbb{E} [\Delta_i^2] = \sum_{i=1}^n \operatorname{Var}(\Delta_i) = \operatorname{tr}(A\Sigma A^T).$$

Since $\Gamma \sim \mathcal{N}(0, I)$, $\Sigma = I$. Further, if A is symmetric, then $AA^T = A^2$. Lastly, following Lemma 2.4, the supremum over binary vectors C_n is achieved when $X = \mathbf{1}_n$. Note that

$$\mathbf{C}^{m-1} \mathbf{X} = (X_1 C_1^{m-1}, \dots, X_n C_n^{m-1}).$$

By Eq. (2.6), each entry of the vector $\mathbf{C}^{m-1} \mathbf{X}$ is thus bounded above by n^{m-2} .

We now turn to generic subgraph counting functions with multiple components. Take f a subgraph counting function of the form (2.1), where m_1, \dots, m_l are distinct positive integers with $m_i \geq 2$ and $\alpha_i \in \mathbb{R}$. Using the subadditivity of the gradient complexity, the conclusion readily follows. The bound on $\mathcal{D}(f)$ may be relaxed further as

$$\mathcal{D}(f) \leq \frac{\sqrt{n}}{2} \sum_{q=1}^l |\alpha_q| m_q (m_q - 1). \quad \blacksquare$$

Next we examine $\operatorname{Lip}(f)$. From Eq. (2.7), we know that for $j \in [n]$,

$$\partial_j f(\alpha, X) = \frac{m\alpha}{2n^{m-1}} C_j^m + \frac{1}{2} \log \left(\frac{p}{1-p} \right) \leq \frac{m\alpha(n-1)_{m-1}}{2n^{m-1}} + \frac{1}{2} \left| \log \left(\frac{p}{1-p} \right) \right|,$$

where $(n)_m$ is the falling factorial,

$$(n)_m = n(n-1)(n-2) \cdots (n-m+1).$$

This upper bound comes from assigning each vertex weight to 1, thus obtaining the number of permutations of $m-1$ distinct variables from a total of $n-1$ variables. Thus for f a general subgraph counting function of the form (2.1), $\text{Lip}(f)$ may be bounded as

$$\text{Lip}(f) \leq \frac{1}{2} \left| \log \left(\frac{p}{1-p} \right) \right| + \frac{1}{2} \sum_{q=1}^l m_q |\alpha_q|.$$

With the necessary bounds on D and L_1 established for Theorem 3.1, we find an upper bound on L_2 , i.e. a bound on the quantity

$$\max_{X \neq Y \in \mathcal{C}_n} \frac{\|\nabla f(\alpha, X) - \nabla f(\alpha, Y)\|_1}{\|X - Y\|_1}.$$

To do so, we first bound the distance between two vertex weight configurations that differ at a single vertex in Lemma 3.3, then use the triangle inequality to bound the 1-norm distance between configurations that differ at any number of vertices.

Lemma 3.3. *Let $X, Y \in \mathcal{C}_n$ be two vectors that differ at a single coordinate i . Then for any $j \in [n]$,*

$$|\partial_j f(\alpha, X) - \partial_j f(\alpha, Y)| \leq \frac{1}{2} \sum_{q=1}^l \frac{|\alpha_q| m_q (m_q - 1)}{n}. \quad (3.2)$$

Proof. Suppose that $X_i = 1$ and $Y_i = 0$. Note that if $i = j$, then

$$|\partial_j f(\alpha, X) - \partial_j f(\alpha, Y)| = 0. \quad (3.3)$$

Assume that $i \neq j$, then from Eq. (2.7),

$$\begin{aligned} |\partial_j f(\alpha, X) - \partial_j f(\alpha, Y)| &\leq \frac{1}{2} \sum_{q=1}^l \frac{m_q |\alpha_q|}{n^{m_q-1}} \left| \sum_{j \neq i_1 \neq \dots \neq i_{m_q-1}} (X_{i_1} \cdots X_{i_{m_q-1}} - Y_{i_1} \cdots Y_{i_{m_q-1}}) \right| \\ &= \frac{1}{2} \sum_{q=1}^l \frac{m_q |\alpha_q|}{n^{m_q-1}} \left| (m_q - 1)! \sum_{\substack{P \subseteq [n] \setminus \{i, j\} \\ |P|=m_q-2}} \prod_{p \in P} X_p \right| \\ &\leq \frac{1}{2} \sum_{q=1}^l \frac{|\alpha_q| m_q! (n-2)(n-3) \cdots (n-m_q+1)}{n^{m_q-1} (m_q-2)!} \\ &\leq \frac{1}{2} \sum_{q=1}^l \frac{|\alpha_q| m_q (m_q - 1)}{n}. \end{aligned} \quad (3.4)$$

The equality from the first display line to the second comes from a simple counting argument. In the first sum, all terms will cancel except those where the index i appears since X and Y differ only at index i . When i appears, the product of Y terms will be zero since $Y_i = 0$. Thus the sum simplifies to counting terms with $m_q - 1$ factors from X where the index i appears and j does not. ■

Using Lemma 3.3, we may bound the 1-norm distance between the gradient vectors of any two vertex-weight configurations.

Lemma 3.4. *Let f be a subgraph counting function of the form (2.1) and $X, Y \in \overline{\mathcal{C}_n}$. Then*

$$\|\nabla f(\alpha, X) - \nabla f(\alpha, Y)\|_1 \leq C \|X - Y\|_1,$$

where

$$C = \frac{1}{2} \sum_{q=1}^l |\alpha_q| m_q (m_q - 1).$$

Proof. Suppose that X and Y differ in a single coordinate i . Holding all other coordinates fixed, ∇f is linear as a function of the i th coordinate. Then using Lemma 3.3,

$$\begin{aligned} \|\nabla f(\alpha, X) - \nabla f(\alpha, Y)\|_1 &= \sum_{j=1}^n |\partial_j f(\alpha, X) - \partial_j f(\alpha, Y)| \\ &\leq \sum_{q=1}^l \sum_{j=1}^n \frac{|\alpha_q| m_q(m_q - 1)}{2n} |X_i - Y_i| \\ &= \sum_{q=1}^l \frac{|\alpha_q| m_q(m_q - 1)}{2} |X_i - Y_i|. \end{aligned}$$

By the triangle inequality, we extend this to vectors that differ in more than one coordinate. The desired result is obtained. ■

4. Main results

With the bounds in Theorem 3.1 established in Lemmas 3.2, 3.3, and 3.4, we are ready to formulate our first main result Theorem 4.1. Fix a subgraph counting function f of the form in Eq. (2.1). Let C_α be the constant

$$C_\alpha = \max \left\{ 2, \frac{1}{2} \left| \log \left(\frac{p}{1-p} \right) \right| + \frac{1}{2} \sum_{q=1}^l |\alpha_q| m_q(m_q - 1) \right\}. \quad (4.1)$$

Furthermore, let $\mathcal{X}_f \subseteq [0, 1]^n$ be defined such that

$$\mathcal{X}_f = \left\{ X \in [0, 1]^n : \left\| X - \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} \right\|_1 \leq 5000 C_\alpha^2 n^{7/8} \right\}. \quad (4.2)$$

Theorem 3.1 gives that vertex-weighted exponential random graphs behave like mixture models. Additionally, the majority of the weight in this mixture model lies on configurations that are almost solutions to the vector equation

$$X = \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2}. \quad (4.3)$$

Theorem 4.1. Let f be a subgraph counting function where the subgraphs are counted as weighted, normalized products of vertex weights as in Eq. (2.1). Then there exists a measure ρ on $[0, 1]^n$ dependent on n and f such that X_n^f is a $(\rho, 80C_\alpha^{1/4} n^{-1/8})$ -mixture such that

$$\rho(\mathcal{X}_f) \geq 1 - 80C_\alpha^{1/4} n^{-1/8}.$$

Proof. By Lemmas 3.2, 3.3, and 3.4, $D \leq C_\alpha \sqrt{n}$, $L_1 \leq C_\alpha$, and $L_2 \leq C_\alpha$, where C_α is defined in Eq. (4.1). Substituting these bounds into Theorem 3.1, the result follows. ■

We will use this preceding theorem to show in Theorem 4.4 that if $X \in \mathcal{X}_f$, then there is a block vector X^* with a small number of communities such that X is close in 1-norm distance to X^* . Toward this end, we will require some statements concerning the preservation of length under orthogonal projections.

Lemma 4.2 (Theorem 2.1 in [19]). Let $0 < \delta < 1$, let $d, k > 0$ be positive integers, let $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^k$ be an orthogonal projection into a uniformly random k dimensional subspace, and let $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ be defined as $g(v) = \sqrt{\frac{d}{k}} \pi(v)$. Then for any vector $v \in \mathbb{R}^d$,

$$\mathbb{P}((1 - \delta) \|v\|_2^2 \leq \|g(v)\|_2^2 \leq (1 + \delta) \|v\|_2^2) \geq 1 - 2e^{-k(\delta^2/2 - \delta^3/3)/2}.$$

Lemma 4.3 (Lemma 30 in [13]). Let $0 < \delta < 1$, d and k be positive integers, and $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ be a linear transformation. If $u, v \in \mathbb{R}^d$ are two vectors of norm smaller than 1 such that

$$(1 - \delta) \|u \pm v\|_2^2 \leq \|g(u \pm v)\|_2^2 \leq (1 + \delta) \|u \pm v\|_2^2,$$

then

$$|\langle g(u), g(v) \rangle - \langle u, v \rangle| \leq 2\delta.$$

Theorem 4.4. Let $0 < \delta < 1$ and let f be a subgraph counting function of the form (2.1). Then there exists a constant $C_\delta > 0$, independent of n , such that for any $X \in \mathcal{X}_f$, there exists a block vector X^* with no more than C_δ communities such that

$$\|X - X^*\|_1 \leq \delta n + 5000C_\alpha^2 n^{7/8}.$$

Proof. Let f be a subgraph counting function of the form defined in Eq. (2.1) and let $X \in \mathcal{X}_f$. That is,

$$f(\alpha, X) = \log \left(\frac{p}{1-p} \right) \|X\|_1 + \sum_{q=1}^l \frac{\alpha_q}{n^{m_q-1}} \sum_{i_1 \neq \dots \neq i_{m_q}} X_{i_1} \cdots X_{i_{m_q}}.$$

As in Lemma 3.2, we may write

$$\nabla f(\alpha, X) = \frac{1}{2} \log \left(\frac{p}{1-p} \right) \mathbf{1}_n + \frac{1}{2} \sum_{q=1}^l \frac{m_q \alpha_q}{n^{m_q-1}} \mathbf{C}^{m_q-1} \mathbf{X} (J - (m_q - 1)I),$$

and the components of the gradient may be written as inner products such that for $j \in [n]$,

$$\partial_j f(\alpha, X) = \frac{1}{2} \log \left(\frac{p}{1-p} \right) + \frac{1}{2} \sum_{q=1}^l m_q \alpha_q \sqrt{1 - \frac{1}{n} + \frac{(2 - m_q)^2}{n}} \langle u^q, v_j^q \rangle,$$

where

$$u^q = \frac{1}{n^{m_q-2} \sqrt{n}} \mathbf{C}^{m_q-1} \mathbf{X}$$

and

$$v_j^q = \frac{1}{\sqrt{n-1+(2-m_q)^2}} (1, 1, \dots, 2-m_q, \dots, 1, 1),$$

with the entry $2-m_q$ in the j th position. Note that v_j^q is a multiple of the j th column of the matrix $J - (m_q - 1)I$. For each $q = 1, \dots, l$ and $j = 1, \dots, n$, the construction yields $\|u^q\|_2^2, \|v_j^q\|_2^2 \leq 1$.

Let $0 < \delta < 1$, $k = \left\lceil 2 \log(1/\delta) (\delta^2/2 - \delta^3/3)^{-1} \right\rceil$, $U \subseteq \mathbb{R}^n$ a uniformly random subspace of dimension k , and denote by $\pi : \mathbb{R}^n \rightarrow U$ an orthogonal projection from \mathbb{R}^n into U . Take $g : \mathbb{R}^n \rightarrow U$ as $g(v) = \sqrt{\frac{n}{k}} \pi(v)$. Temporarily fix an arbitrary pair (j, q) . Define

$$\mathcal{B}_j^q = \{(1-\delta)\|x\|_2^2 \leq \|g(x)\|_2^2 \leq (1+\delta)\|x\|_2^2 : x \in \{u^q, v_j^q, u^q \pm v_j^q\}\}.$$

By Lemma 4.2,

$$\mathbb{P}(\mathcal{B}_j^q) \geq 1 - 8e^{-k(\delta^2/2 - \delta^3/3)/2}.$$

Under this event, both $g(u^q)$ and $g(v_j^q)$ lie in a ball of radius 2 around the origin, since u^q and v_j^q lie in the unit ball and $\delta < 1$. It is also true by Lemma 4.3 that

$$|\langle g(u^q), g(v_j^q) \rangle - \langle u^q, v_j^q \rangle| \leq 2\delta.$$

Let T be a δ -net of the ball of radius 2 about the origin in k dimensions. By Lemma 2.6 of [20], there exists such a net of size no larger than $(1 + 4/\delta)^{k+1}$. For each $q = 1, \dots, l$ and $j = 1, \dots, n$, define

$$\bar{u}^q = \arg \min_{w \in T} \|g(u^q) - w\|_2,$$

$$\bar{v}_j^q = \arg \min_{w \in T} \|g(v_j^q) - w\|_2.$$

Consider $\Delta \bar{u}^q = \bar{u}^q - g(u^q)$ and $\Delta \bar{v}_j^q = \bar{v}_j^q - g(v_j^q)$. Since under \mathcal{B}_j^q , $g(u^q)$ and $g(v_j^q)$ are in the ball of radius 2, both $\|\Delta \bar{u}^q\|_2, \|\Delta \bar{v}_j^q\|_2 \leq \delta$. This gives

$$\begin{aligned} \left| \langle \bar{u}^q, \bar{v}_j^q \rangle - \langle g(u^q), g(v_j^q) \rangle \right| &= \left| \langle \Delta \bar{u}^q + g(u^q), \bar{v}_j^q \rangle - \langle g(u^q), g(v_j^q) \rangle \right| \\ &= \left| \langle \Delta \bar{u}^q, \bar{v}_j^q \rangle + \langle g(u^q), \bar{v}_j^q - g(v_j^q) \rangle \right| \\ &= \left| \langle \Delta \bar{u}^q, \bar{v}_j^q \rangle + \langle g(u^q), \Delta \bar{v}_j^q \rangle \right| \\ &\leq 4\delta. \end{aligned}$$

By the triangle inequality,

$$\begin{aligned} \left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle u^q, v_j^q \rangle \right| &\leq \left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle g(u^q), g(v_j^q) \rangle \right| \\ &\quad + \left| \langle g(u^q), g(v_j^q) \rangle - \langle u^q, v_j^q \rangle \right| \\ &\leq 6\delta. \end{aligned}$$

For each $q = 1, \dots, l$ and $j = 1, \dots, n$, define $\overline{X_j^q} = \langle \overline{u^q}, \overline{v_j^q} \rangle$ and $X_j^q = \langle u^q, v_j^q \rangle$. Then $\overline{X^q}$ is a block vector with no more than $(1 + 4/\delta)^{2(k+1)}$ communities, and

$$\mathbb{E} \|X^q - \overline{X^q}\|_1 = \mathbb{E} \sum_{j=1}^n \left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle u^q, v_j^q \rangle \right| = \sum_{j=1}^n \mathbb{E} \left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle u^q, v_j^q \rangle \right|.$$

We will compute the expectation for each j by conditioning on the event \mathcal{B}_j^q . Under this event, as derived previously, the absolute difference of the inner products is bounded by 6δ . Under the complementary event, which occurs with exponentially small probability in δ and k , the distance is trivially bounded by 5. Thus

$$\begin{aligned} \mathbb{E} \left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle u^q, v_j^q \rangle \right| &= \mathbb{E} \left[\left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle u^q, v_j^q \rangle \right| \mid \mathcal{B}_j^q \right] \mathbb{P}(\mathcal{B}_j^q) \\ &\quad + \mathbb{E} \left[\left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle u^q, v_j^q \rangle \right| \mid \mathcal{B}_j^q \right] \mathbb{P}(\overline{\mathcal{B}_j^q}) \\ &\leq 6\delta + 40e^{-k(\delta^2/2 - \delta^3/3)/2} \\ &\leq 46\delta. \end{aligned}$$

It follows that

$$\mathbb{E} \|X^q - \overline{X^q}\|_1 = \sum_{j=1}^n \mathbb{E} \left| \langle \overline{u^q}, \overline{v_j^q} \rangle - \langle u^q, v_j^q \rangle \right| \leq 46\delta n.$$

This implies that for each q , there must exist a block vector $\overline{X^q}$ with no more than $(1 + 4/\delta)^{2(k+1)}$ communities such that

$$\|X^q - \overline{X^q}\|_1 \leq 46\delta n.$$

Letting $c_q(n) = \sqrt{1 - \frac{1}{n} + \frac{(2-m_q)^2}{n}}$ (note here that $c_q(n) \leq m_q - 1$) for ease of notation, we have

$$\left\| \frac{1}{2} m_q \alpha_q c_q(n) X^q - \frac{1}{2} m_q \alpha_q c_q(n) \overline{X^q} \right\|_1 \leq 23 m_q |\alpha_q| c_q(n) \delta n.$$

Define \overline{X} as a weighted sum of the vectors $\overline{X^q}$ for $q = 1, \dots, l$ with the all ones vector,

$$\overline{X} = \frac{1}{2} \log \left(\frac{p}{1-p} \right) \mathbf{1}_n + \frac{1}{2} \sum_{q=1}^l m_q \alpha_q c_q(n) \overline{X^q}.$$

Then \overline{X} has a number of communities exponentially dependent on δ and l , however, most importantly, independent of n . The triangle inequality implies that

$$\|\nabla f(\alpha, X) - \overline{X}\|_1 \leq 23 \left(\sum_{q=1}^l m_q |\alpha_q| c_q(n) \right) \delta n.$$

Since \tanh is contracting,

$$\left\| \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} - \frac{\mathbf{1}_n + \tanh(\overline{X})}{2} \right\|_1 \leq \frac{23}{2} \left(\sum_{q=1}^l m_q |\alpha_q| c_q(n) \right) \delta n.$$

By assumption, $X \in \mathcal{X}_f$ and by [Theorem 4.1](#),

$$\left\| X - \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} \right\|_1 \leq 5000 C_\alpha^2 n^{7/8}.$$

Letting $X^* = \frac{\mathbf{1}_n + \tanh(\overline{X})}{2}$, the desired conclusion is obtained,

$$\|X - X^*\|_1 \leq \frac{23}{2} \left(\sum_{q=1}^l m_q |\alpha_q| c_q(n) \right) \delta n + 5000 C_\alpha^2 n^{7/8}. \quad \blacksquare$$

Remark 1. Note that the size of the δ -net constructed in the preceding proof affects the final high-probability result. The block vector \bar{X}^q that approximates X^q consists of no more than $(1 + 4/\delta)^{2(k+1)}$ communities. Since k is itself a function of δ , this is purely dependent on δ . It follows that the number of communities of \bar{X} , which is a weighted sum of the vectors \bar{X}^q for $q = 1, \dots, l$ with the all ones vector, depends on both δ and l . Therefore X^* , the approximation block vector for $X \in \mathcal{X}_f$, being a function of \bar{X} , depends on the parameters δ and l , but not on n , as well. We denote the number of communities of X^* implicitly by C_δ in the statement of [Theorem 4.4](#), as l is fixed by the subgraph counting function f defined in [Eq. \(2.1\)](#).

[Theorems 4.1](#) and [4.4](#) together yield the following corollary relating exponential random graphs to mixtures of block vectors.

Corollary 4.5. Let $\alpha_1, \dots, \alpha_l$ be real constants and $\delta > 0$. Then there exists a $C_\delta > 0$ such that for every $n \in \mathbb{N}$, there exists a measure ρ supported on block vectors with at most C_δ communities such that if X_n^f is a random vertex configuration with distribution defined in [Eq. \(1.4\)](#) where the Hamiltonian is as defined in [Eq. \(2.1\)](#), then there is a coupling between X_n^f and $G(n, \rho)$ that satisfies

$$\mathbb{E} \|X_n^f - G(n, \rho)\|_1 \leq \delta n.$$

We now turn to some results concerning subgraph counting functions with positive weights. Define $\varphi_\alpha : [0, 1] \rightarrow \mathbb{R}$ by

$$\varphi_\alpha(x) = \frac{1 + \tanh\left(\frac{1}{2} \log\left(\frac{p}{1-p}\right) + \frac{1}{2} \sum_{q=1}^l m_q \alpha_q \prod_{i=1}^{m_q-1} \left(1 - \frac{i}{n}\right) x^{m_q-1}\right)}{2}.$$

Note that φ is equal to the entry of the constant vector for the fixed point [Eq. \(4.3\)](#). If φ_α has a unique fixed point x , define D_α as

$$D_\alpha = \sup_{\substack{y \in [0, 1] \\ y \neq x}} \frac{|\varphi_\alpha(y) - x|}{|y - x|}.$$

Lemma 4.6. There exists an $x \in (0, 1)$ such that $x = \varphi_\alpha(x)$. Now assume that φ_α is increasing. If the solution x is unique and $\varphi'_\alpha(x) < 1$, then $D_\alpha < 1$.

Proof. Since $\varphi_\alpha : [0, 1] \rightarrow (0, 1)$ is continuous, there exists an $x \in (0, 1)$ such that $x = \varphi_\alpha(x)$. Now assume that φ_α is increasing, x is a unique fixed point and $\varphi'_\alpha(x) < 1$. Consider the function $h(y) = \varphi_\alpha(y) - x$. For $y < x$, $\varphi_\alpha(y) \leq x$ and

$$0 \geq h(y) = \varphi_\alpha(y) - x > y - x.$$

Hence

$$0 \leq \frac{\varphi_\alpha(y) - x}{y - x} < 1.$$

On the other hand, for $y > x$, $\varphi(y) \geq x$ and

$$0 \leq h(y) = \varphi_\alpha(y) - x < y - x.$$

Therefore again

$$0 \leq \frac{\varphi_\alpha(y) - x}{y - x} < 1.$$

Combining these inequalities for $y < x$ and $y > x$ with the fact that $\varphi'_\alpha(x) < 1$, it follows that $D_\alpha < 1$. Note also that if φ_α is not constant, then $D_\alpha > 0$. ■

Theorem 4.7. Let n be a positive integer. Let f be a subgraph counting function of the form [\(2.1\)](#). Assume that $\alpha_q \geq 0$ for all $q \geq 1$, $x = \varphi_\alpha(x)$ has a unique solution x , and $D_\alpha < 1$. Then for any $X \in \mathcal{X}_f$ and any $0 < \varepsilon < 1$,

$$\|X - x\mathbf{1}_n\|_1 \leq \varepsilon n + 10000 C_\alpha^2 \varepsilon^{\frac{\log C_\alpha}{\log D_\alpha}} n^{7/8}.$$

Furthermore, for any constants C_α and D_α , there exist constants $c > 0$ and $0 < \eta < 1/8$ such that

$$\|X - x\mathbf{1}_n\|_1 \leq cn^{1-\eta}.$$

Proof. It will first be shown that if the weights α_q are positive for $q \geq 1$ and x is the unique fixed point of φ_α , then the fixed point [Eq. \(4.3\)](#) has a unique solution $X = x\mathbf{1}_n$. Let X be a solution to this vector equation. For ease of notation let a and b be defined as $a = \min X$ and $b = \max X$. Since the weights are positive,

$$\partial f(\alpha, X) \leq \frac{1}{2} \left(\log\left(\frac{p}{1-p}\right) + \sum_{q=1}^l m_q \alpha_q \prod_{i=1}^{m_q-1} \left(1 - \frac{i}{n}\right) b^{m_q-1} \right).$$

Taking the maximum of the right hand side and then the left hand side in Eq. (4.3), we have

$$b \leq \frac{1 + \tanh\left(\frac{1}{2}\left(\log\left(\frac{p}{1-p}\right) + \sum_{q=1}^l m_q \alpha_q \prod_{i=1}^{m_q-1} \left(1 - \frac{i}{n}\right) b^{m_q-1}\right)\right)}{2}.$$

Similarly for a , since the weights are positive, a parallel reasoning gives

$$a \geq \frac{1 + \tanh\left(\frac{1}{2}\left(\log\left(\frac{p}{1-p}\right) + \sum_{q=1}^l m_q \alpha_q \prod_{i=1}^{m_q-1} \left(1 - \frac{i}{n}\right) a^{m_q-1}\right)\right)}{2}.$$

By assumption, φ_α has a unique solution x . The preceding two inequalities thus give that $a \geq x$ and $b \leq x$. Since b is the maximum of the solution X and a is the minimum, this implies that $X = x\mathbf{1}_n$ is the unique solution to the fixed point Eq. (4.3).

Next we show that for all $X \in \mathcal{X}_f$, X is close to the constant solution $x\mathbf{1}_n$ in 1-norm distance. Let $X \in \mathcal{X}_f$ and $0 < \varepsilon < 1$. Define a sequence of functions $\{\varphi_i(x)\}_{i=1}^\infty$ such that $\varphi_1(x) = \varphi_\alpha(x)$ and $\varphi_{i+1}(x) = \varphi_\alpha(\varphi_i(x))$ for $i \geq 1$. For all $y \in [0, 1]$,

$$|\varphi_1(y) - x| \leq D_\alpha |y - x|.$$

Iterating this preceding inequality yields

$$|\varphi_i(y) - x| \leq D_\alpha^i |y - x|$$

for all $i \in \mathbb{N}$. Set $k = \lceil \log_{D_\alpha} \varepsilon \rceil$. Then for $i = k$ in particular,

$$|\varphi_k(y) - x| \leq D_\alpha^k |y - x| < \varepsilon. \quad (4.4)$$

Let $\Phi : \bar{\mathcal{C}}_n \rightarrow \bar{\mathcal{C}}_n$ be the function

$$\Phi(X) = \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2}.$$

Define a sequence of vectors in \mathbb{R}^n as $Y_0 = X$ and recursively $Y_{i+1} = \Phi(Y_i)$ for $i \geq 0$. Since $\alpha_q \geq 0$ for $q \geq 1$,

$$\begin{aligned} \min Y_1 &= \min \Phi(X) = \min \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} \\ &\geq \min \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, (\min X)\mathbf{1}_n))}{2} = \varphi_\alpha(\min X). \end{aligned}$$

Analogously,

$$\begin{aligned} \max Y_1 &= \max \Phi(X) = \max \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} \\ &\leq \max \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, (\max X)\mathbf{1}_n))}{2} = \varphi_\alpha(\max X). \end{aligned}$$

Iterating these inequalities for min and max gives

$$\varphi_k(\min X) \leq \min Y_k \leq \max Y_k \leq \varphi_k(\max X).$$

Since Inequality (4.4) holds for all $y \in [0, 1]$, it must hold for both $\min X$ and $\max X$, which says that

$$x - \varepsilon < \min Y_k \leq \max Y_k < x + \varepsilon.$$

In other words, all entries of the vector Y_k lie within ε of x in 1-norm. Hence

$$\|Y_k - x\mathbf{1}_n\|_1 < \varepsilon n.$$

We have established that Y_k is close to the constant solution $x\mathbf{1}_n$. In order to show that X is close to the constant solution, we need to show that X is close to Y_k . By Lemma 3.4, for any $U, V \in \bar{\mathcal{C}}_n$,

$$\|\Phi(U) - \Phi(V)\|_1 \leq C_\alpha \|U - V\|_1.$$

Specifically, for consecutive vectors Y_i and Y_{i-1} ,

$$\|Y_i - Y_{i-1}\|_1 = \|\Phi(Y_{i-1}) - \Phi(Y_{i-2})\|_1 \leq C_\alpha \|Y_{i-1} - Y_{i-2}\|_1.$$

Consecutive iterations yield

$$\|Y_i - Y_{i-1}\|_1 \leq C_\alpha^{i-1} \|Y_1 - Y_0\|_1 = C_\alpha^{i-1} \|\Phi(X) - X\|_1.$$

Using this bound, we can bound the distance between X and Y_k as

$$\|X - Y_k\|_1 = \left\| \sum_{i=1}^k (Y_i - Y_{i-1}) \right\|_1 \leq \sum_{i=1}^k \|Y_i - Y_{i-1}\|_1$$

$$\leq \sum_{i=0}^{k-1} C_{\alpha}^i \|\Phi(X) - X\|_1 \leq 2C_{\alpha}^{k-1} \|\Phi(X) - X\|_1.$$

Compiling the above results and applying [Theorem 4.1](#) allows us to bound the distance between X and $x\mathbf{1}_n$ as follows:

$$\begin{aligned} \|X - x\mathbf{1}_n\|_1 &\leq \|X - Y_k\|_1 + \|Y_k - x\mathbf{1}_n\|_1 \\ &< 2C_{\alpha}^{k-1} \|\Phi(X) - X\|_1 + \varepsilon n \\ &\leq 10000\varepsilon^{\frac{\log C_{\alpha}}{\log D_{\alpha}}} C_{\alpha}^2 n^{7/8} + \varepsilon n. \end{aligned} \quad (4.5)$$

The order conclusion on n is deduced by minimizing the last display over $0 < \varepsilon < 1$, which occurs when

$$\varepsilon = \left(\frac{\log \frac{1}{D_{\alpha}}}{10000C_{\alpha}^2 \log C_{\alpha}} n^{1/8} \right)^{\frac{\log D_{\alpha}}{\log \frac{C_{\alpha}}{D_{\alpha}}}}. \quad \blacksquare$$

Remark 2. Note that the exponent of n in the first term of Eq. (4.5) is

$$\frac{7}{8} + \frac{\log C_{\alpha}}{8 \log \frac{C_{\alpha}}{D_{\alpha}}} = \frac{7}{8} + \frac{\log C_{\alpha}}{8 (\log C_{\alpha} - \log D_{\alpha})},$$

and the exponent of n in the second term is

$$1 + \frac{\log D_{\alpha}}{8 \log \frac{C_{\alpha}}{D_{\alpha}}} = 1 + \frac{\log D_{\alpha}}{8 (\log C_{\alpha} - \log D_{\alpha})}.$$

For a fixed C_{α} , as D_{α} approaches 0, the order of both terms tend to $n^{7/8}$; while, as D_{α} limits to 1 from the left, the order of both terms tend to n . The rate at which the order approaches these values is dependent on C_{α} . Under the assumption that $C_{\alpha} \leq \frac{1}{D_{\alpha}}$, we will show explicitly that $\|X - x\mathbf{1}_n\|_1 = o(n)$. From the proof of [Theorem 4.7](#),

$$\|X - x\mathbf{1}_n\|_1 \leq 10000C_{\alpha}C_{\alpha}^k n^{7/8} + D_{\alpha}^k n$$

for any $k \in \mathbb{N}$. Let k be of the form $k = \lceil \log_{D_{\alpha}} n^{-m} \rceil$ where $m > 0$. Then $D_{\alpha}^k \leq n^{-m}$ and, since $C_{\alpha} \leq \frac{1}{D_{\alpha}}$,

$$C_{\alpha}^k \leq C_{\alpha} C_{\alpha}^{\log_{D_{\alpha}} n^{-m}} \leq C_{\alpha} n^m.$$

Therefore

$$\|X - x\mathbf{1}_n\|_1 \leq 10000C_{\alpha}^2 n^{m+7/8} + n^{1-m},$$

and minimizing this over m yields

$$\|X - x\mathbf{1}_n\|_1 \leq (10000C_{\alpha}^2 + 1) n^{15/16}.$$

Remark 3. The unique solution of $x = \varphi_{\alpha}(x)$ is close to the solution to the equation

$$x = \frac{1 + \tanh\left(\frac{1}{2} \left(\log\left(\frac{p}{1-p}\right) + \sum_{q=1}^l m_q \alpha_q x^{m_q-1} \right)\right)}{2},$$

thereby eliminating the n dependence asymptotically. This condition on the positive weights is commonly referred to as the *high temperature regime*. See [Fig. 1](#).

We now investigate a specific example for a family of subgraph counting functions. It demonstrates that the preceding theorem can give meaningful information with respect to sparse graphs and that the condition $C_{\alpha} \leq \frac{1}{D_{\alpha}}$ from [Remark 2](#) is achieved for some substantively interesting models.

Example 4.8. Consider the subgraph counting function

$$f(\alpha, X) = \log\left(\frac{p}{1-p}\right) \|X\|_1 + \frac{\alpha}{n} \sum_{i \neq j} X_i X_j,$$

where vertex weights take value 0 with probability $1-p$ and 1 with probability p and $\alpha \geq 0$.

For notational convenience denote $\log(p/(1-p))$ by γ . Consider $p = p(n) = n^{-d}$ for some $d > 0$ and $|\gamma|/200 \leq \alpha \leq |\gamma|/100$. Ignoring the second term in f , this resembles the $G(n, p)$ model of Gilbert [\[4\]](#). Hence the expected number of vertices will be at least of order np . With p of the preceding form, $\gamma \approx -\log n$ and $\alpha \approx \log n$. For n large enough, $\varphi_{\alpha}(x)$ has a unique fixed point; denote it by \bar{x} . Calculating D_{α} , we have

$$D_{\alpha} \leq \max_{x \in [0, 1]} |\varphi'_{\alpha}(x)|$$

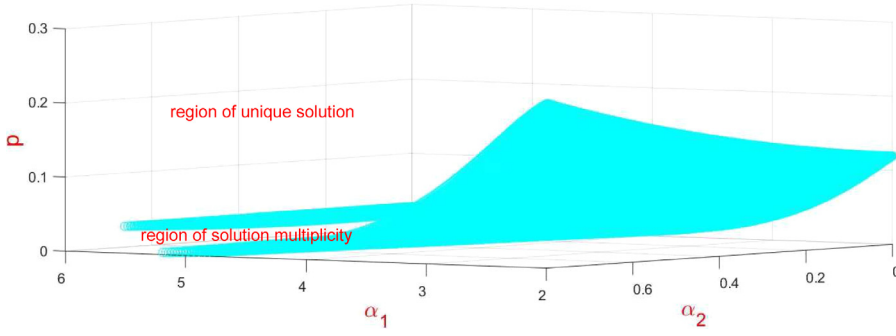


Fig. 1. The parameter space as separated by the region with unique vs. multiple solutions to $x = \varphi_\alpha(x)$. The asymptotic illustration is for the vertex-weighted edge-triangle model.

$$= \max_{x \in [0, 1]} \left| \frac{\alpha \left(1 - \frac{1}{n}\right)}{2 \cosh^2 \left(\frac{1}{2} \gamma + \alpha \left(1 - \frac{1}{n}\right) x\right)} \right|$$

$$\lesssim |\gamma| e^{-|\gamma|/2}.$$

Since $\alpha \rightarrow \infty$, for n sufficiently large $D_\alpha < 1$ and $|\log D_\alpha| \gtrsim |\gamma|$. This allows us to utilize [Theorem 4.7](#) to say that for all $X \in \mathcal{X}_f$,

$$\|X - \bar{x} \mathbf{1}_n\|_1 \leq 10000 \varepsilon^{\frac{\log C_\alpha}{\log D_\alpha}} C_\alpha^2 n^{7/8} + \varepsilon n. \quad (4.6)$$

From Eq. (4.1), $C_\alpha \approx |\gamma|$. Note that in the limit $|\gamma|^2 e^{-|\gamma|/2} \leq 1$, we conclude that the condition $C_\alpha \leq \frac{1}{D_\alpha}$ is satisfied. Examining the exponent on ε in the first term, we find

$$\left| \frac{\log C_\alpha}{\log D_\alpha} \right| \lesssim \frac{\log \log n}{d \log n}.$$

Since n is finite and fixed, let $\varepsilon = n^{-1}$, then

$$\varepsilon^{\frac{\log C_\alpha}{\log D_\alpha}} \approx (\log n)^{1/d}.$$

To yield a meaningful bound in Eq. (4.6), we want the error term to be smaller than the typical number of vertices. Ignoring lower order terms such as logarithms, we require

$$np \gtrsim n^{7/8},$$

which forces the condition that $p \gtrsim n^{-1/8}$. For p in this region, the majority of the probability mass in the model lies on sparse graphs. Furthermore, by [Corollary 4.5](#), there exists a constant \bar{p} and a coupling between $G(n, \bar{p})$ and X_n^f such that

$$\mathbb{E} \|X_n^f - G(n, \bar{p})\|_1 = o(np).$$

[Theorem 4.7](#) shows that for positive weights, any $X \in \mathcal{X}_f$ is at most $cn^{1-\eta}$ distance away from the fixed point of Φ for some constants c and $0 < \eta < 1/8$. This distance can be refined to $cn^{7/8}$ if one assumes that the weights α_q are sufficiently small (either positive or negative). Recall the function $\Phi : \bar{\mathcal{C}}_n \rightarrow \bar{\mathcal{C}}_n$ defined in the proof of this theorem,

$$\Phi(X) = \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2}.$$

Theorem 4.9. Let f be a subgraph counting function of the form (2.1) and let

$$J_\alpha = \frac{1}{4} \sum_{q=1}^l |\alpha_q| m_q (m_q - 1).$$

If $J_\alpha < 1$, then the constant solution X_c is the only solution to the fixed point equation $X = \Phi(X)$ and, for any $X \in \mathcal{X}_f$,

$$\|X - X_c\|_1 \leq \frac{5000 C_\alpha^2}{1 - J_\alpha} n^{7/8}.$$

Proof. By Lemma 3.4, for $X, Y \in \bar{\mathcal{C}}_n$,

$$\|\nabla f(\alpha, X) - \nabla f(\alpha, Y)\|_1 \leq \frac{1}{2} \sum_{q=1}^l |\alpha_q| m_q(m_q - 1) \|X - Y\|_1.$$

Therefore

$$\begin{aligned} \|\Phi(X) - \Phi(Y)\|_1 &= \left\| \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} - \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, Y))}{2} \right\| \\ &\leq \frac{1}{2} \|\nabla f(\alpha, X) - \nabla f(\alpha, Y)\|_1 \\ &\leq \frac{1}{4} \sum_{q=1}^l |\alpha_q| m_q(m_q - 1) \|X - Y\|_1 \\ &= J_\alpha \|X - Y\|_1. \end{aligned}$$

Since $J_\alpha < 1$, Φ is a contraction mapping, further implying that Φ has a unique fixed point: the constant solution $X_c = x\mathbf{1}_n$ provided by Lemma 4.6 and Theorem 4.7. Now let $X \in \mathcal{X}_f$. Then

$$\begin{aligned} \|X - X_c\|_1 &= \|X - \Phi(X) + \Phi(X) - \Phi(X_c) + \Phi(X_c) - X_c\|_1 \\ &\leq \|X - \Phi(X)\|_1 + \|\Phi(X) - \Phi(X_c)\|_1 + \|\Phi(X_c) - X_c\|_1 \\ &\leq \|X - \Phi(X)\|_1 + J_\alpha \|X - X_c\|_1. \end{aligned}$$

From the above inequality, one finds that

$$\|X - X_c\|_1 \leq \frac{\|X - \Phi(X)\|_1}{1 - J_\alpha} \leq \frac{5000C_\alpha^2}{1 - J_\alpha} n^{7/8}. \quad \blacksquare$$

5. Triangle case

Let us take a closer look at the special case of the subgraph counting function consisting solely of triangles

$$f(\alpha, X) = \frac{\alpha}{n^2} \sum_{i \neq k \neq j} X_i X_k X_j. \quad (5.1)$$

Eldan and Gross showed in Theorem 21 of [13] that for a subgraph counting function that counts triangles in simple graphs, there exists a constant $\alpha_0 < 0$ such that for all $\alpha < \alpha_0$ there is a non-constant solution in the form of a stochastic block matrix consisting of two communities. Furthermore, they determined the asymptotic tendency of the edge presence probability as $\alpha \rightarrow -\infty$.

Without loss of generality, take n even. (The odd case follows a similar line of reasoning.) Since we are, implicitly, placing equal probability on vertex weights being 0 and 1 (cf. Eq. (2.4)), some structure of symmetry is expected. We assume that there is a solution to the fixed point equation

$$X = \frac{\mathbf{1}_n + \tanh(\nabla f(\alpha, X))}{2} \quad (5.2)$$

in the form of a vector with the first $n/2$ entries taking value $a \in [0, 1]$ and second $n/2$ entries taking value $b \in [0, 1]$,

$$X = (a, a, \dots, a, b, b, \dots, b).$$

Recall the definition of the discrete derivative. Note that for the subgraph counting function defined above in Eq. (5.1), the discrete derivative at the j th coordinate may be written as

$$\partial_j f(\alpha, X) = \frac{3\alpha}{2n^2} \sum_{i \neq k \neq j} X_i X_k.$$

From here, we examine the possibilities for the product $X_i X_k$ in the sum. There are three types of terms in this sum since there are two different entries in the vector X . Working out the various combinations gives

$$\partial_j f(\alpha, X) = \frac{3\alpha}{n^2} \left(\left(\frac{n}{2} - 1 \right) a^2 + \left(\frac{n}{2} \right) b^2 + \left(\frac{n}{2} - 1 \right) \frac{n}{2} ab \right).$$

Writing this more compactly, it follows that if $X_j = a$, then

$$\partial_j f(\alpha, X) = \frac{3\alpha}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} (a + b)^2 - 2a^2 \right),$$

and similarly, if $X_j = b$, then

$$\partial_j f(\alpha, X) = \frac{3\alpha}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2}(a+b)^2 - 2b^2 \right).$$

From here we obtain a system of two equations in a and b that must be solved simultaneously:

$$a = \frac{1 + \tanh \left(\frac{3\alpha}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2}(a+b)^2 - 2a^2 \right) \right)}{2}$$

and

$$b = \frac{1 + \tanh \left(\frac{3\alpha}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2}(a+b)^2 - 2b^2 \right) \right)}{2}.$$

Instead of considering the previous two equations for $\alpha < 0$, for simplicity, we consider the equations

$$1 - 2a = \tanh \left(\frac{\alpha}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2}(a+b)^2 - 2a^2 \right) \right) \quad (5.3)$$

and

$$1 - 2b = \tanh \left(\frac{\alpha}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2}(a+b)^2 - 2b^2 \right) \right) \quad (5.4)$$

for $\alpha > 0$, where we assimilated the factor of 3 inside α .

Let n be an even integer such that $n > 4$. Fix $b = 1$ in Eq. (5.3). The left hand side of Eq. (5.3) is decreasing from 1 to -1 while the right hand side is increasing with values between 0 and 1. Thus the left side and right side intersect in a unique $a_1 \in (0, 1)$ that solves Eq. (5.3) for $b = 1$. Now let $b = 0$ in the same equation. By similar reasoning, there exists a unique $a_2 \in (0, 1)$ that solves Eq. (5.3) for $b = 0$ with $a_1 < a_2$. For each $b \in [0, 1]$, there is a unique a such that the pair (a, b) is a solution to Eq. (5.3). Let $g : [0, 1] \rightarrow [a_1, a_2]$ be the function that produces a unique a for each b , that is, $g(b) = a$ and the pair (a, b) is a solution to Eq. (5.3). Similarly, we can consider a fixed $a \in [0, 1]$ in Eq. (5.4) to obtain a unique $b \in (0, 1)$ such that the pair (a, b) is a solution to Eq. (5.4). This also yields a corresponding b_1 and b_2 when fixing $a = 1$ and $a = 0$, respectively. Denote by $h : [0, 1] \rightarrow [b_1, b_2]$ the function that produces a unique b for each $a \in [0, 1]$ such that $h(a) = b$ and (a, b) is a solution to Eq. (5.4). Both g and h are decreasing on their respective domains. Note that if the functions g and h intersect at some point (a, b) , then the point (a, b) is a solution to Eqs. (5.3) and (5.4) simultaneously; therefore, the pair (a, b) forms a solution to the fixed point equation.

Consider now the behavior of a_1 and a_2 as α diverges to infinity. We first show that a_2 tends to 0 as $\alpha \rightarrow \infty$ with a rate of decay controlled by a polynomial. This implies that a_1 also approaches 0 as α diverges. We also prove a corresponding result that controls the rate at which a_1 approaches 0, showing that the decay is bounded below by an exponential. Using the symmetry afforded by Eqs. (5.3) and (5.4), similar results hold for b_1 and b_2 .

For the proof of the following Lemma 5.1, we require the Lambert W function. The Lambert W function is a set of functions satisfying the functional equation

$$W(z)e^{W(z)} = z$$

for any $z \in \mathbb{C}$, and may be equivalently written as $W(ze^z) = z$. If we restrict our attention to the reals, there are two branches of this function, denoted by W_{-1} and W_0 , with the latter known as the principal branch. We utilize the branch W_0 in the following proof where $W_0 : [-1/e, \infty) \rightarrow [-1, \infty)$. The principal branch satisfies the inequality

$$\log x - \log \log x + \frac{\log \log x}{2 \log x} \leq W_0(x) \leq \log x - \log \log x + \frac{e}{e-1} \frac{\log \log x}{\log x} \quad (5.5)$$

for all $x \geq e$ as shown in [21]. For more information on the Lambert W function, see [22].

Lemma 5.1. Take α sufficiently large. Let a_2 be defined as above. Then

$$a_2 < \sqrt{\frac{n^2 \log \left(\frac{4\alpha}{n^2} \left(\frac{n}{2} - 1 \right) \right)}{4\alpha \left(\frac{n}{2} - 1 \right)}}. \quad (5.6)$$

Proof. Recall that a_2 is the unique solution to

$$1 - 2a_2 = \tanh \left(\frac{\alpha}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} - 2 \right) a_2^2 \right).$$

Let N be defined as

$$N = \frac{1}{2n^2} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} - 2 \right).$$

Then the preceding equation becomes

$$2a_2 = 1 - \tanh(N\alpha a_2^2).$$

Converting \tanh to exponentials yields

$$a_2 = \frac{1}{e^{2N\alpha a_2^2} + 1} < e^{-2N\alpha a_2^2}.$$

Letting $a = a_2^2$, we wish to solve the inequality

$$\sqrt{a}e^{2N\alpha a} < 1.$$

Squaring both sides and multiplying by $4N\alpha$ gives

$$4N\alpha a e^{4N\alpha a} < 4N\alpha$$

and, since $4N\alpha > e^{e/(e-1)}$ for α large eventually, we use the function W_0 to obtain

$$a < \frac{W_0(4N\alpha)}{4N\alpha} < \frac{\log(4N\alpha)}{4N\alpha},$$

where the final inequality comes from using the bound in Eq. (5.5). Converting back to a_2 and n , the desired Inequality (5.6) is derived. ■

We now show that the decay of a_1 to 0 is slower than a certain exponential.

Lemma 5.2. Take α sufficiently large. Let a_1 be defined as above. Then $a_1 > e^{-c\alpha}$ for all $c > \frac{n-2}{n}$.

Proof. We prove this result in a similar way to the previous Lemma 5.1. Let c be such that

$$c > \frac{n-2}{n}.$$

Recall that a_1 is the unique solution to

$$1 - 2a = \tanh\left(\frac{\alpha}{2n^2}\left(\frac{n}{2} - 1\right)\left(\frac{n}{2}(a+1)^2 - 2a^2\right)\right).$$

We will show that if $a = e^{-c\alpha}$ in the previous display, then the left hand side must be greater than the right hand side. This implies that $a_1 > e^{-c\alpha}$. Consider the limit

$$\lim_{\alpha \rightarrow \infty} \frac{2e^{-c\alpha}}{1 - \tanh\left(\frac{\alpha}{2n^2}\left(\frac{n}{2} - 1\right)\left(\frac{n}{2}(e^{-c\alpha} + 1)^2 - 2e^{-2c\alpha}\right)\right)}$$

and let $y(\alpha)$ denote the expression inside of \tanh . Note that $y(\alpha) \leq \alpha(n-2)/(2n)$. Using L'Hôpital's rule we find

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \frac{2e^{-c\alpha}}{1 - \tanh(y(\alpha))} &= \lim_{\alpha \rightarrow \infty} 2c \left(\frac{dy}{d\alpha}\right)^{-1} e^{-c\alpha} \cosh^2(y(\alpha)) \\ &= \lim_{\alpha \rightarrow \infty} \frac{c}{2} \left(\frac{dy}{d\alpha}\right)^{-1} e^{-c\alpha} (e^{-2y(\alpha)} + e^{2y(\alpha)} + 2) \\ &= 0. \end{aligned}$$

The last line can be justified as follows. The expression $dy/d\alpha$ limits to a nonzero constant. Also, since $c > (n-2)/n$ and $y(\alpha) \leq \alpha(n-2)/(2n)$, it follows that $c\alpha > 2y(\alpha)$. This implies that $a_1 > e^{-c\alpha}$. ■

Lemma 5.1 proves that as $\alpha \rightarrow \infty$, both a_1 and a_2 approach zero since $a_1 < a_2$. Lemma 5.2 gives that the speed with which a_1 tends to zero is bounded below. Taking the limit of n in Eqs. (5.3) and (5.4) yields

$$1 - 2a = \tanh\left(\frac{\alpha(a+b)^2}{8}\right)$$

and

$$1 - 2b = \tanh\left(\frac{\alpha(a+b)^2}{8}\right)$$

by the continuity of \tanh . Thus, for n large enough, a two-block solution to this system is in fact given by a constant solution as $a = b$. By Lemma 33 of [13], the equation

$$1 - 2a = \tanh\left(\frac{\alpha a^2}{2}\right)$$

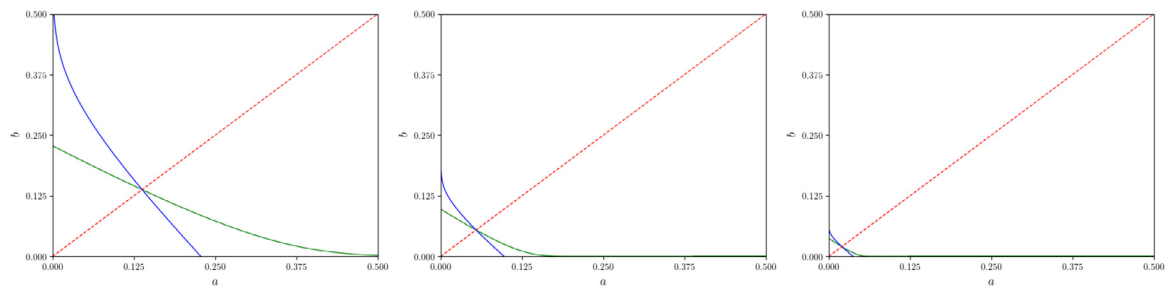


Fig. 2. These figures simulate the movement of the solution to the system of Eqs. (5.3) and (5.4) toward the origin.

has a unique solution $a \in (0, 1)$ for all $\alpha \in \mathbb{R}$. Therefore for this special asymptotic scenario of vertex-weighted graphs, unlike the simple graphs counterpart, the only two-block solution to the fixed point Eq. (5.2) is actually a particular constant vector, which approaches the all zeros vector as $\alpha \rightarrow \infty$ by Lemma 5.1.

Fig. 2 simulates the movement of the solution toward the origin as α diverges to ∞ . The value of n was taken to be sufficiently large and even. From left to right in Fig. 2, the values of α are 100, 1000, and 10000.

CRedit authorship contribution statement

Ryan DeMuse: Conception and design of study, Analysis and/or interpretation of data, Drafting the manuscript.
Mei Yin: Conception and design of study, Analysis and/or interpretation of data, Drafting the manuscript, Revising the manuscript critically for important intellectual content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Mei Yin's research was supported in part by the University of Denver's, United States Faculty Research Fund. She thanks Patrick Shipman for helpful conversations. The authors would also like to thank Ronen Eldan and Renan Gross for enlightening discussions. All authors approved the version of the manuscript to be published.

References

- [1] S.E. Fienberg, Introduction to papers on the modeling and analysis of network data, *Ann. Appl. Statist.* 4 (2010) 1–4.
- [2] S.E. Fienberg, Introduction to papers on the modeling and analysis of network data II, *Ann. Appl. Statist.* 4 (2010) 533–534.
- [3] M. Jackson, The past and future of network analysis in economics, in: Y. Bramoullé, A. Galeotti, B. Rogers (Eds.), *The Oxford Handbook on the Economics of Networks*, Oxford University Press, Oxford, UK, 2016.
- [4] E.N. Gilbert, Random graphs, *Ann. Math. Stat.* 30 (1959) 1141–1144.
- [5] S.J. Cranmer, B.A. Desmarais, Inferential network analysis with exponential random graph models, *Pol. Anal.* 19 (2011) 66–86.
- [6] D. Krioukov, Clustering implies geometry in networks, *Phys. Rev. Lett.* 116 (2016) 208302.
- [7] R. DeMuse, T. Easlick, M. Yin, Mixing time of vertex-weighted exponential random graphs, *J. Comput. Appl. Math.* 362 (2019) 443–459.
- [8] S. Bhamidi, G. Bresler, A. Sly, Mixing time of exponential random graphs, *Ann. Appl. Probab.* 21 (2011) 2146–2170.
- [9] S. Bhamidi, S. Chakraborty, S. Cranmer, B. Desmarais, Weighted exponential random graph models: Scope and large network limits, *J. Stat. Phys.* 173 (2018) 704–735.
- [10] S. Chatterjee, Estimation in spin glasses: A first step, *Ann. Statist.* 35 (2007) 1931–1946.
- [11] B. Bhattacharya, S. Mukherjee, Inference in Ising models, *Bernoulli* 24 (2018) 493–525.
- [12] P. Ghosal, S. Mukherjee, Joint estimation of parameters in Ising model, 2018, arXiv:1801.06570.
- [13] R. Eldan, R. Gross, Exponential random graphs behave like mixtures of stochastic block models, *Ann. Appl. Probab.* 28 (2018) 3698–3735.
- [14] S. Chatterjee, P. Diaconis, Estimating and understanding exponential random graph models, *Ann. Statist.* 41 (2013) 2428–2461.
- [15] D. Aldous, Representations for partially exchangeable arrays of random variables, *J. Multivariate Anal.* 11 (1981) 581–598.
- [16] L. Lovász, *Large Networks and Graph Limits*, American Mathematical Society, Providence, USA, 2012.
- [17] M. Yin, Phase transitions in edge-weighted exponential random graphs, 2016, arXiv:1607.04084.
- [18] R. Eldan, R. Gross, Decomposition of mean-field Gibbs distributions into product measures, *Electron. J. Probab.* 23 (2018) 1–24.
- [19] S. Dasgupta, A. Gupta, An elementary proof of a theorem of Johnson and Lindenstrauss, *Random Struct. Algorithms* 22 (2003) 60–65.
- [20] V.D. Milman, G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces*, Springer-Verlag New York Inc., New York, USA, 1986.
- [21] A. Hoorfar, M. Hassani, Inequalities on the Lambert W function and hyperpower function, *J. Inequal. Pure Appl. Math.* 9 (2008) 1–5.
- [22] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth, On the Lambert W function, *Adv. Comput. Math.* 5 (1996) 329–359.