Logconcave Random Graphs

Alan Frieze*
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA15213
alan@random.math.cmu.edu

Santosh Vempala†
School of Computer Science
Georgia Tech
Atlanta GA 30332
vempala@cc.gatech.edu

Juan Vera
Department of Management Sciences
University of Waterloo
Waterloo, ON N2L 3G1, Canada
jvera@uwaterloo.ca

Submitted: Feb 15, 2010; Accepted: Jun 23, 2010; Published: Aug 9, 2010
Mathematics Subject Classification: 05C80,52A23

Abstract

We propose the following model of a random graph on n vertices. Let F be a distribution in $R_+^{n(n-1)/2}$ with a coordinate for every pair ij with $1 \leqslant i, j \leqslant n$. Then $G_{F,p}$ is the distribution on graphs with n vertices obtained by picking a random point X from F and defining a graph on n vertices whose edges are pairs ij for which $X_{ij} \leqslant p$. The standard Erdős-Rényi model is the special case when F is uniform on the 0-1 unit cube. We examine basic properties such as the connectivity threshold for quite general distributions. We also consider cases where the X_{ij} are the edge weights in some random instance of a combinatorial optimization problem. By choosing suitable distributions, we can capture random graphs with interesting properties such as triangle-free random graphs and weighted random graphs with bounded total weight.

1 Introduction

Probabilistic combinatorics is today a thriving field bridging the classical area of probability with modern developments in combinatorics. The theory of random graphs, pioneered

*Research supported in part by NSF award CCF-0502793.
†Supported in part by NSF award CCF-0721503.

by Erdős-Rényi [7] has given us numerous insights, surprises and techniques and has been used to count, to establish structural properties and to analyze algorithms.

In the standard unweighted model $G_{n,p}$, each pair of vertices ij of an n-vertex graph is independently declared to be an edge with probability p. Equivalently, one picks a random number X_{ij} for each ij in the interval $[0,1]$, i.e., a point in the unit cube, and defines as edges all pairs for which $X_{ij} \leqslant p$. To get a weighted graph, we avoid the thresholding step.

In this paper, we propose the following extension to the standard model. We have a distribution F in $\mathbb{R}_+^N$ where $N = n(n-1)/2$ allows us a coordinate for every pair of vertices. A random point X from F assigns a non-negative real number to each pair of vertices and is thus a random weighted graph. The random graph $G_{F,p}$ is obtained by picking a random point X according to F and applying a p-threshold to determine edges, i.e., the edge set $E_{F,p} = \{ij : X_{ij} \leqslant p\}$. It is clear that this generalizes the standard model $G_{n,p}$ which is the special case when F is uniform over a cube.

In the special case where $F(x) = 1_{x\in K}$ is the indicator function for some convex subset K of $\mathbb{R}_+^N$ we use the notation $G_{K,p}$ and $E_{K,p}$. Thus to obtain $G_{K,p}$ we let X be a random point in K. It includes the restriction of any L_p ball to the positive orthant. The case of the simplex

$$K = \{x \in \mathbb{R}^N : \forall e, x_e \geqslant 0, \sum_e \alpha_i x_e \leqslant L\}$$

for some set of coefficients α appears quite interesting by itself and we treat it in detail in Section 4. In the weighted graph setting, it corresponds to a random graph with a bound on the total edge weight. In general, F could be any distribution, but we will consider a further generalization of the cube and simplex, namely, when F has a logconcave density f. We call this a logconcave distribution. A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *logconcave* if for any two points $x, y \in \mathbb{R}^n$ and any $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \geqslant f(x)^\lambda f(y)^{1-\lambda},$$

i.e., $\ln f$ is concave. We discuss the motivation presently along with a precise definition.

The model appears to be considerably more general than $G_{n,p}$. Nevertheless, can we recover interesting general properties including threshold phenomena?

The average case analysis of algorithms for NP-hard problems was pioneered by Karp [13] and in the context of graph algorithms, the theory of random graphs has played a crucial role (see [9] for a somewhat out-dated survey). To improve on this analysis, we need tractable distributions that provide a closer bridge between average case and worst-case. We expect the distributions described here to be a significant platform for future research.

We end this section with a description of the model and a summary of our main results.

1.1 The model and motivation

We consider logconcave density functions f whose support lies in the positive orthant. For such a density f, let $\sigma_e^2(f) = \mathsf{E}_{X\sim f}(X_e^2)$ denote the second moment along each axis e. We just use σ_e when f is fixed and simply σ when the second moment is the same along every axis. We will also use $\sigma_{\min} = \sigma_{\min}(f) := \min \sigma_e(f)$ and $\sigma_{\max} = \sigma_{\max}(f) := \max \sigma_e(f)$. We also restrict f to be downmonotone, i.e., for any $x, y \in \mathbb{R}^N$ such that $x \leqslant y$ coordinate-wise, we have $f(x) \geqslant f(y)$. We denote by F the distribution obtained from f. Given such an F, we generate a random graph $G_{F,p}$ by picking a point X from F and including as edges all pairs ij for which $X_{ij} \leqslant p$.

We now give some rationale for the model. First, it is clear that we need the distribution to have some "spread" in order to avoid focusing on essentially a single graph. Fixing only the standard deviations along the axes allows highly restricted distributions, e.g., the line from the origin to the vector of all 1's. To avoid this, we require that the density is down-monotone. When f corresponds to the uniform density over a convex body K, this means that when $x \in K$, the box with 0 and x at opposite corners is also in K. It also implies that f can be viewed as the restriction to the positive orthant of a 1-unconditional distribution for which the density $f(x_1, \ldots, x_N)$ stays fixed when we reflect on any subset of axes, i.e., negating subset of coordinates keeps f the same. Such distributions include, e.g., the L_p ball for any p but also much less symmetric sets, e.g., the uniform distribution over any down-monotone convex body.

To generalize further, we allow logconcave densities. Allowing arbitrary densities with down-monotone supports would lead to the same problem as before, and we need a concavity condition on the density. Logconcavity is particularly suitable since products and marginals of logconcave functions remain logconcave. So, e.g., the distribution restricted to a particular pair ij is also logconcave.

The model departs from the standard $G_{n,p}$ model by allowing for dependencies, i.e., the joint distribution for a subset of coordinates is not a product distribution and could be quite far from any product distribution. Moreover the coordinates are neither positively correlated nor negatively correlated in general. Nevertheless, there is a significant literature on the geometry and concentration of logconcave distributions and we leverage these ideas in our proofs.

We note briefly that sampling logconcave distributions efficiently requires only a function oracle, i.e., for any point x, we can compute a function proportional to the density at x (see e.g., [17]).

Following our presentation for general monotone logconcave densities, we focus our attention on an interesting special case: a simplex in the positive orthant with unequal edge lengths, i.e., there is a single defining constraint of the form $a \cdot X \leqslant 1$, $a \geqslant 0$, in addition to the nonnegativity constraints. This can be interpreted as a budget constraint for a random graph.

2 Results

2.1 Random graphs from logconcave densities.

We prove asymptotic results that require $n \to \infty$. As such we we need to deal with a sequence of distributions F_n, but for notational convenience we always refer to F.

Our first result estimates the point at which $G_{F,p}$ is connected in general in terms of n and σ, the standard deviation in any direction. Our main result is that after fixing the second moments along every axis, the threshold for connectivity can be narrowed down to within an $O(\log n)$ factor.

Theorem 2.1 *Let F be distribution in the positive orthant with a down-monotone log-concave density. Then there exist absolute constants $0 < c_1 < c_2$ such that*

$$\lim_{n\to\infty} \mathsf{P}(G_{F,p} \text{ is connected}) = \begin{cases} 0 & p < \dfrac{c_1\sigma_{\min}}{n} \\ 1 & p > \dfrac{c_2\sigma_{\max}\ln n}{n} \end{cases}$$

F being so general makes this theorem quite difficult to prove. It requires several results that are trivial in $G_{n,p}$.

The reader will notice the disparity between the upper and lower bound.

Conjecture 2.2 [1] *Let F be as in Theorem 2.1. Then there exists a constant c_0 such that if $p < c_0\sigma_{\min}\ln n/n$ then* **whp**[2] *$G_{F,p}$ has isolated vertices.*

Having proven Theorem 2.1 it becomes easy to prove other similar results.

Theorem 2.3 *Let F be as in Theorem 2.1. Then there exist absolute constants $c_3 < c_4$ such that*

$$\lim_{\substack{n\to\infty \\ n \text{ even}}} \mathsf{P}(G_{F,p} \text{ has a perfect matching}) = \begin{cases} 0 & p < \dfrac{c_3\sigma_{\min}}{n} \\ 1 & p > \dfrac{c_4\sigma_{\max}\ln n}{n} \end{cases}$$

Finally, for this section, we mention a result on Hamilton cycles that can be obtained quite simply from a result of Hefetz, Krivelevich and Szabó [10].

Theorem 2.4 *Let F be as in Theorem 2.1. Then there exists an absolute constant c_6 such that if*

$$p \geqslant c_6\sigma_{\max}\frac{\ln n}{n}\cdot\frac{\ln\ln\ln n}{\ln\ln\ln\ln n}$$

then $G_{F,p}$ is Hamiltonian **whp**.

[1] In an early version of this paper, an abstract of which appeared in FOCS 2008, we incorrectly claimed this conjecture as a theorem.

[2] A sequence of events $\mathcal{E}_n$ is said to occur *with high probability* **whp**, if $\lim_{n\to\infty}\mathsf{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$

2.2 Random Graphs from a Simplex

We now turn to a specific class of convex bodies K for which we can prove fairly tight results. We consider the special case where X is chosen uniformly at random from the simplex

$$\Sigma = \Sigma_{n,L,\alpha} = \left\{ X \in \mathbb{R}_+^N : \sum_{e \in E_n} \alpha_e X_e \leqslant L \right\}.$$

Here $N = \binom{n}{2}$ and $E_n = \binom{[n]}{2}$ and L is a positive real number and $\alpha_e > 0$ for $e \in E_n$.

We observe first that $G_{\Sigma_{n,L,\alpha},p}$ and $G_{\Sigma_{n,N,\alpha N/L},p}$ have the same distribution and so we assume, unless otherwise stated, that $L = N$. The special case where $\alpha = \mathbf{1}$ (i.e. $\alpha_e = 1$ for $e \in E_n$) will be easier than the general case. We will see that in this case $G_{\Sigma,p}$ behaves a lot like $G_{n,p}$.

Although it is convenient to phrase our theorems under the assumption that $L = N$, we will not always assume that $L = N$ in the main body of our proofs. It is informative to keep the L in some places, in which case we will use the notation Σ_L for the simplex. In general, when discussing the simplex case, we will use Σ for the simplex. On the other hand, we will if necessary subscript Σ by one or more of the parameters α, L, p if we need to stress their values.

We will not be able to handle completely general α. We will restrict our attention to the case where

$$\frac{1}{M} \leqslant \alpha_e \leqslant M \qquad \textit{for } e \in E_n \tag{1}$$

where $M = M(n)$. An α that satisfies (1) will be called *M-bounded*.

This may seem restrictive, but if we allow arbitrary α then by choosing $E \subseteq E_n$ and making α_e, $e \notin E$ very small and $\alpha_e = 1$ for $e \in E$ then $G_{\Sigma,p}$ will essentially be a random subgraph of $G = ([n], E)$, perhaps with a difficult distribution.

We first discuss the connectivity threshold: We need the following notation.

$$\alpha_v = \sum_{w \neq v} \alpha_{vw}, \qquad \textit{for } v \in [n],$$

where, if $e = \{v, w\}$ then $\alpha_{vw} = \alpha_e$.

Theorem 2.5

(a) *Let* $p = \frac{\ln n + c_n}{n}$. *Then if* $\alpha = \mathbf{1}$,

$$\lim_{n \to \infty} \mathsf{P}(G_{\Sigma,p} \textit{ is connected}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-e^{-c}} & c_n \to c \\ 1 & c_n \to \infty \end{cases}.$$

(b) *Suppose that* α *is* M*-bounded and* $M \leqslant (\ln n)^{1/4}$. *Let* p_0 *be the solution to*

$$\sum_{v \in [n]} \xi_v(p) = 1$$

where $\xi_v(p) = \left(1 - \frac{\alpha_v p}{N}\right)^N$. *Then for any fixed* $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}(G_{\Sigma,p} \textit{ is connected}) = \begin{cases} 0 & p \leqslant (1-\varepsilon)p_0 \\ 1 & p \geqslant (1+\varepsilon)p_0 \end{cases}.$$

Our proof of part (a) of the above theorem relies on the following:

Lemma 2.6 *If* $\alpha = \mathbf{1}$ *and* m *is the number of edges in* $G_{\Sigma,p}$ *then*

(a) *Conditional on* m, $G_{\Sigma,p}$ *is distributed as* $G_{n,m}$ *i.e. it is a random graph on vertex set* $[n]$ *with* m *edges.*

(b) **Whp** m *satisfies*

$$\mathsf{E}(m) + \sqrt{\mathsf{E}(m)\omega} \leqslant m \leqslant \mathsf{E}(m) + \sqrt{\mathsf{E}(m)\omega}$$

for any $\omega = \omega(n)$ *which tends to infinity with* n.

So to prove part Theorem 2.5(a) all we have to verify is that $\mathsf{E}(m) \sim \frac{1}{2}n(\ln n + c_n)$ and apply known results about the connectivity threshold for random graphs, see for example Bollobás [4] or Janson, Łuczak and Ruciński [11]. (We do this explicitly in Section 4.2). Of course, this implies much more about $G_{\Sigma,p}$ when $\alpha = \mathbf{1}$. It turns out to be $G_{n,m}$ in disguise, where $m = m(p)$.

Our next theorem concerns the existence of a giant component i.e. one of size linear in n. It is somewhat weak.

Theorem 2.7 *Let* $\varepsilon > 0$ *be a small positive constant and* α *be* M*-bounded.*

(a) *If* $p \leqslant \frac{(1-\varepsilon)}{Mn}$ *then* **whp** *the maximum component size in* $G_{\Sigma,p}$ *is* $O(\ln n)$.

(b) *If* $p \geqslant \frac{(1+\varepsilon)M}{n}$ *then* **whp** *there is a unique giant component in* $G_{\Sigma,p}$ *of size* $\geqslant \kappa n$ *where* $\kappa = \kappa(\varepsilon, M)$.

Next, we turn our attention to the diameter of $G_{\Sigma,p}$.

Theorem 2.8 *Let* $k \geqslant 2$ *be a fixed integer. Suppose that* α *is* M*-bounded and assume that* $M = n^{o(1)}$. *Suppose that* θ *is fixed and satisfies* $\frac{1}{k} < \theta < \frac{1}{k-1}$. *Suppose that* $p = \frac{1}{n^{1-\theta}}$. *Then* **whp** $\mathbf{diam}(G_{\Sigma,p}) = k$.

2.3 Randomly weighted graphs

We will also consider the use of X as weights for an optimisation problem. In particular, we will consider the Minimum Spanning Tree (MST) and the Asymmetric Traveling Salesman Problem (ATSP) in which the weights $X : [n]^2 \to \mathbb{R}_+$ are randomly chosen from a simplex.

Our next theorem concerns spanning trees. We say that α is *decomposable* if there exist d_v, $v \in [n]$ such that $\alpha_{vw} = d_v d_w$. In which case we define

$$d_S = \sum_{v \in S} d_v \ for \ S \subseteq V \ and \ D = d_V.$$

Let Λ_X be weight of the minimum weight spanning tree of the complete graph K_n when the edge weights are given by a random point X from $\Sigma_{n,\alpha}$.

Theorem 2.9 *If α is decomposable and $d_v \in [\omega^{-1}, \omega]$, $\omega = (\ln n)^{1/10}$ for $v \in V$ and X is chosen uniformly at random from $\Sigma_{n,\alpha}$ then*

$$\mathsf{E}[\Lambda_X] \sim \sum_{k=1}^{\infty} \frac{(k-1)!}{D^k} \sum_{\substack{S \subseteq V \\ |S|=k}} \frac{\prod_{v \in S} d_v}{d_S^2}.$$

(The notation $a_n \sim b_n$ means that $\lim_{n \to \infty}(a_n/b_n) = 1$*, assuming that $b_n > 0$ for all n.)*

Note that if $d_v = 1$ for all $v \in [n]$ then the expression in the theorem yields $\mathsf{E}[\Lambda_X] \sim \zeta(3)$.

Now we consider the Asymmetric Traveling Salesman Problem. We will need to make an extra assumption about the simplex. We assume that

$$\alpha_{v_1,w} = \alpha_{v_2,w} \ for \ all \ v_1, v_2, w.$$

Under this assumption, the distribution of the weights of edges leaving a vertex v is independent of the particular vertex v. We call this *row symmetry.* We show that a simple patching algorithm based on that in [14] works **whp**.

Theorem 2.10 *Suppose that the cost matrix X of an instance of the ATSP is drawn from a row symmetric M-bounded simplex where $M \leqslant n^\delta$, for sufficiently small δ. Then there is an $O(n^3)$ algorithm that* **whp** *finds a tour that is asymptotically optimal, i.e.,* **whp** *the ratio of cost of the tour found to the optimal tour cost tends to one.*

3 Proofs: logconcave densities

We consider logconcave distributions restricted to the positive orthant. We also assume they are down-monotone, i.e., if $x \geqslant y$ then the density function f satisfies $f(y) \geqslant f(x)$. We begin by collecting some well-known facts about logconcave densities and proving some additional properties. These properties will be the main tools for our subsequent analysis and allow us to deal with the non-independence of edges. In particular, they will allow us to estimate the probability that certain sets of edges are included or excluded from $G_{F,p}$. We specifically assume the following about F:

Assumption A: $F : \mathbb{R}_+^N \to \mathbb{R}_+$ *is a distribution with a down-monotone logconcave density function f with support in the positive orthant.*

The two main lemmas of this section are

Lemma 3.1 *Let F satisfy Assumption A. Let $G = (V, E)$ be a random graph from $G_{F,p}$ and $S \subseteq V \times V$ with $|S| = s$. Then*

$$e^{-a_1 ps/\sigma_{\min}} \leqslant \mathsf{P}(S \cap E = \emptyset) \leqslant e^{-a_2 ps/\sigma_{\max}}$$

where a_1, a_2 are some absolute constants and the lower bound requires $p < \sigma_{\min}/4$.

Lemma 3.2 *Let F satisfy Assumption A. Let $G = (V, E)$ be a random graph from $G_{F,p}$ and $S \subseteq V \times V$ with $|S| = s$. There exist constants $b_1 < b_2$ such that*

$$\left(\frac{b_1 p}{\sigma_{\max}}\right)^s \leqslant \mathsf{P}(S \subseteq E) \leqslant \left(\frac{b_2 p}{\sigma_{\min}}\right)^s.$$

The lower bound requires $p \leqslant \sigma_{\min}/4$.

Note how these lemmas approximate what happens in $G_{n,p}$ and note the absence of an inequality for $\mathsf{P}(S \cap E = \emptyset, T \subseteq E)$ where $S \cap T = \emptyset$. The lower bounds are not used in this paper, but we hope to be able to use them in any subsequent paper.

3.1 Properties

The following classical theorem summarizing basic properties of logconcave functions was proved by Dinghas [5], Leindler [15] and Prékopa [19, 20].

Theorem 3.3 *All marginals as well as the distribution function of a logconcave function are logconcave. The convolution of two logconcave functions is logconcave.*

We will use several results from [16]. In order to state them we need some additional notation. A logconcave function $f : \mathbb{R}^m \to \mathbb{R}_+$ is *isotropic* if (i) it has mean 0 and (ii) its co-variance matrix is the identity. It is a *density* if $\int_x f(x)dx = 1$. If f is a density then so is $f_\lambda(x) = \lambda^m f(\lambda x)$. Also $\sigma_e(f_\lambda) = \sigma_e(f)/\lambda$ for all e. These identities are useful for translating results on the *isotropic* case to a more general case. For a function f we denote its maximum value by M_f.

Lemma 3.4

(a) *Let $f : \mathbb{R} \to \mathbb{R}_+$ be a logconcave density function with mean μ_f. Then*

$$\frac{1}{8\sigma_f} \leqslant f(\mu_f) \leqslant M_f \leqslant \frac{1}{\sigma_f}.$$

(For a one dimensional function f, it is appropriate to use $\sigma_f = \sigma(f)$).

(b) *Let X be a random variable with a logconcave density function $f : \mathbb{R} \to \mathbb{R}_+$.*

(i) *For every $c > 0$,*

$$\mathsf{P}(f(X) \leqslant c) \leqslant \frac{c}{M_f}.$$

(ii)

$$\mathsf{P}(X \geqslant \mathsf{E}(X)) \geqslant \frac{1}{e}.$$

(c) *Let X be a random point drawn from a logconcave distribution in $\mathbb{R}^m$. Then*

$$\mathsf{E}(|X|^k)^{1/k} \leqslant 2k\mathsf{E}(|X|).$$

(d) *If $f : \mathbb{R}^s \to \mathbb{R}_+$ is an isotropic logconcave density function then*

$$M_f \geqslant (4e\pi)^{-s/2}.$$

Proof The above lemma is from [16]. Part (a) of this lemma is from Lemma 5.5. Part (bi) is Lemma 5.6(a) and Part (bii) is Lemma 5.4. Part (c) is Lemma 5.22. Part (d) is Lemma 5.14(c). □

We prove the next two lemmas with our theorems in mind.

Lemma 3.5 *Let X be a random variable with a non-increasing logconcave density function $f : \mathbb{R}_+ \to \mathbb{R}_+$.*

(a) *For any $p \geqslant 0$,*

$$\mathsf{P}(X \leqslant p) \leqslant pM_f \leqslant \frac{p}{\sigma_f}.$$

(b) *For any $0 \leqslant p \leqslant \sigma_f$,*

$$\mathsf{P}\left(X \leqslant p\right) \geqslant \frac{p}{2\sigma_f}.$$

Proof For part (a) use $\mathsf{P}(x \leqslant p) = \int_{x=0}^{p} f(x)dx \leqslant pM_f$ and then apply Lemma 3.4(a).

For part (b), we check the value of $f(p)$. If $f(p) \geqslant M_f/2$, then the claim follows from monotonicity. If not, by Lemma 3.4(bi),

$$\mathsf{P}\left(f(X) \leqslant \frac{M_f}{2}\right) \leqslant \frac{1}{2}$$

and so

$$\mathsf{P}(X \leqslant p) \geqslant \mathsf{P}\left(f(X) \geqslant \frac{M_f}{2}\right) \geqslant \frac{1}{2} \geqslant \frac{p}{2\sigma_f}$$

as required. □

Lemma 3.6 *Let $v = (v_1, \dots, v_s)$ where*

$$v_i = \int_{\mathbb{R}_+^s} x_i f(x)\, dx$$

be the centroid of F. Then $v_i \geqslant \sigma_i/4$ for all $i \leqslant s$ and $f(v) \geqslant e^{-A_1 s}/\sigma_\Pi$, where $\sigma_\Pi = \prod_{i=1}^{s} \sigma_i$ and $A_1 > 0$ is some absolute constant.

Proof Applying Lemma 3.4(c) with $k = 2$ gives

$$v_i \geqslant \frac{1}{4}\left(\int_{\mathbb{R}_+^s} x_i^2 f(x)\,dx\right)^{\frac{1}{2}} \geqslant \frac{\sigma_i}{4}.$$

We next prove that

$$f(v) \geqslant 2^{-2s-4} f(0). \tag{2}$$

Let $H \subseteq \mathbb{R}^s$ be a hyperplane through v that is tangent to the set $\{x : f(x) \geqslant f(v)\}$. Let a be the unit normal to H. The down-monotonicity of f implies that a is non-negative. Let $H(t)$ denote the hyperplane parallel to H at distance t from the origin. Let

$$h(t) = \int_{H(t)} f(y)dy$$

be the marginal of f along a. The function h is also a logconcave density and observe that its mean $\mu_h = a \cdot v$.

Let x be a point on $H = H(a \cdot v)$. Since H is a tangent plane $f(x) \leqslant f(v)$. Using logconcavity,

$$f(x/2)^2 \geqslant f(0)f(x)$$

and so

$$f(x/2) \geqslant \sqrt{\frac{f(0)}{f(x)}} f(x) \geqslant \sqrt{\frac{f(0)}{f(v)}} f(x).$$

Therefore

$$h(a \cdot v/2) = \int_{H(a \cdot v/2)} f(y)\,dy \geqslant \frac{1}{2^{s-1}}\sqrt{\frac{f(0)}{f(v)}} h(\mu_h) \geqslant \frac{1}{2^{s-1}}\sqrt{\frac{f(0)}{f(v)}} \frac{1}{8\sigma(h)}$$

where we have used Lemma 3.4(a) for the last inequality.

On the other hand, using Lemma 3.4(a) we have $h(a \cdot v/2) \leqslant M_h \leqslant \frac{1}{\sigma(h)}$ and (2) follows.

Applying Lemma 3.4(d) to the isotropic logconcave function

$$\hat{f}(y_1, y_2, \ldots, y_s) = 2^{-s}\sigma_\Pi f(|\sigma_1 y_1|, |\sigma_2 y_2|, \ldots, |\sigma_s y_s|)$$

we see that $f(0)$ which is the maximum of $\hat{f}$ is at least $(2\pi e)^{-s}/\sigma_\Pi$. The lemma follows from (2). □

3.1.1 Proofs of the Main lemmas

Proof of Lemma 3.1 We consider the projection of F to the subspace spanned by S. For $x \in \mathbb{R}_+^S$ let

$$f_S(x) = \int_{y \in \mathbb{R}^{\bar{S}}} f(x, y)dy.$$

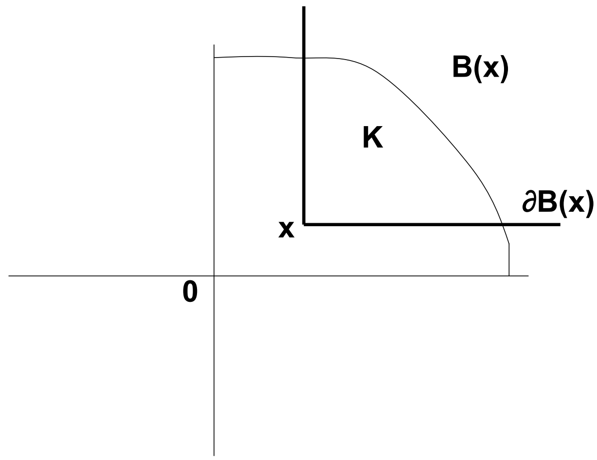

Figure 1: Proof of Lemma 3.1: the ratio of the measure of $\partial B(x) \cap K$ to the measure of $B(x) \cap K$ is a nonincreasing function of each coordinate x_e.

It is logconcave by Theorem 3.3. For a point $x \in \mathbb{R}^s_+$, let $B(x)$ be the positive orthant at x, i.e.,

$$B(x) = \{y \in \mathbb{R}^s_+ \,:\, y \geqslant x\}.$$

Let $g(x)$ be the integral of f_S over $B(x)$. Then by Theorem 3.3, g is also logconcave. The function $h(x) = \ln g(x)$ is concave and so for $e \in S$,

$$\frac{\partial h(x)}{\partial x_e} = \frac{\frac{\partial g(x)}{\partial x_e}}{g(x)}$$

is nonincreasing, see Figure 1. Therefore, it achieves its maximum at $x_e = 0$, i.e.,

$$\frac{\partial h(x)}{\partial x_e} \leqslant \frac{\partial g(0)}{\partial x_e}$$

since $g(0) = 1$. The derivative of g at $x_e = 0$ is simply the probability mass at $x_e = 0$, i.e.,

$$\frac{\partial g(0)}{\partial x_e} = -\int_{x_e=0} f_S(x)\,dx \leqslant -\frac{1}{8\sigma_{\max}}$$

where the inequality is from Lemma 3.4(a). (Consider $\phi_e(x) = \int_{y\in\mathbb{R}^{S\setminus\{e\}}} f_S(x,y)dy$. Lemma 3.4(a) implies that $\phi_e(0) \geqslant 1/\sigma_e$). Thus, by concavity,

$$h(x) \leqslant h(0) - \frac{1}{8\sigma_{\max}} \sum_{e\in S} x_e$$

and so

$$g(x) \leqslant e^{-\sum_{e=1}^{s} x_e/8\sigma_{\max}}.$$

Setting $x_e = p$ for all $e \in S$, we get the first inequality of the lemma.

For the lower bound, first assume that $\sigma_{\max} = \sigma_{\min} = \sigma$. Let f_S be the marginal of f in R_+^S and let $v = (v_1, \ldots, v_s)$, $s = |S|$ be the centroid of f_S. Consider the box induced by the origin and v. From Lemma 3.6

$$g(\sigma/4, \sigma/4, \ldots, \sigma/4) \geqslant f_S(v)(\sigma/4)^s \geqslant e^{-(A_1+2)s}.$$

For $p < \sigma/4$, by the logconcavity of g along the line from 0 to $(\sigma/4, \ldots, \sigma/4)$,

$$g(p, \ldots, p) \geqslant g(0)^{1-4p/\sigma} g(\sigma/4, \ldots, \sigma/4)^{4p/\sigma} = g(\sigma/4, \ldots, \sigma/4)^{4p/\sigma} \geqslant e^{-A_2 ps/\sigma}.$$

We now remove the assumption $\sigma_{\max} = \sigma_{\min}$ using scaling. Define

$$\hat{g}(y_1, y_2, \ldots, y_s) = \sigma_\Pi f(\sigma_1 y_1, \sigma_2 y_2, \ldots, \sigma_s y_s).$$

$\hat{g}$ is the density of the vector Y defined by $Y_e = X_e/\sigma_e$ for all $e \in S$. Thus $\mathsf{E}(Y_e^2) = 1$ for all $e \in S$ and

$$\mathsf{P}(X_e \geqslant p,\, e \in S) = \mathsf{P}(Y_e \geqslant p/\sigma_e,\, e \in S) \geqslant \mathsf{P}(Y_e \geqslant p/\sigma_{\min},\, e \in S) \geqslant e^{-A_2 ps/\sigma_{\min}}.$$

□

Proof of Lemma 3.2 We prove the lemma in the case where $\sigma_{\min} = \sigma_{\max} = \sigma$. The general case follows by scaling as at the end of the proof of Lemma 3.1. Consider the projection to the span of S and the induced density f_S. From Lemma 3.6, we see that for $p \leqslant \sigma/4$, for any point x with $0 \leqslant x_e \leqslant p$ for all $e \in S$, $f_S(x) \geqslant (4e^{A_1}\sigma)^{-s}$. The lower bound follows.

For the upper bound, assume $\sigma_{\min} = \sigma_{\max} = \sigma$ and project to S as before. Then consider the origin symmetric function g obtained by reflecting f on each axis and scaling to keep it a density, i.e.,

$$g(x_1, \ldots, x_n) = 2^{-s} f(|x_1|, \ldots, |x_n|).$$

This function is 1-unconditional (i.e., reflection-invariant for the axis planes) and its covariance matrix is $\sigma^2 I$. By a result of Ball (Theorem 7 in [1]), we have that

$$g(0)^{1/s} \leqslant eL/\sigma$$

where L is the supremum of L_K over all 1-unconditional convex bodies K of volume 1 with covariance matrix equal to $L_K^2 I$. It is a famous open conjecture that $L_K = O(1)$ for any convex body K of unit volume with covariance matrix $L_K^2 I$. This has been verified for 1-unconditional bodies (via the results of [3] and [18]). Thus, in our setting, $g(0) \leqslant (c/\sigma)^s$ for an absolute constant c. The upper bound on f_S follows. □

3.2 Proof of Theorem 2.1

For a set S, $|S| = k$, the probability that it forms a component of $G_{F,p}$, is by Lemma 3.1, at most $e^{-a_2pk(n-k)/\sigma_{\max}}$. Therefore,

$$\mathsf{P}(G \text{ is not connected}) \leqslant \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} e^{-a_2pk(n-k)/\sigma_{\max}}.$$

It follows that for $p \geqslant 3\sigma_{\max} \ln n/(a_2 n)$, the random graph is connected **whp**.

We show next that if $p \leqslant \sigma_{\min}/(2eb_2n)$ then **whp** $|E_{F,p}| \leqslant n/2$ and so $G_{F,p}$ cannot be connected. Indeed, if $p \leqslant \sigma_{\min}/(2eb_2n)$ where b_2 is as in Lemma 3.2 and $N = \binom{n}{2}$,

$$\mathsf{P}(|E_{F,p}| \geqslant n/2) \leqslant \binom{N}{n/2} \left(\frac{b_2 p}{\sigma_{\min}}\right)^{n/2} \leqslant \frac{1}{2^{n/2}}.$$

□

3.3 Proof of Theorem 2.3

The proof of Theorem 2.1 shows that if $p < c_1\sigma_{\min}/n$ then there are isolated vertices and so we can take $c_3 = c_1$. We have no hope of getting the constants a_1, a_2 right here for all F and so we will be content with finding a perfect matching between $V_1 = [n/2]$ and $V_2 = [n] \setminus V_1$. Applying Hall's Theorem we see that

$$\begin{aligned}
\mathsf{P}(G_{F,p} \textit{ has no p.m.}) &\leqslant 2\sum_{k=1}^{n/4} \binom{n/2}{k}\binom{n/2}{k-1} e^{-a_2k(n/2-k+1)p/\sigma_{\max}} \\
&\leqslant 2\sum_{k=1}^{n/4} \left(\frac{n^2 e^{2-a_2np/4\sigma_{\max}}}{4k^2}\right)^k \\
&= o(1)
\end{aligned}$$

provided $p \geqslant 9\sigma_{\max} \ln n/(a_2 n)$. □

3.4 Proof of Theorem 2.4

We use the following result from [10]: Let $G = (V, E)$ have n vertices and let $12 \leqslant d = d(n) \leqslant e^{\ln^{1/3} n}$ be a parameter such that with $n_0 = \frac{n \ln\ln n \ln d}{\ln n \ln\ln\ln n}$:

P1 For every $S \subset V$, if $|S| \leqslant n_0/d$ then $|N(S)| \geqslant d|S|$.
($N(S)$ denotes the set of vertices not in S that have at least one neighbor in S).

P2 There is an edge in G between any two disjoint subsets $A, B \subset V$ such that $|A|, |B| \geqslant n_0/4130$.

If G satisfies $\mathbf{P_1}, \mathbf{P_2}$ then G is Hamiltonian.

So let $p = \frac{\gamma \sigma_{\max} \ln n}{n}$ where lower bounds on $\gamma = \gamma(n)$ will be exposed below. We will use $d = \frac{\ln \ln \ln n}{\ln \ln \ln \ln n}$. First of all, if $\gamma \geqslant 2d/a_2$, then

$$\begin{aligned}
\mathsf{P}(\mathbf{P_1} \textit{ fails}) &\leqslant \sum_{s=1}^{n_0/d} \binom{n}{s} \binom{n}{ds} e^{-a_2 \gamma s(n-s) \ln n/n} \\
&\leqslant \sum_{s=1}^{n_0/d} \left(\frac{ne}{s} \cdot \frac{n^d e^d}{d^d s^d} \cdot n^{-a_2(1-o(1))\gamma} \right)^s \\
&= o(1).
\end{aligned}$$

Then, if $\gamma \geqslant \frac{3 \ln \ln \ln n}{B_2 \ln d}$, we have

$$\begin{aligned}
\mathsf{P}(\mathbf{P_2} \textit{ fails}) &\leqslant \binom{n}{n_0/4130}^2 e^{-a_2 \gamma (n_0/4130)^2 \ln n/n} \\
&\leqslant \left(\frac{B_1 n}{n_0} \cdot n^{-\gamma B_2 n_0/n} \right)^{2n_0} \qquad \textit{for some } B_1, B_2 > 0 \\
&= o(1).
\end{aligned}$$

The theorem follows.

4 Proofs: Simplex

The case of the uniform distribution over a simplex is much easier to analyse, taking into account its precise structure. Here, to some extent, probabilities can be estimated very precisely.

The following lemma represents a sharpening of Lemmas 3.1 and 3.2 for the simplex case. For $S \subseteq E$, let

$$\alpha(S) = \sum_{e \in S} \alpha_e.$$

Lemma 4.1

(a) *If $S \subseteq E_n$ and $E_p = E(G_{\Sigma_L, p})$ and $\alpha(S)p \leqslant L$ then,*

$$\mathsf{P}(S \cap E_p = \emptyset) = \left(1 - \frac{\alpha(S)p}{L} \right)^N.$$

(b) *If $S, T \subseteq E_n$ and $S \cap T = \emptyset$ and $|T| = o(n)$ and $\alpha(S)|T|p, \alpha(T)Np, MNp = o(L)$ and $\alpha(S)p \leqslant L$ then*

$$\mathsf{P}(S \cap E_p = \emptyset, T \subseteq E_p) = (1 + o(1)) \left(\prod_{e \in T} \alpha_e \right) \left(\frac{Np}{L} \right)^{|T|} \left(1 - \frac{\alpha(S)p}{L} \right)^N.$$

(If $\alpha(S)p > L$ then the above probabilities are all zero).

Proof

(a) With $\mathbf{vol}_N$ denoting N-dimensional volume, we have

$$\mathbf{vol}_N(\Sigma_L) = \frac{L^n}{N! \prod_{e\in E_n} \alpha_e}$$

and so

$$\begin{aligned} \mathsf{P}(S \cap E_p = \emptyset) &= \frac{\mathbf{vol}_N(\Sigma_L \cap \{X_e \geqslant p : \; e \in S\})}{\mathbf{vol}_N(\Sigma_L)} \\ &= \frac{(L-\alpha(S)p)^N/(N! \prod_{e\in E_n} \alpha_e)}{L^n/(N!\prod_{e\in E_n}\alpha_e)} \\ &= \left(1 - \frac{\alpha(S)p}{L}\right)^N . \end{aligned} \tag{3}$$

(b) Assume first that $S = \emptyset$. For $T' \subseteq T$ and $e \notin T'$ we have

$$\begin{aligned} \mathsf{P}(e \in E_p \mid X_f, f \in T') = 1 - \left(1 - \frac{\alpha_e p}{L - \sum_{f\in T'} \alpha_f X_f}\right)^{N-|T'|} \\ \leqslant \frac{\alpha_e(N-|T'|)p}{L - \sum_{f\in T'}\alpha_f X_f} \leqslant \frac{\alpha_e N p}{L}\left(1 + \frac{2\alpha(T')p}{L}\right). \end{aligned}$$

Hence

$$\mathsf{P}(T \subseteq E_p) \leqslant \left(\prod_{e\in T}\alpha_e\right)\left(\frac{Np}{L}\right)^{|T|} \exp\left\{\frac{2\alpha(T)|T|p}{L}\right\}. \tag{4}$$

Similarly,

$$\begin{aligned} \mathsf{P}(e \in E_p \mid X_f, f \in T') \geqslant \frac{\alpha_e(N-|T'|)p}{L - \sum_{f\in T'}\alpha_f X_f}\left(1 - \frac{\alpha_e(N-|T'|)p}{2(L - \sum_{f\in T'}\alpha_f X_f)}\right) \\ \geqslant \frac{\alpha_e N p}{L}\left(1 - \frac{|T'|}{N} - \frac{\alpha_e N p}{L}\right). \end{aligned}$$

It follows that

$$\mathsf{P}(T \subseteq E_p) = \left(\prod_{e\in T}\alpha_e\right)\left(\frac{Np}{L}\right)^{|T|} \exp\left\{O\left(\frac{|T|^2}{N} + \frac{\alpha(T)Np}{L}\right)\right\}. \tag{5}$$

Now

$$\mathsf{P}(S \cap E_p = \emptyset \mid X_e,\, e \in T) = \left(1 - \frac{\alpha(S)p}{L - \sum_{e\in T}\alpha_e X_e}\right)^{N-|T|}.$$

So, if $T \subseteq E_p$ then

$$\mathsf{P}(S \cap E_p = \emptyset \mid X_e,\, e \in T) \geqslant \left(1 - \frac{\alpha(S)p}{L}\right)^N \left(1 - \frac{2\alpha(S)\alpha(T)Np^2}{L(L-\alpha(T)p)}\right).$$

and

$$\mathsf{P}(S \cap E_p = \emptyset \mid X_e,\, e \in T) \leqslant \left(1 - \frac{\alpha(S)p}{L}\right)^N \left(1 + \frac{2\alpha(S)|T|p}{L}\right)$$

Part (b) follows by combining the above two inequalities with (5). □

4.1 Coupling $G_{\Sigma,p}$ and $G_{n,m}$ when $\alpha = 1$: Proof of Lemma 2.6.

The distribution $G_{\Sigma,p}$ conditioned on any fixed number of edges m is uniform over graphs with m edges i.e. is distributed as $G_{n,m}$. This is because Σ is *axis-symmetric* i.e. it is invariant under permutation of coordinates.

Let e_{ij} be the indicator random variable for the event that ij is an edge of $G_{\Sigma,p}$ and let $m = \sum_{i,j} e_{ij}$. Let $q = \mathsf{E}(e_{ij})$ so that $\mathsf{E}(m) = qN$. We bound the variance of m.

$$\begin{aligned} \mathsf{E}(m^2) - \mathsf{E}(m)^2 &= \sum_{ij} \mathsf{E}(e_{ij}^2) - \mathsf{E}(e_{ij})^2 + \sum_{ij \neq kl} (\mathsf{E}(e_{ij}e_{kl}) - \mathsf{E}(e_{ij})\mathsf{E}(e_{kl})) \\ &\leqslant qN + \sum_{ij \neq kl} \mathsf{P}(X_{ij} \leqslant p \text{ and } X_{kl} \leqslant p) - \mathsf{P}(X_{ij} \leqslant p)\mathsf{P}(X_{kl} \leqslant p). \qquad (6) \end{aligned}$$

It follows from Lemma 4.1 that,

$$q = \mathsf{P}(X_{ij} \leqslant p) = 1 - \left(1 - \frac{p}{L}\right)^N.$$

Furthermore, if $p \leqslant L/2$ then

$$\begin{aligned} \mathsf{P}(X_{kl} \leqslant p \text{ and } X_{ij} \leqslant p) = 1 - \mathsf{P}(X_{ij} \geqslant p) - \mathsf{P}(X_{kl} \geqslant p) + \mathsf{P}(X_{ij} \geqslant p \textit{ and } X_{kl} \geqslant p) = \\ 1 - 2\left(1 - \frac{p}{L}\right)^N + \left(1 - \frac{2p}{L}\right)^N. \end{aligned}$$

Using these identities, we see that if $p \leqslant L/2$ then

$$\begin{aligned} \mathsf{E}(m^2) - \mathsf{E}(m)^2 &\leqslant qN + \frac{N(N-1)}{2}\left(1 - 2\left(1-\frac{p}{L}\right)^N + \left(1-\frac{2p}{L}\right)^N - \left(1-\left(1-\frac{p}{L}\right)^N\right)^2\right) \\ &= qN + \frac{N(N-1)}{2}\left(\left(1 - \frac{2p}{L}\right)^N - \left(1 - \frac{p}{L}\right)^{2N}\right) \\ &\leqslant qN. \qquad (7) \end{aligned}$$

If $p > L/2$ then $\mathsf{P}(X_{kl} \leqslant p \text{ and } X_{ij} \leqslant p) = 1 - 2\left(1 - \frac{p}{L}\right)^N$ and so (7) is still true.

Using Chebyshev's inequality,

$$\mathsf{P}(qN + \sqrt{qN\omega} \leqslant m \leqslant qN + \sqrt{qN\omega}) = 1 - o(1). \qquad (8)$$

This completes the proof of Lemma 2.6. □

4.2 Connectivity for $G_{\Sigma,p}$ when $\alpha = 1$: Proof of Theorem 2.5 (a)

Suppose first that $c_n \to c$. Let now $L = N$ and let $p = \frac{\ln n + c_n}{n}$ and let $m = |E_p|$. Then q in Section 4.1 satisfies

$$p - \frac{p^2}{2} \leqslant q \leqslant p. \tag{9}$$

Let $m_0 = Np - n^{2/3}$ and $m_1 = Np + n^{2/3}$. Now (8) implies that **whp**, $m_0 \leqslant m \leqslant m_1$. But then

$$o(1) + e^{-e^{-c}} = o(1) + \mathsf{P}(G_{n,m_1} \textit{ is connected}) \leqslant \mathsf{P}(S_{p,\mathbf{1}} \textit{ is connected})$$
$$\leqslant o(1) + \mathsf{P}(G_{n,m_2} \textit{ is connected}) = o(1) + e^{-e^{-c}}.$$

Taking limits gives the result for $c_n \to c$ and the result for $c_n \to \pm\infty$ follows by monotonicity.

4.3 Connectivity for $G_{\Sigma,p}$: Proof of Theorem 2.5 (b)

Applying Lemma 4.1(a) with $L = N$ we see that for $v, w \in [n]$,

$$\mathsf{P}(v \textit{ is isolated}) = \xi_v(p), \tag{10}$$

where $\xi_v = \xi_v(p) = \left(1 - \frac{\alpha_v p}{N}\right)^N$,

$$\mathsf{P}(v, w \textit{ are isolated}) = \left(1 - \frac{(\alpha_v + \alpha_w - \alpha_{vw})p}{N}\right)^N \tag{11}$$

Let $p = (1-\varepsilon)p_0$. We observe first that

$$\frac{1}{2M^2}\ln n \leqslant \alpha_v p_0 \leqslant 2M^2 \ln n \qquad \textit{for all } v \in [n]. \tag{12}$$

If the upper bound breaks for some $v \in V$, then we have $\alpha_w p_0 \geqslant 2\ln n$ and $\xi_w(p_0) \leqslant n^{-2}$ for all $w \in [n]$ and this contradicts the definition of p_0. On the other hand, if the lower bound breaks for some $v \in V$ then $\alpha_w p_0 \leqslant \frac{1}{2}\ln n$ and $\xi_w(p_0) \geqslant (1 - o(1))n^{-1/2}$ for all $w \in [n]$ and this also contradicts the definition of p_0. Define a_v by $\xi_v(p_0) = n^{-a_v}$. It follows that

$$\frac{1}{3M^2} \leqslant a_v \leqslant 3M^2 \textit{ for } v \in [n]. \tag{13}$$

Consider the function

$$\phi(x) = \sum_{v \in [n]} n^{-x a_v}.$$

We know that $\phi(1) = 1$ and $\phi'(1) = -\ln n \sum_v a_v n^{-a_v} \leqslant -\ln n/3M^2$. It follows that $\phi(1-\varepsilon) = \Omega((\ln n)^{1/2})$ for small ε and this implies that if Z_0 is the expected number of isolated vertices in $G_{\Sigma,p}$ then $\mathsf{E}(Z_0) = \Omega((\ln n)^{1/2})$.

Since $M = o(\ln n)$, (10) and (11) imply that

$$\mathsf{P}(v, w \textit{ are isolated}) \sim \mathsf{P}(v \textit{ is isolated})\mathsf{P}(w \textit{ is isolated})$$

and then Chebyshev's inequality implies that $Z_0 \neq 0$ **whp** and hence **whp** $S_{n,p,\alpha}$ is not connected.

Suppose now that $p = (1+\varepsilon)p_0$. It follows from (12) that the expected number of isolated vertices A_1 in $G_{\Sigma,p}$ satisfies

$$A_1 = \sum_{v\in[n]} \xi_v(p) \leqslant \sum_{v\in[n]} \left(1 - \frac{\varepsilon a_v p}{N}\right)^N \xi_v(p_0) \leqslant n^{-\varepsilon/2M^2} \sum_{v\in[n]} \xi_v(p_0) = n^{-\varepsilon/2M^2}.$$

Thus **whp** $G_{\Sigma,p}$ has no isolated vertices. Let A_k denote the expected number of components of size $1 \leqslant k \leqslant n/2$ in $G_{\Sigma,p}$. Let $\pi_k = \mathsf{P}(A_k \neq 0)$ and $k_0 = n/M^6(\ln n)^2$. Then for $2 \leqslant k \leqslant k_0$,

$$\begin{aligned}
\pi_k &\leqslant \sum_{|S|=k} \left(1 - \frac{\sum_{v\in S, w\notin S} \alpha_{vw} p}{N}\right)^N && (14)\\
&\leqslant \sum_{|S|=k} \left(1 - \frac{\sum_{v\in S} \alpha_v p}{N}\right)^N \Bigg/ \left(1 - \frac{\sum_{v,w\in S} \alpha_{vw} p}{N}\right)^N \\
&\leqslant \frac{e^{2k^2 Mp} A_1^k}{k!} \\
&\leqslant \left(\frac{e^{2kM(1+\varepsilon)(2M^3 \ln n/n)} n^{-\varepsilon/2M^2} e}{k}\right)^k \\
&\leqslant \left(\frac{e^{1+o(1)} n^{-\varepsilon/2M^2}}{k}\right)^k
\end{aligned}$$

after using $p_0 \leqslant 2M^3 \ln n/n$ from (12). Thus $\sum_{k=1}^{k_0} A_k = o(1)$ and so **whp** there are no components of size $1 \leqslant k \leqslant k_0$ in $G_{\Sigma,p}$.

For $k > k_0$ we use

$$\begin{aligned}
\sum_{k=k_0}^{n/2} \pi_k &\leqslant \sum_{k=k_0}^{n/2} \sum_{|S|=k} \left(1 - \frac{knp}{2MN}\right)^N \\
&\leqslant \sum_{k=k_0}^{n/2} \binom{n}{k} e^{-k \ln n/(4M^3)} \\
&\leqslant \sum_{k=k_0}^{n/2} \left(\frac{ne}{k} \cdot n^{-1/4M^3}\right)^k \\
&\leqslant \sum_{k=k_0}^{n/2} (M^6 (\ln n)^2 n^{-1/4M^3})^k \\
&= o(1).
\end{aligned}$$

Thus **whp** there are no components of size $1 \leqslant k \leqslant n/2$ in $G_{\Sigma,p}$. This completes the proof of part (b) of Theorem 2.5. □

4.4 Giant Component in $G_{\Sigma,p}$: Proof of Theorem 2.7

We use a simple coupling argument. For a vector $\mathbf{p} \in \mathbb{R}_+^N$ we define $G_{\alpha,\mathbf{p}}$ to be the random graph where X is chosen uniformly from Σ_α and an edge e is taken iff $X_e \leqslant p_e$. Suppose first that $\lambda_e > 0$ for all $e \in E_n$. Define α' by $\alpha'_e = \alpha_e \lambda_e$ and define $\mathbf{p}'$ by $p'_e = p_e/\lambda_e$. We claim that $G_{\alpha,\mathbf{p}} = G_{\alpha',\mathbf{p}'}$ in distribution. Indeed, for a fixed graph $G = (V, E)$ we have

$$\begin{aligned}
&\mathsf{P}(G_{\alpha,\mathbf{p}} = G) = \\
&\frac{1}{\mathbf{vol}_N(\Sigma_N)} \int_{\substack{0 \leqslant x_e \leqslant p_e \\ e \in E}} \mathbf{vol}_{N-|E|} \left(\left\{ x_f \geqslant p_f,\, f \notin E,\, \sum_{f \notin E} \alpha_f x_f \leqslant N - \sum_{e \in E} \alpha_e x_e) \right\} \right) \prod_{x \in E} dx_e \\
&= \left(\prod_{e \in E} \alpha_e\right) \frac{N!}{(N - |E|)! L^N} \int_{\substack{0 \leqslant x_e \leqslant p_e \\ e \in E}} \left(\max\left\{0, N - \sum_{e \in E} \alpha_e x_e - \sum_{e \notin E} \alpha_e p_e \right\}\right)^{N-|E|} \prod_{x \in E} dx_e \\
&= \left(\prod_{e \in E} \alpha'_e\right) \frac{N!}{(N - |E|)! L^N} \int_{\substack{0 \leqslant y_e \leqslant p'_e \\ e \in E}} \left(\max\left\{0, N - \sum_{e \in E} \alpha'_e y_e - \sum_{e \notin E} \alpha'_e p'_e \right\}\right)^{N-|E|} \prod_{e \in E} dy_e \\
&= \mathsf{P}(G_{\alpha',\mathbf{p}'} = G)
\end{aligned}$$

So for (a) we start with $\mathbf{p} = p\mathbf{1}$ and take $\lambda_e = 1/\alpha_e$ to get $G_{\Sigma,p} = G_{\mathbf{1},\mathbf{p}'}$ in distribution. Note that $p'_e \leqslant (1-\varepsilon)/n$ and so we can couple so that $G_{\mathbf{1},\mathbf{p}'} \subseteq G_{\mathbf{1},\frac{1-\varepsilon}{n}\mathbf{1}}$. Part (a) follows from (8) as in Section 4.1. Part (b) is similar.

4.5 Diameter of $G_{\Sigma,p}$: Proof of Theorem 2.9

Recall that $p = \frac{1}{n^{1-\theta}}$ where $\frac{1}{k} < \theta < \frac{1}{k-1}$. We show first that **whp** the diameter exceeds $k-1$. Let Z_t denote the number of paths of length $t \leqslant k-1$ from vertex 1 to vertex 2. We consider the existence of t edges making up a path. Applying Lemma 4.1(b): $S = \emptyset$ and $|T| = k$,

$$\begin{aligned} \mathsf{E}[Z_t] &\leqslant (1+o(1))n^{t-1}(Mp)^t \\ &\leqslant 2n^{t-1}\left(\frac{M}{n^{1-\theta}}\right)^t \\ &= 2M^t n^{\theta t-1} \\ &= o(1). \end{aligned}$$

Case 1: $k \geqslant 3$.
We must now show that the diameter is at most k. The following lemma provides some structure:

Lemma 4.2 *The following hold* **whp***:*

(a) *The maximum degree* $\Delta \leqslant \Delta_0 = 10Mn^\theta$.

(b) *Suppose that* $S \subseteq V$ *with* $|S| \leqslant n^{1-\theta-\varepsilon}$ *for some fixed* ε. *Then*

$$|N(S)| \geqslant n^\theta |S|/(10M\ln n).$$

Proof (a) We consider the existence of $t = 10Mn^\theta$ edges incident with a fixed vertex. Applying Lemma 4.1(b): $S = \emptyset$ and $|T| = \Delta_0$. ($k \geqslant 3$ is needed here to ensure that $\alpha(T)p = o(1)$).

$$\mathsf{P}[\Delta \geqslant \Delta_0] \leqslant (1+o(1))n\binom{n}{\Delta_0}(Mp)^{\Delta_0} \leqslant 2n\left(\frac{e}{10}\right)^{\Delta_0} = o(1).$$

(b) Using Lemma 4.1(a) we see that the probability that this fails to hold can be bounded by

$$\sum_{|S|=1}^{n^{1-\theta-\varepsilon}} \sum_{|T|=0}^{n^\theta s/(10M\ln n)} \left(1 - \frac{|S|(n-|S|-|T|)p}{MN}\right)^N \leqslant$$

$$\sum_{s=1}^{n^{1-\theta-\varepsilon}} \sum_{t=0}^{n^\theta s/(10M\ln n)} n^{s+t}\exp\left\{-s(n-s-t)n^{\theta-1}/M\right\} \leqslant$$

$$\sum_{s=1}^{n^{1-\theta-\varepsilon}} \sum_{t=0}^{n^\theta s/(10M\ln n)} n^{s+t}e^{-sn^\theta/2M} = o(1).$$

$\square$

For a vertex v let $N_r(v)$ be the set of vertices at distance r from v. Let $r_0 = \left\lfloor \frac{k-1}{2} \right\rfloor$ and $r_1 = \left\lfloor \frac{k}{2} \right\rfloor$. It follows from Lemma 4.2 that **whp** we have

$$(n^\theta/(10M\ln n))^r \leqslant |N_r(v)| \leqslant (10Mn^\theta)^r \qquad for\ 1 \leqslant r \leqslant r_1.$$

Furthermore, we have $r_0+r_1 \leqslant k-1$. So suppose that $v, w \in V$ and $N_{r_0}(v) \cap N_{r_1}(w) = \emptyset$. (If the intersection is non-empty then their distance is already $\leqslant k$). Now condition on the sets T, S of edges and non-edges exposed in the construction of $N_{r_0}(v), N_{r_1}(w)$. Then **whp** we have $|S| = O(n(M\Delta_0)^{r_1})$ and $|T| = O((M\Delta_0)^{r_1})$.

Let $\nu_v = |N_{r_0}(v)|, \nu_w = |N_{r_1}(w)|$. Given S, T let $R = \{xy : x \in N_{r_0}(v), y \in N_{r_1}(w)\}$. Using Lemma 4.1(b), the conditional probability that there is no edge between $N_{r_0}(v)$ and $N_{r_1}(w)$ is bounded as follows: $|R| + |S| = O(n^{r_1\theta+1+o(1)})$ and $|T| = O(n^{r_1\theta+o(1)})$.

$$\frac{\mathsf{P}((R \cup S) \cap E_p = \emptyset,\, T \subseteq E_p)}{\mathsf{P}(S \cap E_p = \emptyset,\, T \subseteq E_p)} = (1+o(1)) \left(1 - \frac{\alpha(R)p}{N}\right)^N$$
$$\leqslant 2e^{-\nu_v\nu_w p/M} = \exp\left\{-\Omega(n^{(r_0+r_1+1)\theta-1-o(1)})\right\}. \quad (15)$$

Now $(r_0 + r_1 + 1)\theta - 1 = \Omega(1)$ and this completes the proof for the case $k \geqslant 3$.

Case 2: $k = 2$.
This is much simpler. We show that if $p = n^{-\beta}$ where $\beta = 1/2 - \varepsilon$ then $\mathbf{diam}(G_{\Sigma,p}) = 2$ **whp**. Here ε is an arbitrarily small positive constant.

We first argue that the minimum degree in $G_{\Sigma,p}$ is at least $\Delta_1 = n^{1/2+\varepsilon}/(10M\ln n)$. Indeed, if δ denotes minimum degree then from Lemma 4.1(a),

$$\mathsf{P}[\delta \leqslant \Delta_1] \leqslant n\binom{n}{n-\Delta_1}\left(1 - \frac{(n-\Delta_1)p}{MN}\right)^N = o(1).$$

Then by conditioning on $N(v)$, we argue as in (15) that **whp** every pair of distinct vertices v, w have a common neighbour. More precisely,

$$\frac{\mathsf{P}(N(v) \cap N(w) = \emptyset, N(v) = X)}{\mathsf{P}(N(v) = X)} = (1+o(1))\left(1 - \frac{\Delta_1 p}{MN}\right)^N \leqslant e^{-n^\varepsilon}.$$

4.6 Minimum Spanning Tree: Proof of Theorem 2.8

Suppose that T is our minimum length spanning tree. Then we can write its length $\ell(T)$ as

$$\begin{aligned}
\ell(T) &= \sum_{e \in T} X_e \\
&= \sum_{e \in T} \int_{p=0}^{N} 1_{X_e \geqslant p} dp \\
&= \int_{p=0}^{N} \sum_{e \in T} |\{e:\ X_e \geqslant p\}| dp \\
&= \int_{p=0}^{N} (\kappa(G_{\Sigma,p}) - 1) dp
\end{aligned}$$

where κ denotes the number of components. The final equation is the only place where we need to assume that T is a minimum length spanning tree.

So,

$$\Lambda_X = \int_{p=0}^{N} (\mathsf{E}[\kappa(G_{\Sigma,p})] - 1])dp \tag{16}$$

The general strategy from now on is to show that the integral in (16) is dominated by small values of p and the expectation $\mathsf{E}[\kappa(G_{\Sigma,p})]$ is dominated by the expected number of small components. We then try and carefully estimate the expected number of small components when p is small. So, a lot of the proof involves showing that certain quantities can be ignored.

Going back to (14) (with $M = \omega^2$) we see that

$$\pi_k \leqslant \binom{n}{k}\left(1 - \frac{knp}{2\omega^2 N}\right)^N \leqslant \left(\frac{ne}{k} \cdot e^{-np/2\omega^2}\right)^k \tag{17}$$

for $1 \leqslant k \leqslant n/2$.

So, $p \geqslant p_0 = \frac{5\omega^2 \ln n}{n}$ implies $Pr[G_{\Sigma,p}$ is not connected$] = o(N^{-2})$. Therefore,

$$\Lambda_X = \int_{p=0}^{p_0} (\mathsf{E}[\kappa(G_{\Sigma,p})] - 1])dp + o(N^{-1}). \tag{18}$$

Next let $\kappa_{k,p}$ denote the number of components with k vertices. $\kappa_{1,p}$ is the number of isolated vertices and

$$\mathsf{E}[\kappa_{1,p}] = \sum_{v \in V}\left(1 - \frac{d_v(D - d_v)p}{N}\right)^N.$$

It follows that

$$\Lambda_X \geqslant \int_{p=0}^{p_0} \sum_{v \in V}\left(1 - \frac{d_v(D - d_v)p}{N}\right)^N dp - p_0 + o(N^{-1}) \geqslant \Lambda_0 = \frac{1}{2D}\sum_{v \in V}\frac{1}{d_v} \geqslant \frac{1}{2\omega^2}. \tag{19}$$

Using Lemma 4.1(b) to tighten (17), we see that for $k \leqslant n^{1/2}$ and $p \leqslant p_0$,

$$\mathsf{E}[\kappa_{k,p}] \leqslant \sum_{|S|=k} k^{k-2}(\omega^2 p)^{k-1}\left(1 - \frac{knp}{2\omega^2 N}\right)^N \leqslant \frac{1}{\omega^2 p}\left(ne \cdot \omega^2 p e^{-np/2\omega^2}\right)^k. \tag{20}$$

Explanation: Choose a set S of k vertices and then a tree H on these vertices in k^{k-2} ways. $(\omega^2 p)^{k-1}\left(1 - \frac{kn}{2\omega^2 N}\right)^N$ bounds the probability that H exists and there are no edges from S to $V \setminus S$.

So if $p_1 = \frac{20\omega^2 \ln \omega}{n}$ then for $k \leqslant n^{1/2}$,

$$\begin{aligned}\int_{p=p_1}^{p_0} (\mathsf{E}[\kappa_{k,p}] - 1)dp &\leqslant \frac{1}{\omega^2 p_1} \left(2e\omega^4\right)^k \int_{p=p_1}^{\infty} \left(\frac{np}{2\omega^2} e^{-np/2\omega^2}\right)^k dp \\ &= \frac{2}{np_1} \left(2e\omega^4\right)^k \int_{x=10\ln\omega}^{\infty} (xe^{-x})^k dx \\ &\leqslant \frac{2}{np_1} \left(2e\omega^4\right)^k \int_{x=10\ln\omega}^{\infty} e^{-2kx/3} dx \\ &\leqslant \frac{2}{np_1} \left(2e\omega^4\right)^k \frac{1}{k\omega^{6k}} \\ &\leqslant \frac{1}{\omega^{k+2}}.\end{aligned}$$

Now for any k there are fewer than n/k components of size $\geqslant k$. So,

$$\sum_{k \geqslant n^{1/2}} \int_{p=p_1}^{p_0} (\mathsf{E}[\kappa_{k,p}] - 1)dp \leqslant n^{1/2} p_0.$$

It follows from (18) and (19) that

$$\begin{aligned}\Lambda_X &= \int_{p=0}^{p_1} (\mathsf{E}[\kappa(G_{\Sigma,p})] - 1])dp + O\left(\sum_{k=1}^{\infty} \frac{1}{\omega^{k+2}} + n^{1/2} p_0\right) + o(N^{-1}) \qquad (21) \\ &\sim \int_{p=0}^{p_1} \mathsf{E}[\kappa(G_{\Sigma,p})]dp \\ &= \sum_{k=1}^{\omega^5} \int_{p=0}^{p_1} \mathsf{E}[\kappa_{k,p}]dp + O(np_1/\omega^5) \\ &\sim \sum_{k=1}^{\omega^5} \int_{p=0}^{p_1} \mathsf{E}[\kappa_{k,p}]dp, \qquad (22)\end{aligned}$$

Now let $\tau_{k,p}$ denote the number of components of $G_{\Sigma,p}$ that are isolated trees with k vertices. For $X \subseteq V$ we let $A_k = \left\{a \in [1,k]^k : \sum_{j=1}^k a_j = 2k-2\right\}$. Then, where $q = e^{-Dp}$,

$$\mathsf{E}[\tau_{k,p}] \sim (k-2)! p^{k-1} \sum_{a \in A_k} \sum_{\substack{f:[k] \to V \\ f \text{ an injection}}} \prod_{j=1}^{k} \frac{d_{f(j)}^{a_j} q^{d_{f(j)}}}{(a_j - 1)!} \qquad \textit{for } k \leqslant \omega^5. \qquad (23)$$

Explanation: We choose a degree sequence a_j, $j = 1, 2, \ldots, k$ for our tree. Then we choose f to assign vertices to the degrees. The number of trees with this degree sequence is $\frac{(k-2)!}{\prod_{v \in X}(a_v - 1)!}$. Let H be such a tree. Going back to Lemma 4.1(b) with $T = E(H)$ and $|S| = k(n-k) + \binom{k}{2} - k + 1$ we see that the probability H is an isolated tree component is $\sim p^{k-1} \prod_{v \in X} d_v^{a_v} \left(1 - \frac{d_v D p}{N}\right)^N \sim p^{k-1} \prod_{v \in X} d_v^{a_v} q^{d_v}$.

We will show that the expression (23) can be written

$$\mathsf{E}[\tau_{k,p}] \sim (k-2)!p^{k-1} \sum_{a \in A_k} \prod_{i=1}^{k} \sum_{v=1}^{n} \frac{d_v^{a_i} q^{d_v}}{(a_i-1)!}. \tag{24}$$

Observe that the sum Σ on the RHS of (24) can be expressed

$$\Sigma = \Sigma_1 + \cdots + \Sigma_k$$

where

$$\Sigma_j = \sum_{a \in A_k} \sum_{f \in \mathcal{F}_j} \psi(a, f)$$

and $\mathcal{F}_j$ is the set of functions from $[k] \to V$ with a range of size j and $\psi(a,f) = \prod_{i=1}^{k} \frac{d_{f(i)}^{a_j} q^{d_{f(i)}}}{(a_i-1)!}$.

Thus the sum on the RHS of (23) is equal to Σ_k. We show next that

$$\frac{\Sigma_{j+1}}{\Sigma_j} \geqslant n^{1-o(1)} \qquad\qquad 1 \leqslant i < k. \tag{25}$$

Observe first that

$$\frac{1}{\omega^{2k} e^{k\omega Dp} k!} \leqslant \psi(a,f) \leqslant \omega^{2k}.$$

Our bounds $\omega^{10} \leqslant \ln n, k \leqslant \omega^5, p \leqslant p_1$ imply that $\psi(a,f) = n^{o(1)}$ for all a, f. So, $\Sigma_j = |\mathcal{F}_j| n^{o(1)} = n^{j+o(1)}$. This confirms (25), which implies that $\Sigma \sim \Sigma_k$ and confirms (24).

We re-write (24) as

$$\begin{aligned}
\mathsf{E}[\tau_{k,p}] &\sim (k-2)!p^{k-1}[x^{2k-2}] \left(\sum_{v=1}^{n} \sum_{r=1}^{\infty} \frac{q^{d_v} d_v^r}{(r-1)!} x^r \right)^k \\
&= (k-2)!p^{k-1}[x^{k-2}] \left(\sum_{v=1}^{n} q^{d_v} d_v e^{d_v x} \right)^k \\
&= (k-2)!p^{k-1} \sum_{\substack{S \subseteq V \\ |S|=k}} q^{d_S} \frac{d_S^{k-2}}{(k-2)!} \prod_{v \in S} d_v
\end{aligned} \tag{26}$$

where $d_S = \sum_{v \in S} d_v$.

So,

$$\sum_{k=1}^{\omega^5}\int_{p=0}^{p_1}\mathsf{E}[\tau_{k,p}]dp \sim \sum_{k=1}^{\omega^5}\sum_{\substack{S\subseteq V\\|S|=k}} d_S^{k-2}\prod_{v\in S}d_v\int_{p=0}^{p_1}p^{k-1}e^{-d_S Dp}dp$$

$$= \sum_{k=1}^{\omega^5}\sum_{\substack{S\subseteq V\\|S|=k}}\frac{\prod_{v\in S}d_v}{d_S^2D^k}\int_{x=0}^{d_SDp_1}x^{k-1}e^{-x}dx \tag{27}$$

$$\sim \sum_{k=1}^{\omega^5}\sum_{\substack{S\subseteq V\\|S|=k}}\frac{\prod_{v\in S}d_v}{d_S^2D^k}\int_{x=0}^{\infty}x^{k-1}e^{-x}dx \tag{28}$$

$$= \sum_{k=1}^{\omega^5}\frac{(k-1)!}{D^k}\sum_{\substack{S\subseteq V\\|S|=k}}\frac{\prod_{v\in S}d_v}{d_S^2} \tag{29}$$

$$\sim \sum_{k=1}^{\infty}\frac{(k-1)!}{D^k}\sum_{\substack{S\subseteq V\\|S|=k}}\frac{\prod_{v\in S}d_v}{d_S^2} \tag{30}$$

(27) **to** (28): $d_SDp_1 \geqslant 20k\ln\omega$ and $x \geqslant 20k\ln\omega$ implies that $x^{k-1} \leqslant e^{x/2}$. Hence

$$\int_{x=d_SDp_1}^{\infty}x^{k-1}e^{-x}dx \leqslant \int_{x=20k\ln\omega}^{\infty}e^{-x/2}dx = 2\omega^{-10k}.$$

(29) **to** (30):

$$\sum_{k=\omega^5}^{\infty}\frac{(k-1)!}{D^k}\sum_{\substack{S\subseteq V\\|S|=k}}\frac{\prod_{v\in S}d_v}{d_S^2} \leqslant \sum_{k=\omega^5}^{\infty}\frac{(k-1)!\omega^2}{k^2D^k}\sum_{\substack{S\subseteq V\\|S|=k}}\prod_{v\in S}d_v \leqslant \sum_{k=\omega^5}^{\infty}\frac{\omega^2}{k^3} = O(\omega^{-13})$$

which must be compared with (19).

It only remains to show that if $\sigma_{k,p} = \kappa_{k,p} - \tau_{k,p}$ then

$$\sum_{k=1}^{\omega^5}\int_{p=0}^{p_1}\mathsf{E}[\sigma_{k,p}]dp = o(\omega^{-2}). \tag{31}$$

But, arguing as in (20) we see that for $k \leqslant n/2$,

$$\mathsf{E}[\sigma_{k,p}] \leqslant \sum_{|S|=k}k^k(\omega^2p)^k\left(1-\frac{knp}{2\omega^2N}\right)^N \leqslant \left(nek\cdot\omega^2pe^{-np/2\omega^2}\right)^k.$$

Hence,

$$\sum_{k=1}^{\omega^5}\int_{p=0}^{p_1}\mathsf{E}[\sigma_{k,p}]dp \leqslant \sum_{k=1}^{\omega^5}(2ek\omega^4)^k\int_{p=0}^{p_1}\left(\frac{np}{2\omega^2}e^{-np/2\omega^2}\right)^k dp \leqslant \sum_{k=1}^{\omega^5}(2ek\omega^4)^kp_1 = n^{o(1)-1}$$

and (31) follows. □

5 TSP algorithm: Proof of Theorem 2.10

A digraph is a set of edges (i, j) and these can equally well be viewed as the set of edges of a bipartite graph. So we consider there to be a *digraph view* and a *bipartite view*. The algorithm consists of the following:

Step 1 Solve the assignment problem with cost matrix X i.e. find a minimum cost perfect matching in the bipartite view. The edges $(i, \mathbf{a}(i))$ of the optimal assignment form a set of vertex disjoint cycles $C_1, C_2, \ldots, C_k$ in the digraph view.

Step 2 Assume that $|C_1| \geqslant |C_2| \geqslant \cdots \geqslant |C_k|$.
For $i = k$ down to 2: $C_1 \leftarrow C_1 \oplus C_i$. (*Patch* C_i *into* C_1).

Here $C_1 \oplus C_i$ is obtained by removing an edge (a, b) from C_1 and an edge (c, d) from C_i and adding edges $(a, d), (c, b)$ to make one cycle. These two edges are chosen to minimise the cost $X_{ad} + X_{cb}$.

Each patch reduces the number of cycles by one and so the procedure ends with a tour.
Analysis:

(a) The row symmetry assumption implies that the matching found in Step 1 is uniformly random and so in the digraph view it has $O(\ln n)$ cycles **whp**. We prove this as follows: For any two permutations π_1, π_2 we have

$$\mathsf{P}(\mathbf{a}(X) = \pi_1) = \mathsf{P}(\mathbf{a}(\pi_1\pi_2^{-1}X) = \pi_1) = \mathsf{P}(\mathbf{a}(X) = \pi_2).$$

It follows that **whp** $|C_1| = \Omega(n/\ln n)$.

(b) We next put a high probability bound on the length of the longest edge in the solution to Step 1. There are several steps:

(1) We let $\omega = KM(\ln n)^2$ for some large constant K and argue that **whp** every vertex in G_{Σ,p_1}, $p_1 = \omega/n$, has in-degree and out-degree at least $\omega_0 = L \ln n$ where $L = K^{1/2}$.

To verify the degree bounds, fix a vertex v and partition $[n] \setminus \{v\}$ into sets $V_1, \ldots, V_{\omega_0}$ of size $\sim n/\omega_0$. Using Lemma 4.1(a) we see that

$$\mathsf{P}(\exists i : \ d_{p_1}(v, V_i) = 0) \leqslant e^{-np_1/(M\omega_0)} = n^{-L}$$

where $d_p(v, V_i)$ is the number of $G_{\Sigma,p}$ neighbors of v in V_i.

Thus with probability at least $1 - n^{-L}$, v has one out-neighbor in each part of the partition. This gives an out-degree of at least $L \ln n$ as required. In-degree is treated similarly. If $L \geqslant 2$ then the failure probability is sufficient to give the result for all v.

(2) We use Lemma 4.1(b) and a simple first moment argument to argue that if in the bipartite view we have two sets S, T contained in different sides of the partition and $|S| \leqslant n^{2/3}$ and $|T| \leqslant L|S|\ln n/4$ then **whp** the induced bipartite sub-graph on $S \cup T$ contains at most $L|S|\ln n/2$ edges of length $\leqslant p_1$. Indeed, if $\mathcal{B}$ is the event that there are S, T with more edges, then

$$\begin{aligned}
\mathsf{P}(\mathcal{B}) &\leqslant (1+o(1)) \sum_{s=1}^{n^{2/3}} \sum_{t=1}^{Ls\ln n/4} \binom{n}{s}\binom{n}{t}\binom{st}{Ls\ln n/2}\left(\frac{KM^2(\ln n)^2}{n}\right)^{Ls\ln n/2} \\
&\leqslant 2n \sum_{s=1}^{n^{2/3}} \left(\frac{ne}{s}\right)^s \left(\frac{4en}{Ls\ln n}\right)^{Ls\ln n/4} \left(\frac{KM^2e(\ln n)^2 s}{2n}\right)^{Ls\ln n/2} \\
&= 2n \sum_{s=1}^{n^{2/3}} \left(\frac{ne}{s} \cdot \left(\frac{M^4L^3e^3(\ln n)^3 s}{n}\right)^{L\ln n/4}\right)^s \\
&= o(1).
\end{aligned}$$

(3) Now suppose that the optimum solution to Step 1 contains an edge (x, y) of length greater than $2Mn^{-1/2}$. We grow alternating paths from x, y in a breadth first manner using edges of length $\leqslant p_1$. Using (b1) and (b2) we see that the levels grow at a rate $\geqslant L\ln n/5$ until they are of size at least $n^{3/5}$ say. This will happen regardless of the matching $\mathbf{a}$ produced by Step 1. Indeed, let $S_0 = \{x\}$ and in general, let $S_{i+1} = \mathbf{a}^{-1}(N_p(S_i) \setminus S_0 \cup \cdots \cup S_i$. $N_p(S)$ denotes the neighbors in G_{F,p_1} of a set S contained in one side of the partition. It follows from (b1) and (b2) that $|S_{i+1}| \geqslant L|S_i|\ln n/5$, as long as $|S_i| \leqslant n^{2/3}$. So **whp** there exists i_0 such that $|S_{i_0}| \geqslant n^{3/5}$. Similarly, if $T_0 = \{y\}$ and $T_{j+1} = \mathbf{a}(N_p(T_j)) \setminus T_0 \cup \cdots \cup T_j$ then **whp** there exists j_0 such that $|T_{j_0}| \geqslant n^{3/5}$. We can then use Lemma 4.1(a) to argue that **whp** there is an edge of length at most $Mn^{-1/2}$ joining the final two levels S, T. Indeed

$$\begin{aligned}
&\mathsf{P}(\exists |S|, |T| \geqslant n^{3/5} : \textit{there is no } S, T \textit{ edge of length } \leqslant Mn^{1/2}) \\
\leqslant &\binom{n}{n^{3/5}}^2 e^{-n^{7/10}} \\
= &o(1).
\end{aligned}$$

Then exchanging along the alternating path adds edges of total cost at most $Mn^{-1/2} + o(p_1 \ln n) \leqslant 2Mn^{-1/2}$ and removes an edge of length strictly greater than this, a contradiction.

(b) It follows from the above that we can **whp** "ignore" the edges of length $> p_2 = Mn^{-1/4}$ in our construction in Step 1. Let the edges of length $\leqslant p_2$ be denoted E_1 and the edges of length in the range $[p_2, 2p_2]$ be denoted E_2. We observe next that

whp $|E_1| \leqslant 10M^2n^{7/4}$. Indeed, applying (4) we see that if $t = 10M^2n^{7/4}$ then

$$\begin{aligned} \mathsf{P}(|E_1| \geqslant t) &\leqslant \binom{N}{t} M^t \left(\frac{M}{n^{1/4}}\right)^t \exp\left\{\frac{2M^3t^2}{Nn^{1/4}}\right\} \\ &\leqslant \left(\frac{Ne}{t} \cdot \frac{M^2}{n^{1/4}} \cdot \exp\left\{\frac{2M^3t}{Nn^{1/4}}\right\}\right)^t \\ &= o(1). \end{aligned}$$

Let us now condition on the exact lengths of the edges in E_1. The distribution of remaining edges can now **whp** be written as $X'_e = p_2 + Y'_e$ where Y' is chosen uniformly from a simplex Σ' in at least $N' \geqslant N - 10M^2n^{7/4}$ dimensions and with RHS $L' \geqslant N - 10M^3n^{7/4} - Np_2$.

(1) We can now argue very simply: Choose for each $2 \leqslant i \leqslant k$ an edge (a_i, b_i) of cycle C_i. (If $|C_i| = 1$ then $a_i = b_i$). Then divide C_1 into k paths $P_1, \ldots, P_k$ of length $\sim |C_1|/k$. Arguing as in (a1) we can show that **whp**

$$\text{each } a_i \text{ has at least } n_0 = n^{3/4}/(2(\ln n)^3)\ E_1 \cup E_2 \text{ out-neighbors } Q_i \text{ in } P_i. \tag{32}$$

Indeed, fix i and divide P_i into $|P_i|/(2n^{1/4}\ln n) \geqslant n^{3/4}/(2(\ln n)^3)$ disjoint pieces, each of size $\geqslant 2n^{1/4}\ln n$. The (conditional) probability that there is no $(E_1 \cup E_2)$-edge from a_i to any one of these pieces is at most $e^{-2\ln n} = n^{-2}$. This follows by applying Lemma 4.1(a) to Σ'.

Thus (32) holds **whp**. Now further condition on the lengths of the E_2-edges from the a_i to C_1. The lengths of the unconditioned edges are now determined by the uniform selection from a simplex Σ" with $\sim N$ coordinates and $RHS \sim N$. Let R_i be the in-neighbors of the Q_i on C_1. Applying Lemma 4.1(a) once more, we see that

$$\mathsf{P}(\exists i : \textit{ there is no } R_i : b_i \textit{ edge}) \leqslant (\ln n)e^{-n_0p_2/M} = o(1).$$

(2) In summary, **whp** the cost of the patching is $O(p_2 \ln n) = o(1/M)$. Finally, the cost of the minimum tour is $\Omega(1/M)$ **whp**. We can for example show that if we only consider edges of length at most $\varepsilon/(Mn)$ for small constant ε then **whp** at least half of the vertices have out-degree zero. Lemma 4.1(a) shows that the expected number of isolated vertices is $\Omega(n)$. We can then use the Chebyshev inequality to argue that there $\Omega(n)$ isolated vertices **whp**.

6 Discussion

Our work raises several open questions.

0. **Connectivity Threshold.** Is $\ln n/n$ the threshold for connectivity? E.g. prove Conjecture 2.2. An analysis of the second moment raises an interesting question about conditional probabilities of logconcave marginals. Namely, for $X \in \mathbb{R}^N$ drawn from an isotropic down-monotone logconcave density, is it true that

$$\mathsf{P}(X_{k+1} \geqslant p \mid X_1, \ldots, X_k \geqslant p) \leqslant (1 + cp^2)\mathsf{P}(X_{k+1} \geqslant p)$$

for some constant c?

1. **Random graphs with prescribed structure.** We can generate interesting classes of random graphs with prescribed structure. For example, let us consider H-free subgraphs of a fixed graph G. Let $P_H \subseteq [0,1]^{E(G)}$ be defined as follows: Let $H_1, H_2, \ldots, H_s$ be an enumeration of the copies of H in G. Fix some p_0. P_H is the set of solutions to a linear program.

$$\sum_{e \in E(H_i)} X_e > |E(H)|p_0 \quad for\ i = 1, 2, \ldots, s. \tag{33}$$
$$0 \leqslant X_e \leqslant 1, \quad \forall e \in E(G).$$

It is easy to see that G_{P_H,p_0} is H-free, indeed $\sum_{e\in E(H)} X_e \leqslant |E(H)|p_0$ for any H in G_{P_H,p_0}. It would be interesting to analyze important properties of G_{P_H,p_0}. For example, when H is the list of all triangles of the complete graph, we get triangle-free graphs. Similarly when H is a path of length 2, we get matchings (and we can get matchings of any fixed graph by including only the edges as coordinates).

A related question is whether this formulation can be used to generate such H-free graphs uniformly at random. Logconcave distributions can be sampled, but the thresholding process might give a (slightly?) nonuniform distribution.

2. **Thresholds for monotone properties** Do monotone graph properties have sharp thresholds for logconcave densities as they do for Erdős-Rényi random graphs?

3. **Giant Component.** When does $G_{F,p}$ have a giant component? We have barely scratched the surface of this problem.

4. **Smoothed Analysis.** Smoothed Analysis as proposed by Spielman and Teng [21] can be viewed as choosing the costs X uniformly from a unit ball. This is a special case of what we are proposing and it is natural to ask what can be proved about this generalisation, e.g. for Linear Programming.

5. **Hamilton Cycles.** Can we remove the $\frac{\ln\ln\ln n}{\ln\ln\ln\ln n}$ factor from the proof of Theorem 2.4?

6. **Degree Sequence.** This is a fundamental parameter and we know very little about it.

References

[1] K. Ball: Logarithmically concave functions and sections of convex sets in $\mathbb{R}^n$, *Studia Math.* 88, no. 1 (1988) 6984.

[2] K. Ball: Normed spaces with a weak Gordon-Lewis property, *Functional Analysis, Proc. of the Seminar at UT Austin*, Lecture Notes in Math. 1470 (1987-89), 36-47.

[3] J. Bourgain, On high-dimensional maximal functions associated to convex bodies, *American Journal of Mathematics* 108 (1986) 1467-1476.

[4] B. Bollobás: Random Graphs, *Academic Press*, 1985.

[5] A. Dinghas: Über eine Klasse superadditiver Mengenfunktionale von Brunn–Minkowski–Lusternik-schem Typus, *Math. Zeitschr.* **68** (1957), 111–125.

[6] D. Dubhashi and D. Ranjan, Balls and Bins: A Study in Negative Dependence, *Random Structures and Algorithms* 13 (1998) 99-124.

[7] P. Erdős and A. Rényi: On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960) 17-61.

[8] A.M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathemaics* 10 (1985) 47 - 56.

[9] A.M. Frieze and C. McDiarmid, Algorithmic theory of random graphs, *Random Structures and Algorithms* 10 (1997) 5-42.

[10] D. Hefeta, M. Krivelevich and T. Szábo, Hamilton cycles in highly connected and expanding graphs, to appear.

[11] S. Janson, T. Łuczak and A Rucinski: Random Graphs, *Wiley-Interscience*, 2000

[12] R. Kannan, L. Lovász and M. Simonovits, Isoperimetric Problems for Convex Bodies and a Localisation Lemma, *Discrete and Computational Geometry* 13 (1995) 541-559.

[13] R.M. Karp, Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane, *Mathematics of Operations Research*, Mathematics of Operations Research 2 (1977) 209-24.

[14] R.M. Karp and J.M. Steele, Probabilistic analysis of heuristics, in The traveling salesman problem: a guided tour of combinatorial optimization, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys Eds. (1985) 181–206.

[15] L. Leindler: On a certain converse of Hölder's Inequality II, *Acta Sci. Math. Szeged* 33 (1972), 217–223.

[16] L. Lovász and S. Vempala: The geometry of logconcave functions and sampling algorithms, *Random Structures and Algorithms*, 30(3), (2007), 307-358.

[17] L. Lovász and S. Vempala: Fast Algorithms for Logconcave Functions: Sampling, Rounding, Integration and Optimization, *Proc. of FOCS*, (2006), 57-68.

[18] V. D. Milman and A. Pajor: Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space, *Geometric Aspects of Functional Analysis*, Lecture Notes in Math. 1376 (1987-88), 64-104.

[19] A. Prékopa: Logarithmic concave measures and functions, *Acta Sci. Math. Szeged* 34 (1973), 335–343.

[20] A. Prékopa: On logarithmic concave measures with applications to stochasic programming, *Acta Sci. Math. Szeged* 32 (1973), 301–316.

[21] D. Spielman and S. Teng, Smoothed Analysis: Why The Simplex Algorithm Usually Takes Polynomial Time, *Journal of the ACM* 51 (2004) 385 - 463.