Logconcave Random Graphs

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

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

Juan Vera
Department of Management Sciences
University of Waterloo
Waterloo, Ontario
jvera@uwaterloo.ca

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 $1 \leq i, j \leq n$. 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 $X_{ij} \leq p$. The standard Erdős-Rényi model is the special case when F is uniform on the 0-1 unit cube. We determine 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.

Categories and Subject Descriptors: F.2

General Terms: Algorithms.

Keywords: Random Graphs, LogConcave Functions.

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 by Erdős-Rényi [6] 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 $X_{ij} \leq p$. To get a weighted graph, we avoid the thresholding step.

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

STOC'08, May 17–20, 2008, Victoria, British Columbia, Canada.

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 $E_{F,p} = \{ij : X_{ij} \leq p\}$. 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

$$K = \{X \in \mathbb{R}^N : \forall e, X_e \geq 0, \sum_e \alpha_i x_e \leq L\}$$

for some set of coefficients α appears quite interesting by itself and we treat it in detail in Section 1.4. In the weighted graph setting, it corresponds to a random graph with a bound on the total edge weight. In general, F be could be any distribution, but we will consider a further generalization of the cube and simplex, namely, 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 $\lambda \in [0, 1]$,

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

i.e., $\log f$ is concave.

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 [12] and in the context of graph algorithms, the theory of random graphs has played a crucial role (see [8] 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 generalized model

We consider logconcave density functions whose support lies in the positive orthant. Let F be a distribution with such a

density and mean μ. The second moment along each axis, $\sigma_{ij}^2(F)$ will be important. We just use σ_{ij} when F is fixed and simply σ when the standard deviation is the same along every axis.
Fixing only the second moments along the axes allows highly restricted distributions, e.g., the line from the origin to the vector of all 1's. To ensure greater "spread", we require that the density is down-monotone, i.e., for any $x, y \in \mathbb{R}^N$ such that $x \leq y$, we have $f(x) \geq f(y)$. 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 $f(x_1, \dots, x_n)$ 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. We note that sampling such 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., [16]).

1.2 Results

1.3 Logconcave densities.

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 standard deviation σ along every axis, the threshold for connectivity can be narrowed down to within a constant factor.

THEOREM 1.1. *Let F be distribution in the positive orthant with a down-monotone logconcave density and second moment σ^2 along every axis. There exist absolute constants $c_1 < c_2$ such that*

$$\lim_{n\to\infty} \mathsf{P}(G_{F,p} \textit{ is connected}) = \begin{cases} 0 & p < \frac{c_1 \sigma \ln n}{n} \\ 1 & p > \frac{c_2 \sigma \ln n}{n} \end{cases}$$

F being so general makes this theorem quite difficult to prove. It requires several results that are trivial in $G_{n,p}$. Having proven them, it becomes easy to prove similar results e.g.

THEOREM 1.2. *Let F be distribution in the positive orthant with a down-monotone logconcave density and second moment σ^2 along every axis. There exist absolute constants $c_3 < c_4$ such that*

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

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 [9].

THEOREM 1.3. *Let F be distribution in the positive orthant with a down-monotone logconcave density and second moment σ^2 along every axis. There exists an absolute constant c_5 such that if*

$$p \geq \frac{c_5 \sigma \ln n}{n} \cdot \frac{\ln \ln \ln n}{\ln \ln \ln \ln n}$$

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

1.4 Random Graphs from a Simplex

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

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

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

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

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.

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

THEOREM 1.4.

(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 $M \leq (\ln n)^{1/4}$. 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* $\varepsilon > 0$,

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

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

LEMMA 1.5. *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*

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

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

So to prove part (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 [3] or Janson, Łuczak and Ruciński [10]. (We do this explicitly in Section 3.2). 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 1.6. *Let* $\varepsilon > 0$ *be a small positive constant.*

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

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

Let $\mathcal{P}$ be a monotone increasing graph property. p_0 is a threshold for $\mathcal{P}$ if $p/p_0 \to 0$ implies that $\mathsf{P}(G_{\Sigma,p} \in \mathcal{P}) \to 0$ and $p/p_0 \to \infty$ implies that $\mathsf{P}(G_{\Sigma,p} \in \mathcal{P}) \to 1$. It is an open question as to whether every monotone property has a threshold. We can make the following rather weak statement.

THEOREM 1.7. *If* $M = O(1)$ *and* α *is* M*-bounded then every monotone property* $\mathcal{P}$ *has a threshold in the model* $G_{\Sigma,p}$.

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

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

Our next theorem concerns spanning trees. Let Λ_X be weight of the minimum length spanning tree of the complete graph K_n when the edge weights are given by X.

THEOREM 1.8. *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 $\mathsf{E}[\Lambda_X] \sim \zeta(3)$.

We turn our attention next to the diameter of in $G_{\Sigma,p}$.

THEOREM 1.9. *Let* $k \ge 2$ *be a fixed integer. Suppose that* α *is* M*-bounded and for simplicity assume only 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 diam**$(G_{\Sigma,p}) = k$.

1.5 Random Travelling Salesman Problems

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

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

Under this assumption, the distribution of the weights of edges leaving a vertex v is independent of of the particular vertex v. We call this *row symmetry*. We show that a simple patching algorithm based on that in [13] works **whp**.

THEOREM 1.10. *Suppose that the cost matrix* X *of an instance of the ATSP is drawn from a row symmetric simplex where* $M \le n^\delta$, *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.*

2. PROOFS: LOGCONCAVE DENSITIES

We consider logconcave distributions restricted to the positive orthant. We also assume they are down-monotone, i.e., if $x \ge y$ then the density function f satisfies $f(y) \ge f(x)$. We begin by collecting some well-known facts about logconcave densities and proving some additional properties. The new properties will be the main tools for our subsequent analyses and allow us to deal with the nonindependence of edges.

2.1 Properties

The following classical theorem summarizing basic properties of logconcave functions was proved by Dinghas [4], Leindler [14] and Prékopa [18, 19].

THEOREM 2.1. *All marginals as well as the distribution function of a logconcave function are logconcave. The convolution of two logconcave functions is logconcave.*

We will need the following results, Lemmas 5.5(a) and 5.6(a) from [15]: A logconcave function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *isotropic* if (i) it has mean 0 and (ii) its co-variance matrix is the identity.

LEMMA 2.2. *Let* $g : \mathbb{R} \to \mathbb{R}_+$ *be an isotropic logconcave density function. Then* $g(x) \le 1$ *for all* x.

□

LEMMA 2.3. *Let* X *be a random point drawn from a logconcave density function* $g : \mathbb{R} \to \mathbb{R}+$. *For every* $c > 0$,

$$\mathsf{P}(g(X) \le c) \le \frac{c}{M_g}.$$

□

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

LEMMA 2.4. *Let X be a random variable with a non-increasing logconcave density f, support in $\mathbb{R}_+$, $\max f = f(0) = M_f$ and $\mathsf{E}(X^2) = \sigma^2$.*

(a) $\frac{1}{\sqrt{3}\sigma} \leq M_f \leq \frac{2}{\sigma}$.

(b) *For any $p \geq 0$,*

$$\mathsf{P}(x \leq p) \leq pM_f \leq \frac{2p}{\sigma}.$$

(c) *For any $0 \leq p \leq \sigma$,*

$$\mathsf{P}\left(x \leq p\right) \geq \frac{pM_f}{2} \geq \frac{p}{2\sqrt{3}\sigma}.$$

Proof Let g be the symmetric density with $g(x) = g(-x) = \sigma f(\sigma x)/2$. Then g has mean 0 and variance 1. Lemma 2.2(a) implies that $g(x) \leq 1$ andthe upper bound in part (a) follows.
For the lower bound, we claim that M_f is minimized by the constant density on an interval. Let us assume w.l.o.g. that $\sigma = 1$. Suppose f satisfies the conditions of the lemma, has minimum M_f and is not constant. Then we can replace f by another function g that has the same integral as f and $M_g = M_f$ but g is constant. Since this effectively moves mass closer to the origin, the second moment w.r.t. to g is smaller than that w.r.t. f. To make the second moment one, we scale up along the x-axis and scale down along the y-axis (the density). This gives a function with $\sigma = 1$ and smaller maximum, contradicting the assumption. For the constant function, it follows the interval must have length $1/\sqrt{3}$ and therefore $M_f = 1/\sqrt{3}$.
For part (b) use $\mathsf{P}(x \leq p) = \int_{x=0}^{p} f(x)dx \leq pM_f$.
To obtain part (c) let $p_0 = p/3$ and $f(p_0) = \alpha M_f$. As $\ln f$ is concave let $h(x) = c - \gamma x$ be a tangent line to the graph of $\ln f(x)$ at the point $x = p_0$. Then $c = \gamma p_0 + \ln(\alpha M_f)$ and $c - \gamma x \geq \ln f(x)$ for all $x \geq 0$. Considering the secant to $\ln f$ through the points $(0, \ln(M_f))$ and $(p_0, \ln(\alpha M_f))$ we obtain that $\gamma \geq \frac{\ln(1/\alpha)}{p_0}$.
For part (c), we check the value of $f(p)$. If $f(p) \geq M_f/2$, then the claim follows. If not, by Lemma 2.3,

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

and so

$$\mathsf{P}(X \leq p) \geq \mathsf{P}\left(f(X) \geq \frac{M_f}{2}\right) \geq \frac{1}{2} \geq \frac{p}{2\sigma}$$

as required. □

LEMMA 2.5. *Let f be a down-monotone logconcave function in $\mathbb{R}^2$. Let $C = f(0,0)$ and suppose that $p_1, p_2 \geq 0$ are such that $Cp_1p_2 \leq 1/2$. Then*

$$\int_{x_1 \geq 0} \int_{x_2 \geq 0} f(x_1, x_2)\, dx_1\, dx_2 \times \int_{x_1 \geq p_1, x_2 \geq p_2} f(x_1, x_2)\, dx_1\, dx_2 \leq (1 + cp_1p_2) \int_{x_1 \geq p_1, x_2 \geq 0} f(x_1, x_2)\, dx_1\, dx_2 \times \int_{x_1 \geq 0, x_2 \geq p_2} f(x_1, x_2)\, dx_1\, dx_2$$

where c is an absolute constant.

Proof Let

$$F_0 = \int_{x_1 \geq 0} \int_{x_2 \geq 0} f(x_1, x_2)\, dx_1\, dx_2.$$

Let $A_1 = [0, p_1] \times [0, p_2]$, $A_2 = [0, p_2] \times [p_1, \infty]$, $A_3 = [p_1, \infty] \times [0, p_2]$ and $A_4 = [p_1, \infty] \times [p_2, \infty]$. We note that $A_1 \leq Cp_1p_2$. Define

$$F_i = F_0^{-1} \int_{A_i} f(x, y)\, dx\, dy$$

as the integrals in each of these four disjoint regions. Note that $F_1 + F_2 + F_3 + F_4 = 1$. The inequality we want to prove is

$$F_4(F_1 + F_2 + F_3 + F_4) \leq (1 + cp_1p_2)(F_2 + F_4)(F_3 + F_4)$$

which is implied by

$$F_4 F_1 \leq cp_1p_2 F_4(F_2 + F_3 + F_4)$$

and this in turn by,

$$F_1 \leq cp_1p_2(1 - Cp_1p_2).$$

We choose $c = 2C$ so that

$$F_1 \leq Cp_1p_2 \leq cp_1p_2(1 - Cp_1p_2).$$

□

We extend Lemma 2.5 to

LEMMA 2.6. *Let f be a down-monotone logconcave function in $\mathbb{R}^m$ and let $p_1, p_2, \ldots, p_m \geq 0$. Let $\rho = \max_{i,j} p_i p_j$. Let $C = f(0, \ldots, 0)$ and suppose that $C\rho \leq 1/2$. Suppose that $\phi_i = 1_{x_i \geq p_i}$, $i = 1, 2, \ldots, m$ For $S \subseteq [m]$ let*

$$g(S) = \int_{x_1, \ldots, x_m \geq 0} \prod_{i \in S} \phi_i(x_i)\ f(x_1, x_2, \ldots, x_m) \prod_{i=1}^{m} dx_i.$$

Suppose that $t \in [m]$ and $T = [t]$ and $\bar{T} = [m] \setminus T$. Then there exists $c > 0$ such that

$$g(\emptyset) g([m]) \leq (1 + c\rho)^{2m} g(T) g(\bar{T}). \tag{2}$$

Proof We prove the lemma by induction on m. The base case $m = 1$ is trivial and $m = 2$ follows from Lemma 2.5. Applying Lemma 2.5 to the logconcave function

$$h(x_1, x_{t+1}) = \int_{x_i \geq 0, i \neq 1, t+1} f(x_1, \ldots, x_m) \prod_{i \neq 1, t+1} \phi_i(x_i) dx_i$$

we see that if $T_1 = \{2, \ldots, t\}$ and $\bar{T}_1 = \{t+2, \ldots, m\}$ then

$$g(T_1 \cup \bar{T}_1) g([m]) \leq (1 + c\rho) g(T_1 \cup \bar{T}) g(T \cup \bar{T}_1). \tag{3}$$

We can apply Lemma 2.5 inductively to each of the terms on the RHS of (3). We apply it to the coordinates $2, t+1$ and $1, t+2$ respectively. We obtain, with $T_2 = T_1 \setminus \{2\}$, $\bar{T}_2 = \bar{T}_1 \setminus \{t+2\}$,

$$g(T_1 \cup \bar{T}_1) g([m]) \leq (1 + c\rho)^3 \frac{g(T_2 \cup \bar{T}) g(T_1 \cup \bar{T}_1)}{g(T_2 \cup \bar{T}_1)} \frac{g(T_1 \cup \bar{T}_1) g(T \cup \bar{T}_2)}{g(T_1 \cup \bar{T}_2)}$$

which implies

$$\begin{aligned} g([m]) &\le (1+c\rho)^3 \frac{g(T_2\cup\bar{T})}{g(T_2\cup\bar{T}_1)}\frac{g(T\cup\bar{T}_2)}{g(T_1\cup\bar{T}_2)} g(T_1\cup\bar{T}_1) \\ &\vdots \\ &\le (1+c\rho)^{m+1}\frac{g(\bar{T})}{g(\bar{T}_1)}\frac{g(T)}{g(T_1)} g(T_1\cup\bar{T}_1), \end{aligned}$$

where the final inequality is derived by a repeated use of the inequality

$$\frac{g(A\cup B)}{g(A\cup B')} \le \frac{g(A'\cup B)}{g(A'\cup B')}. \quad (4)$$

Here A, B are disjoint and A' and B' are obtained from A, B respectively by deleting a single element.
Inequality (4) follows directly from Lemma 2.5. Now the inequality

$$g(\emptyset)g(T_1\cup\bar{T}_1) \le g(T_1)g(\bar{T}_1)$$

follows from the inductive hypothesis for $\mathbb{R}^{m-2}$ (after integrating over x_1, x_{t+1}). Using this in the previous inequality completes the proof.

□

One can also prove the following generalisation of Lemma 2.6

COROLLARY 2.7. *Let f be a down-monotone logconcave function in $\mathbb{R}^m$. Suppose that ϕ_i, $i = 1, 2, \ldots, m$ are monotone increasing functions. Then*

$$\begin{aligned} \mathsf{E}\left(\prod_{i=1}^m \phi_i(x_i)\right) &\le \\ &(1+c\rho)^{2m}\mathsf{E}\left(\prod_{i=1}^t \phi_i(x_i)\right)\mathsf{E}\left(\prod_{i=t+1}^m \phi_i(x_i)\right) \\ &\le (1+c\rho)^{2m\log_2 m}\prod_{i=1}^m \mathsf{E}\left(\phi_i(x_i)\right). \quad (5) \end{aligned}$$

Proof We can approximate each ϕ_i by a non-negative linear combination of indicator functions and then use linearity of expectation to obtain the result. □
We remark next that using the full power of Lemma 2.7 enables us to prove some strong upper tail bounds. In particular,

LEMMA 2.8. *If $0 \le \varepsilon \le 1$ then*

$$\mathsf{P}(e_{S,p} - \mathsf{E}(e_{S,p}) \ge \varepsilon\mathsf{E}(e_{S,p})) \le (1+c\rho)^{2|S|\log_2|S|} e^{-\varepsilon^2\mathsf{E}(e_{S,p})/3}.$$

□

LEMMA 2.9. *Let $F : \mathbb{R}_+^s \to \mathbb{R}_+$ be a distribution with a down-monotone logconcave density function f with support in the positive orthant. Let $\mathsf{E}(X_i^2) = \sigma_i^2$ for coordinate i and let $\sigma_\Pi = \prod_{i=1}^s \sigma_i$. Let $v = (v_1, \ldots, v_s)$ be the centroid of F. Then $v_i \ge \sigma_i/4$ for all $i \le s$ and $f(v) \ge e^{-A_1 s}/\sigma_\Pi$ for some absolute constant $A_1 > 0$.*

Proof Omitted □

LEMMA 2.10. *Let F be as in Lemma 2.9. Let $\sigma_{\min} = \min\sigma_i$ and $\sigma_{\max} = \max\sigma_i$. Let $G = (V, E)$ be a random graph from $G_{F,p}$ and $S \subseteq V\times V$ with $|S| = s$. Then*

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

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

Proof We consider the projection of F to the subspace spanned by S. Let f_S be the resulting density function. It is logconcave by Theorem 2.1. For a point $x \in \mathbb{R}_+^s$, let $B(x)$ be the positive orthant at x, i.e.,

$$B(x) = \{y \in \mathbb{R}_+^s \,:\, y \ge x\}.$$

Let $g(x)$ be integral of f_S over $B(x)$. Then by Theorem 2.1, g is also logconcave. The function $h(x) = \ln g(x)$ is concave and

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

is nonincreasing. Therefore, it achieves its maximum at $x_i = 0$, i.e.,

$$\frac{\partial h(x)}{\partial x_i} \le \frac{\partial g(0)}{\partial x_i}$$

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

$$-\int_{x_i=0} f_S(x)\,dx \le -\frac{1}{\sqrt{3}\sigma_{\max}}$$

where the inequality is from Lemma 2.4(a). Thus,

$$h(x) \le h(0) - \frac{1}{\sqrt{3}\sigma_{\max}}\sum_{i=1}^s x_i$$

and so

$$g(x) \le e^{-\sum_{i=1}^s x_i/\sqrt{3}\sigma_{\max}}.$$

Setting $x_i = p$, 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 $v = (v_1, \ldots, v_s)$ be the centroid of F_s. Consider the box induced by the origin and v. From Lemma 2.9,

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

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

$$\begin{aligned} g(p,\ldots,p) &\ge g(0)^{1-4p/\sigma} g(\sigma/4,\ldots,\sigma/4)^{4p/\sigma} \\ &= g(\sigma/4,\ldots,\sigma/4)^{4p/\sigma} \ge e^{-A_2 ps/\sigma}. \end{aligned}$$

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 $\mathsf{E}(Y_i^2) = 1$ for all $i \le s$ and

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

□

LEMMA 2.11. *Let F be as in Lemma 2.10. 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*

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

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

Proof 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 2.10. Consider the projection to the span of S and the induced density f_S. From Lemma 2.9, we see that for $p \le \sigma/4$, for any point x with $0 \le x_i \le p$, $f_S(x) \ge (4e^{A_1}\sigma)^{-s}$. The lower bound follows.

For the upper bound, assume $\sigma_{\min} = \sigma_{\max} = s$ 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 theorem of Bobkov and Nazarov [2], its maximum, $g(0) \le c^s$ for an absolute constant c. The bound follows. □

2.2 Proof of Theorem 1.1

For a set S, $|S| = k$, the probability that it forms a component of $G_{F,p}$, is by Lemma 2.10, at most $e^{-a_2 pk(n-k)/\sigma_{\max}}$. Therefore,

$$\mathsf{P}(G \text{ is not connected}) \le \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} e^{-a_2 pk(n-k)/\sigma_{\max}}.$$

It follows that for $p \ge 3\sigma_{\max} \ln n/(a_2 n)$, the random graph is connected with probability $1 - o(1)$.

Assume next that $p \le (1-\varepsilon)\sigma_{\min} \ln n/(a_1(n-1))$ where a_1 is as in Lemma 2.10. Now fix a vertex v. Then we have,

$$\mathsf{P}(v \textit{ is isolated}) \ge n^{\varepsilon - 1}. \tag{6}$$

Now consider two vertices v, w. Then,

$$\begin{aligned}
&\mathsf{P}(v, w \textit{ isolated}) = \qquad (7)\\
&\mathsf{P}(v \textit{ is isolated and } w \textit{ has no edges to } V \setminus \{v\})\\
&\le (1 + O(p^2))^{2n} \mathsf{P}(v \textit{ is isolated})\\
&\times \mathsf{P}(w \textit{ has no edges to } V \setminus \{v\}), \textit{by Lemma } 2.6\\
&\le (1 + o(1)) \mathsf{P}(v \textit{ is isolated})(\mathsf{P}(w \textit{ is isolated}) +\\
&\qquad \mathsf{P}(x_{vw} \le p))\\
&\le (1 + o(1)) \mathsf{P}(v \textit{ is isolated}) \times\\
&\qquad (\mathsf{P}(w \textit{ is isolated}) + 2p/\sigma_{\min}) \quad \textit{from Lemma } 2.4\\
&\le (1 + o(1)) \mathsf{P}(v \textit{ is isolated})(\mathsf{P}(w \textit{ is isolated}) +\\
&\qquad 2 \ln n/n)\\
&= (1 + o(1)) \mathsf{P}(v \textit{ is isolated}) \mathsf{P}(w \textit{ is isolated}). \qquad (8)
\end{aligned}$$

Let Z_1 denote the number of isolated vertices of $G_{F,p}$. It follows from (6) that $\mathsf{E}(Z_1) \ge n^{\varepsilon}$ and from (8) that $\mathsf{E}(Z_1^2) \le \mathsf{E}(Z_1) + (1 + o(1))\mathsf{E}(Z_1)^2 = (1 + o(1))\mathsf{E}(Z_1)^2$. The Chebyshev inequality implies that $Z_1 \neq 0$ **whp**. □

2.3 Proof of Theorem 1.2

Omitted

2.4 Proof of Theorem 1.3

We use the following result from [9]: Let $G = (V, E)$ with n vertices and let $d = d(n) \in [12, e^{\ln^{1/3} n}]$ 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 $|S| \le n_0/d$ then $|N(S) \ge d|S|$. ($N(S)$ 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 $|A|, |B| \ge n_0/4130$.

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

Let $d = \frac{\ln \ln \ln n}{\ln \ln \ln \ln n}$ and $\gamma = \Omega(d/\ln d)$ and $p = \frac{\gamma\sigma \ln n}{n}$ to obtain the theorem

3. PROOFS: SIMPLEX

The following lemma represents a sharpening of Lemmas 2.10 and 2.11 for the simplex case.

LEMMA 3.1.

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

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

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

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

Proof Omitted. □

3.1 Coupling $G_{\Sigma,p}$ and $G_{n,m}$ when $\alpha = 1$: Proof of Lemma 1.5.

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 $G_{n,m}$. Rest of proof is omitted. □

3.2 Connectivity for $G_{\Sigma,p}$ when $\alpha = 1$: Proof of Theorem 1.4 (a)

Omitted

3.3 Connectivity for $G_{\Sigma,p}$: Proof of Theorem 1.4 (b)

Applying Lemma 3.1 we see that for $v, w \in [n]$,

$$\mathsf{P}(v \textit{ is isolated}) = \xi_v(p) \tag{9}$$

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

$$\mathsf{P}(v, w \textit{ are isolated}) = \left(1 - \frac{(\alpha_v + \alpha_w - \alpha_{vw})p}{N}\right)^N. \tag{10}$$

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

$$\frac{1}{2M^2} \ln n \le \alpha_v p_0 \le 2M^2 \ln n \textit{ for all } v \in [n]. \tag{11}$$

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

$$\frac{1}{3M^2} \le a_v \le 3M^2 \textit{ for } v \in [n]. \tag{12}$$

Consider the function

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

We know that $\phi(1) = 1$ and $\phi'(1) = -\ln n \sum_v a_v n^{-a_v} \leq -\ln n/3M^2$. 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 $M = o(\ln n)$, (9) and (10) imply that

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

and then the Chebyshev 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

$$A_1 = \sum_{v\in[n]} \xi_v(p) \leq n^{-\varepsilon/6M^2} \sum_{v\in[n]} \xi_v(p_0) = n^{-\varepsilon/6M^2}.$$

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

$$\begin{aligned} \pi_k &\leq \sum_{|S|=k} \left(1 - \frac{\alpha(S:\bar{S})p}{N}\right)^N \qquad (13) \\ &\leq e^{k^2Mp} \sum_{|S|=k} \exp\left\{-\sum_{v\in S} \alpha_v p\right\} \\ &\leq \frac{e^{k^2Mp} e^{o(k)} A_1^k}{k!} \\ &\leq \left(\frac{e^{kM(1+\varepsilon)(2M^3 \ln n/n)} n^{-\varepsilon/6M^2} e^{1+o(1)}}{k}\right)^k \\ &\leq \left(\frac{e^{1+o(1)} n^{-\varepsilon k/6M^2}}{k}\right)^k \end{aligned}$$

for $k \leq k_0$, after using $p_0 \leq 2M^3 \ln n/n$ from (11). Thus $\sum_{k=1}^{k_0} A_k = o(1)$ and so **whp** there are no components of size $1 \leq k \leq k_0$ in $G_{\Sigma,p}$.
For $k > k_0$ we use

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

Thus **whp** there are no components of size $1 \leq k \leq n/2$ in $G_{\Sigma,p}$. This completes the proof of part (b) of Theorem 1.4. □

3.4 Giant Component in $G_{\Sigma,p}$: Proof of Theorem 1.6

Omitted

3.5 Thresholds: Proof of Theorem 1.7

Omitted

3.6 Diameter of $G_{\Sigma,p}$: Proof of Theorem 1.8

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 $t \leq k-1$ from vertex 1 to vertex 2. We consider the existence of t edges making up a path. Applying Lemma 3.1(b): $S = \emptyset$ and $|T| = k$,

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

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

LEMMA 3.2. *The following hold* **whp**:

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

(b) *If* $S \subseteq V$ *with* $|S| \leq n^{1-\theta-\varepsilon}$ *for some fixed* ε. *Then* $|N(S)| \geq n^\theta|S|/(10M \ln n)$ *where* $N(S)$ *is the set of vertices, not in* S, *that are neighbors of* S.

Proof (a) We consider the existence of $t = 10Mn^\theta$ edges incident with a fixed vertex. Applying Lemma 3.1(b): $S = \emptyset$ and $|T| = \Delta_0$. ($k \geq 3$ is needed here to ensure that $\alpha(T)p = o(1)$).

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

(b) Using Lemma 3.1(a) we see that the probability that this fails to hold can be bounded by

$$\begin{aligned} &\sum_{|S|=1}^{n^{1-\theta-\varepsilon}} \sum_{|T|=0}^{n^\theta s/(10M\ln n)} \left(1 - \frac{|S|(n-|S|-|T|)p}{MN}\right)^N \leq \\ &\sum_{s=1}^{n^{1-\theta-\varepsilon}} \sum_{t=0}^{n^\theta s/(10M\ln n)} n^{s+t} \exp\left\{-s(n-s-t)n^{\theta-1}/M\right\} \\ &\leq \sum_{s=1}^{n^{1-\theta-\varepsilon}} \sum_{t=0}^{n^\theta s/(10M\ln n)} n^{s+t} e^{-sn^\theta/2M} = o(1). \end{aligned}$$

□

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 3.2 that **whp** we have for $1 \leq r \leq r_1$,

$$(n^\theta/(10M\ln n))^r \leq |N_r(v)| \leq (10Mn^\theta)^r.$$

Furthermore, we have $r_0+r_1 \leq k-1$. 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 $\leq k$). 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 3.1(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 $|T| = O(n^{r_1\theta+o(1)})$.

$$\frac{\mathsf{P}((R\cup S)\cap E_p=\emptyset,\, T\subseteq E_p)}{\mathsf{P}(S\cap E_p=\emptyset,\, T\subseteq E_p)} = (1+o(1))(1-\alpha(R)p)^N \le 2e^{-\nu_v\nu_w p/M} = \exp\left\{-\Omega(n^{(r_0+r_1+1)\theta-1-o(1)})\right\}. \quad (14)$$

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

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 3.1(a),

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

By conditioning on $N(v), N(w)$, we argue as in (14) that **whp** every pair of vertices v,w have a common neighbor.

3.7 Minimum Spanning Tree: Proof of Theorem 1.9

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

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

where κ denotes the number of components.

So,

$$\Lambda_X=\int_{p=0}^N(\mathsf{E}[\kappa(G_{\Sigma,p})]-1])dp \quad (15)$$

Going back to (13) (with $M=\omega^2$) we see that

$$\pi_k\le\binom{n}{k}\left(1-\frac{knp}{2\omega^2N}\right)^N\le\left(\frac{ne}{k}\cdot e^{-np/2\omega^2}\right)^k \quad (16)$$

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

So, if $p_0=5\omega^2\ln n\nu$ then

$$p\ge p_0 \textit{ implies } \mathsf{P}[G_{\Sigma,p} \textit{ is not connected}]=o(N^{-2}).$$

So,

$$\Lambda_X=\int_{p=0}^{p_0}(\mathsf{E}[\kappa(G_{\Sigma,p})]-1])dp+o(N^{-1}). \quad (17)$$

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

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

It follows that

$$\Lambda_X\ge\frac{1}{2\omega^2}. \quad (18)$$

Using Lemma 3.1(b) to tighten (16), we see that for $k\le n^{1/2}$ and $p\le p_0$,

$$\mathsf{E}[\kappa_{k,p}]\le\frac{1}{\omega^2p}\left(ne\cdot\omega^2pe^{-np/2\omega^2}\right)^k. \quad (19)$$

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

$$\int_{p=p_1}^{p_0}(\mathsf{E}[\kappa_{k,p}]-1)dp\le\frac{1}{\omega^{k+2}}.$$

It follows from (17) and (18) that

$$\Lambda_X=\sum_{k=1}^{\omega^5}\int_{p=0}^{p_1}\mathsf{E}[\kappa_{k,p}]dp, \quad (20)$$

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 $q=e^{-Dp}$, for $k\le\omega^5$

$$\mathsf{E}[\tau_{k,p}]\sim(k-2)!p^{k-1}\sum_{a\in A_k}\sum_{\substack{f:[k]\to V\\ f\ an\ injection}}\prod_{j=1}^k\frac{d_{f(j)}^{a_j}q^{d_{f(j)}}}{(a_j-1)!} \quad (21)$$

Putting $d_S=\sum_{v\in S}d_v$, this can be re-expressed

$$\begin{aligned}\mathsf{E}[\tau_{k,p}] &\sim (k-2)!p^{k-1}\sum_{a\in A_k}\prod_{i=1}^k\sum_{v=1}^n\frac{d_v^{a_i}q^{d_v}}{(a_i-1)!}\\ &\sim (k-2)!p^{k-1}[x^{2k-2}]\left(\sum_{v=1}^n\sum_{r=1}^\infty\frac{q^{d_v}d_v^r}{(r-1)!}x^r\right)^k\\ &= (k-2)!p^{k-1}[x^{k-2}]\left(\sum_{v=1}^n q^{d_v}d_ve^{d_vx}\right)^k\\ &= (k-2)!p^{k-1}\sum_{\substack{S\subseteq V\\|S|=k}}q^{d_S}\frac{d_S^{k-2}}{(k-2)!}\prod_{v\in S}d_v. \end{aligned} \quad (22)$$

So,

$$\sum_{k=1}^{\omega^5}\int_{p=0}^{p_1}\mathsf{E}[\tau_{k,p}]dp\sim\sum_{k=1}^\infty\frac{(k-1)!}{D^k}\sum_{\substack{S\subseteq V\\|S|=k}}\frac{\prod_{v\in S}d_v}{d_S^2} \quad (23)$$

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

$$\sum_{k=1}^{\omega^5}\int_{p=0}^{p_1}\mathsf{E}[\sigma_{k,p}]dp=o(\omega^{-2}). \quad (24)$$

But, arguing as in (19) we see that for $k\le n/2$,

$$\mathsf{E}[\sigma_{k,p}]\le\left(ne\cdot\omega^2pe^{-np/2\omega^2}\right)^k.$$

Hence,

$$\sum_{k=1}^{\omega^5}\int_{p=0}^{p_1}\mathsf{E}[\sigma_{k,p}]dp\le\sum_{k=1}^{\omega^5}(2e\omega^4)^kp_1=n^{o(1)-1}$$

and (24) follows. □

4. TSP ALGORITHM: PROOF OF THEOREM 1.10

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 $|C_1| \geq |C_2| \geq \cdots \geq |C_k|$.
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 put a 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 3.1(a) we see that

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

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 $L \geq 2$ then the failure probability is sufficient to give the result for all v.

(2) We use Lemma 3.1(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 $|S| \leq n^{2/3}$ and $|T| \leq L|S|\ln n/4$ then **whp** the induced bipartite sub-graph on $S \cup T$ contains at most $L|S|\ln n/2$ edges of length $\leq p_1$. Indeed, if $\mathcal{B}$ is the event that there are S,T with more edges, then

$$\mathsf{P}(\mathcal{B}) \leq \tag{25}$$

$$(1+o(1))\sum_{s=1}^{n^{2/3}}\sum_{t=1}^{Ls\ln n/4}\binom{n}{s}\binom{n}{t}\binom{st}{Ls\ln n/2} \times \left(\frac{KM^2(\ln n)^2}{n}\right)^{Ls\ln n/2} \tag{26}$$

$$= o(1). \tag{27}$$

(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 $\leq p_1$. Using (1) and (2) we see that the levels grow at a rate $L\ln n/5$ 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 (1) and (2) that $|S_{i+1}| \geq L|S_i|\ln n/5$, as long as $|S_i| \leq n^{2/3}$. So **whp** there exists i_0 such that $|S_{i_0}| \geq n^{3/5}$. 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 $|T_{j_0}| \geq n^{3/5}$.

We can then use Lemma 3.1(a) to argue that **whp** there is an edge of length at most $Mn^{-1/2}$ joining the final two levels S,T. Indeed

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

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

(c): 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 $\leq p_2$ 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** $|E_1| \leq 10M^2n^{7/4}$. Indeed, if $t = 10M^2n^{7/4}$ then

$$\mathsf{P}(|E_1| \geq t) \leq \binom{N}{t}M^t\left(\frac{M}{n^{1/4}}\right)^t \exp\left\{\frac{2M^3t^2}{Nn^{1/4}}\right\} \leq \left(\frac{Ne}{t}\cdot\frac{M^2}{n^{1/4}}\cdot\exp\left\{\frac{2M^2t}{Nn^{1/4}}\right\}\right)^t = o(1).$$

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 $N' \geq N - 10M^2n^{7/4}$ dimensions and with RHS $L' \geq N - 10M^3n^{7/4} - Np_2$.

(1) We can now argue very simply: Choose for each $2 \leq i \leq k$ 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**

$$\text{each } a_i \text{ has at least } n_0 = n^{3/4}/(2(\ln n)^3)\ E_2 \text{ out-neighbors } Q_i \text{ in } P_i. \quad (28)$$

As a check, fix i and divide P_i into $|P_i|/(2n^{1/4}\ln n) \geq n^{3/4}/(2(\ln n)^3)$ disjoint pieces, each of size $\geq 2n^{1/4}\ln n$. The probability that there is no E_2-edge from a_i to any one of these pieces is at most $e^{-(2-o(1))\ln n} = n^{-2+o(1)}$. This follows by applying Lemma 3.1(a) to Σ'.

Thus (28) holds **whp**. Now condition on the lengths of the E_{2p}-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 $RSS \sim N$. Let R_i be the in-neighbors of the Q_i on C_1. Applying Lemma 3.1(a) once more, we see that

$$\mathsf{P}(\exists i : \textit{there is no } R_i, b_i \textit{ edge}) \leq (\ln n)e^{-n_0 p_2/M} = o(1).$$

(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 3.1(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**.

5. OPEN QUESTIONS

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.

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

G_{P_H, p_0} is H-free and it would be interesting to analyze important properties of G_{P_H, p_0}. We can for example generate triangle-free graphs. 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).

Can we uniformly generate H-free graphs in this way?

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 [20] 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 see 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 1.3.

6. **Degree Sequence** This is an important parameter but we know relatively little about it.

6. REFERENCES

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

[2] S. G. Bobkov and F. L. Nazarov: On convex bodies and log-concave probability measures with unconditional basis, *Geom. Aspects of Funct. Analysis*, (Milman-Schechtman eds.), Lecture Notes in Math. 1807 (2003), 53-69.

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

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

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

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

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

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

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

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

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

[12] 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.

[13] 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.

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

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

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

[17] 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 Mathematics 1376 (1987-88), 64-104.

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

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

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