A note on log-concave random graphs

Alan Frieze*    Tomasz Tkocz
Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh PA 15213,
U.S.A.

Submitted: Jun 22, 2018; Accepted: Jul 24, 2019; Published: Aug 30, 2019
© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

We establish sharp threshold for the connectivity of certain random graphs whose (dependent) edges are determined by uniform distributions on generalized Orlicz balls, crucially using their negative correlation properties. We also show existence of a unique giant component for such random graphs.

Mathematics Subject Classifications: 05C80

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 [2], [3] has given us numerous insights, surprises and techniques and has been used to count, to establish structural properties and to analyze algorithms. There are by now several texts [1], [6], [4] that deal exclusively with the subject. The most heavily studied models being $G_{n,m}$ and $G_{n,p}$. Both have vertex set $[n]$ and in the first we choose $m$ random edges and in the second we include each possible edge independently with probability $p$.

Let $X$ be a random vector in $[0, \infty)^{n \choose 2}$ with a log-concave down-monotone density $f$, that is (i) $\log f$ is concave and (ii) $f(x) \geq f(y)$ if $x \leq y$ (coordinate-wise). For $0 < p < 1$, let $G_{X,p}$ be a random graph with vertices $1, \ldots, n$ and edges determined by $X$: for $1 \leq i < j \leq n$, $\{i,j\}$ is an edge if and only if $X_{i,j} \leq p$. Such log-concave random graphs were introduced by Frieze, Vempala and Vera in [5]. For instance, when $X$ is uniform on $[0,1]^{n \choose 2}$, $G_{X,p}$ is the random graph $G_{n,p}$.

The paper [5], among other things, studied the connectivity of $G_{X,p}$ and found a logarithmic gap for the threshold. There is no gap when $G_{X,p}$ is defined by uniform

*Research supported in part by NSF Grant DMS1661063
sampling from a “well-behaved” regular simplex\textsuperscript{1} and we extend this case to Generalized Orlicz Balls GOBs: that is sets of the form \( \{ x \in \mathbb{R}^d : \sum_{i=1}^d f_i(|x_i|) \leq 1 \} \) for some nondecreasing lower semicontinuous convex functions \( f_1, \ldots, f_d : [0, \infty) \to [0, \infty] \) with \( f_i(0) = 0 \), which are not identically 0 or \( +\infty \) on \( (0, \infty) \) (such functions are sometimes called Orlicz functions). Note that we allow \( f_i \) to take the value \( +\infty \); for instance, the cube \([0, 1]^d\) is a GOB with
\[
f_1(t) = \cdots = f_d(t) = \begin{cases} 
0, & t \in [0, 1]. \\
+\infty, & t \in (1, \infty).
\end{cases}
\]

The key property of Orlicz balls is negative correlation. We say that a random vector \( X \) in \( \mathbb{R}^d \) has negatively correlated coordinates if for any disjoint subsets \( I, J \) of \( \{1, \ldots, d\} \) and nonnegative numbers \( s_i, t_j \), we have
\[
P(\forall i \in I |X_i| > s_i, \forall j \in J |X_j| > t_j) \leq P(\forall i \in I |X_i| > s_i)P(\forall j \in J |X_j| > t_j).
\]
It was shown in [8] that this property holds for random vectors uniformly distributed on GOBs (see also [9] for a first such result treating two coordinates and [10] for a simpler proof of the general result).

Notation: Throughout the paper we will let \( \sigma_{\min} \) and \( \sigma_{\max} \) be defined by
\[
\sigma_{\min}^2 = \min_{1 \leq i < j \leq n} \mathbb{E} X_{i,j}^2 \quad \text{and} \quad \sigma_{\max}^2 = \max_{1 \leq i < j \leq n} \mathbb{E} X_{i,j}^2.
\]

Our result concerning connectivity is the following theorem.

\textbf{Theorem 1.} Let \( X = (X_{i,j})_{1 \leq i < j \leq n} \) be a log-concave random vector in \( [0, \infty)\binom{n}{2} \) with a down-monotone density and negatively correlated coordinates.

(a) For every \( \delta \in (0, 1) \), there are constants \( c_1 \) and \( c_2 \) dependent only on \( \delta \) such that for \( p < c_1 \sigma_{\min} \log n \), we have
\[
P(G_{X,p} \text{ has isolated vertices}) > 1 - c_2 n^{-\delta}.
\]

(b) For every \( \delta \in (0, 1) \), there are constants \( C_1 \) and \( C_2 \) dependent only on \( \delta \) such that for \( p > C_1 \sigma_{\max} \log n \), we have
\[
P(G_{X,p} \text{ is connected}) > 1 - C_2 n^{-\delta}.
\]

We will also discuss the existence of a giant component for smaller values of \( p \).

Notation: Let
\[
M = \max_T \sup_{y \in [0, \infty]^{|T|}} \max_{(i,j) \notin T} \mathbb{E}(X_{i,j}^2|X_T = y),
\]
where the first maximum is over all nonempty subsets \( T \) of the index set \( \{(i,j), 1 \leq i < j \leq n\} \) and we denote \( X_T = (X_{i,j})_{(i,j) \in T} \).

\textsuperscript{1}A regular simplex \( \{ x \in \mathbb{R}^d : a \cdot x \leq 1 \} \) for some \( a \geq 0 \) if \( a_i/a_j \leq K \) for some not too large \( K \).
Theorem 2. Let $X = (X_{i,j})_{1 \leq i < j \leq n}$ be a log-concave random vector in $[0, \infty)^{\binom{n}{2}}$ with a down-monotone density. Assume that $M = O(1)$. There are constants $c_1$ and $c_2$ such that for every $\beta > 1$, we have

(i) If $p < \frac{c_1 \sigma_{\min}}{n}$, then

$$P(G_{X,p} \text{ has a component of order } \geq \beta \log n) < \frac{12}{n^{\beta - 1}}.$$

(ii) If $p > \frac{c_2 M \log \left( \frac{M \sigma_{\min}}{n} \right)}{n}$, then

$$P(G_{X,p} \text{ has a component of order } \in [\beta \log n, n/2]) < \frac{1}{n^{\beta - 1}}$$

and

$$P(G_{X,p} \text{ has a unique giant component of order } > n/2) > 1 - \frac{5 \beta + (\beta - 1)^{-1}}{\log n}.$$

Note that we have dropped the assumption of negative correlation.

Remark 3. For our theorem on the existence of a giant component we need to have $M = O(1)$. For instance, for a GOB $\{ x \in \mathbb{R}^{\binom{n}{2}} : \sum_{1 \leq i < j \leq n} f_{i,j}(|x_{i,j}|) \leq 1 \}$, this assumption is satisfied in a fairly general situation. Let

$$a_{i,j} = \sup \{ t > 0 : f_{i,j}(t) \leq 1 \}.$$

Since the $f_{i,j}$ are Orlicz functions, the $a_{i,j}$’s are finite. Furthermore, $M \leq \max_{i,j} a_{i,j}^2$ and so $\max_{i,j} a_{i,j} = O(1)$ implies that $M = O(1)$.

2 Preliminaries

We shall use several results from [5], which we now include for convenience. Let $X$ be a random vector in $[0, \infty)^{\binom{n}{2}}$ with a log-concave down-monotone density $f$. Let $G = (V, E)$ be a $G_{X,p}$ random graph. Lemma 3.1 and 3.2 from [5] say that there are universal constants $a_1, a_2, b_1, b_2$ such that for $S \subset V \times V$ with $|S| = s$, we have

$$e^{-a_1 ps/\sigma_{\min}} \leq P(S \cap E = \emptyset) \leq e^{-a_2 ps/\sigma_{\max}}, \quad (2)$$

and

$$\left( \frac{b_1 p}{\sigma_{\max}} \right)^s \leq P(S \cap E = \emptyset) \leq \left( \frac{b_2 p}{\sigma_{\min}} \right)^s, \quad (3)$$

where both lower bounds require $p \leq \frac{\sigma_{\min}}{4}$.

Moreover, Lemma 3.5 from [5] asserts that for a nonnegative random variable $X$ with a nonincreasing log-concave density, for any $0 \leq p \leq \sqrt{\mathbb{E}X^2}$,

$$P(X \leq p) \leq \frac{p}{\sqrt{\mathbb{E}X^2}}. \quad (4)$$
Recalling the proof of a part of Theorem 2.1 from [5], for a set \( S \) with \( |S| = k \), the probability that it forms a component of \( G_{X,p} \) is by (2) at most \( e^{-a_2pk(n-k)/\sigma_{\text{max}}} \) and thus

\[
P(G_{X,p} \text{ is not connected}) \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} e^{-a_2pk(n-k)/\sigma_{\text{max}}}. \tag{5}
\]

### 3 Connectivity: Proof of Theorem 1

**Proof.** Part (b) follows directly from (5). For (a), we adapt the standard second moment argument used for the Erdős-Rényi model. For \( 1 \leq i \leq n \), let \( Y_i \) be equal to 1 if the vertex \( i \) is isolated and 0 otherwise. Let \( Y = Y_1 + \ldots + Y_n \) be the number of isolated vertices.

We have,

\[
P(G_{X,p} \text{ has isolated vertices}) = P(Y > 0) \geq \left( E[Y] \right)^2/E[Y]^2.
\]

Thus, if we show that \( E[Y]^2 \leq (1 + \varepsilon)(E[Y])^2 \), then \( P(Y > 0) \geq 1 - \varepsilon \). Clearly,

\[
E[Y]^2 = \sum_k E[Y_k^2] \geq \sum_k E[Y_k] = \sum_k \sum_{k \neq l} P(Y_k = 1 = Y_l) = \sum_k \sum_{k \neq l} P(Y_k = 1 = Y_l)
\]

and our goal is to show that

\[
E[Y] \leq \frac{\varepsilon}{2}(E[Y])^2 \quad \text{and} \quad \sum_{k \neq l} P(Y_k = 1 = Y_l) \leq \left( 1 + \frac{\varepsilon}{2} \right)(E[Y])^2.
\]

From the negative correlation of coordinates of \( X \) as well as an elementary inequality \( P(A) \leq P(A \cap B) + 1 - P(B) \), we get

\[
P(Y_k = 1 = Y_l) = P(\forall i \neq k, X_{ik} > p, X_{il} > p, X_{kl} > p)
\]

\[
\leq P(\forall i \neq k, X_{ik} > p)P(\forall i \neq k, l X_{il} > p)
\]

\[
\leq P(\forall i \neq k, X_{ik} > p)[P(\forall i \neq l X_{il} > p) + 1 - P(X_{kl} > p)]
\]

\[
= P(Y_k = 1)[P(Y_l = 1) + P(X_{kl} \leq p)].
\]

By (4), \( P(X_{kl} \leq p) \leq \frac{p}{\sigma_{\text{min}}} \) (recall that by the Prékopa-Leindler inequality, marginals of log-concave vectors are log-concave; clearly, marginals of down-monotone densities are down-monotone). Therefore,

\[
\sum_{k \neq l} P(Y_k = 1 = Y_l) \leq \sum_{k \neq l} P(Y_k = 1)P(Y_l = 1) + \sum_{k \neq l} P(Y_k = 1)\frac{p}{\sigma_{\text{min}}}
\]

\[
\leq \left( \sum_k P(Y_k = 1) \right)^2 + \frac{np}{\sigma_{\text{min}}} \sum_k P(Y_k = 1)
\]

\[
\leq \left( 1 + \frac{np}{\sigma_{\text{min}}E[Y]} \right)(E[Y])^2 < \left( 1 + \frac{c_1 \log n}{E[Y]} \right)(E[Y])^2,
\]
so it suffices to take $\varepsilon$ such that

$$
\varepsilon \geq \frac{2}{EY} \quad \text{and} \quad \varepsilon \geq \frac{2c_1 \log n}{EY}.
$$

By (2), $P(Y_k = 1) \geq e^{-apn/\sigma_{\min}}$, for some universal constant $a$ (the assumption $p < \frac{1}{4} \sigma_{\min}$ of that lemma is clearly satisfied if $p < c_1 \sigma_{\min}^{\log n/n}$), so

$$
EY = \sum_k P(Y_k = 1) \geq ne^{-apn/\sigma_{\min}} > n^{1-ac_1}.
$$

Thus, $\varepsilon = c_2 n^{ac_1-1} \log n$ will suffice. \hfill \Box

4 Giant Component: Proof of Theorem 2

Lemma 4. Let $X = (X_{i,j})_{1 \leq i < j \leq n}$ be a log-concave random vector in $[0, \infty)^{(n \choose 2)}$ with a down-monotone density. There are universal constants $a$ and $b$ such that for $S,T \subset \{(i,j), \ 1 \leq i < j \leq n\}$ and $p > 0$, we have

$$
P(\forall s \in S X_s > p, \forall t \in T X_t \leq p) \leq e^{-ap|S|M} \left( \frac{bp}{\sigma_{\min}} \right)^{|T|}.
$$

Proof. Fix disjoint sets $S,T \subset \{(i,j), \ 1 \leq i < j \leq n\}$ (if they are not disjoint, the probability in question is 0) and $y \in [0, \infty)^{|T|}$. Let $f$ be the density of $(X_S, X_T)$. The conditional density of the vector $X_S$ given $X_T = y$,

$$
f_{X_S|X_T}(x|y) = \frac{f(x,y)}{\int f(x',y)dx'}
$$

is down-monotone and log-concave. Therefore, by (2),

$$
P(\forall s \in S X_s > p|X_T = y) \leq e^{-ap|S|M}.
$$

We denote the density of $X_T$ by $f_{X_T}$ and get

$$
P(\forall s \in S X_s > p, \forall t \in T X_t \leq p) = \int_{[0,p]^{|T|}} P(\forall s \in S X_s > p|X_T = y)f_{X_T}(y)dy
$$

$$
\leq \int_{[0,p]^{|T|}} e^{-ap|S|M} f_{X_T}(y)dy
$$

$$
= e^{-ap|S|M} P(\forall t \in T X_t \leq p)
$$

$$
\leq e^{-ap|S|M} \left( \frac{bp}{\sigma_{\min}} \right)^{|T|},
$$

where the final inequality follows directly from (3). \hfill \Box
With this lemma in hand, we can prove Theorem 2.

Proof. Let $Z_k$ be the number of components of order $k$ (that is, on $k$ vertices) in $G_{X,p}$. As for the Erdős-Rényi model, looking at a spanning tree for each component and bounding the corresponding in-out edge probabilities using Lemma 4 yields

$$
\mathbb{E}Z_k \leq \left( \frac{n}{k} \right)^{k-2} e^{-apk(n-k)/M} \left( \frac{bp}{\sigma_{\min}} \right)^{k-1}
$$

$$
\leq \left( \frac{en}{k} \right)^{k} e^{-apk(n-k)/M} \left( \frac{bp}{\sigma_{\min}} \right)^{k-1}
$$

$$
= \frac{\sigma_{\min}}{bp} \frac{1}{k^2} \left[ \frac{eb}{\sigma_{\min}^2} \left( \frac{en}{bp} \right)^{n-k} \right]^k.
$$

If $p = \frac{M}{a} \frac{c}{n}$, with $c$ being a constant (chosen soon), this becomes

$$
\mathbb{E}Z_k \leq \frac{e}{A} \frac{1}{c} \frac{n}{k^2} \left[ Ace^{-ck/n} \right]^k,
$$

where we put $A = \frac{eb}{\sigma_{\min}}$.

Case 1. If $c$ is a small constant, say $c \leq \frac{1}{eA}$ (equivalently, $p \leq \frac{\sigma_{\min}}{e^{1+bp} n}$), then we bound $e^{-ck/n}$ crudely by 1 and get that

$$
\mathbb{E}Z_k \leq \frac{e}{A} \frac{1}{c} \frac{n}{k^2} (Ac)^k \leq en(Ac)^{k-1} \leq e^2 ne^{-k}.
$$

Thus,

$$
\mathbb{E} \left( \sum_{k \geq \beta \log n} Z_k \right) \leq e^2 n \cdot \sum_{k \geq \beta \log n} e^{-k} \leq e^2 ne^{-\beta \log n} \cdot \frac{1}{1-e^{-1}} = \frac{e^2}{e-1} \frac{1}{n^{\beta-1}} < \frac{12}{n^{\beta-1}}.
$$

By the first moment method, this gives (i).

Case 2. Let $c$ be a large constant, say such that $Ace^{-c/2} \leq \frac{1}{e}$ and $Ac \geq e^2$, which holds when, say $c \geq 4 \log A$, provided that $A$ is large enough, which leads to the assumption on $p$ in (ii). Then for $k \leq n/2$, we have

$$
\mathbb{E}Z_k \leq \frac{en}{Ac} (Ace^{-c/2})^k \leq ne^{-k}.
$$

Thus,

$$
\mathbb{E} \left( \sum_{\beta \log n \leq k \leq n/2} Z_k \right) \leq ne^{-1} \sum_{k \geq \beta \log n} e^{-k} \leq \frac{1}{e-1} \frac{1}{n^{\beta-1}} < \frac{1}{n^{\beta-1}}.
$$

By the first moment method, this gives the first part of (ii).
To go about the second part and show that there is a giant component, we shall simply count the number of vertices on the small components and show that with high probability, there are strictly less $n$ such vertices. The uniqueness of a giant component plainly follows from the fact that it has more than $n/2$ vertices, so there cannot be more than one such components. Fix $1 \leq k \leq \beta \log n$ and set $t = ne^{-k-1}$. For any positive integer $l \leq et + 1$, we have

$$P(Z_k \geq et) \leq P(Z_k(Z_k - 1) \ldots (Z_k - l + 1) \geq et(et - 1) \ldots (et - l + 1))$$
$$\leq \frac{E Z_k(Z_k - 1) \ldots (Z_k - l + 1)}{et(et - 1) \ldots (et - l + 1)}$$

As for the upper bound for $E Z_k$, looking at spanning trees for each $l$-tuple of distinct components of order $k$ and bounding the corresponding in-out edge probabilities using Lemma 4 yields

$$E Z_k(Z_k - 1) \ldots (Z_k - l + 1) \leq \binom{n}{k} \binom{n-k}{k} \ldots \binom{n-(m-1)k}{k} (k^{k-2})^l e^{-\frac{m}{k}kl(n-kl)} \left( \frac{bp}{\sigma_{\min}} \right)^{(k-1)l}$$

Provided that $kl \leq n/2$, under our assumption $c \geq 4 \log A$, this is further upper bounded by $(t/k^2)^l$, which gives

$$P(Z_k \geq ne^{-k}) = P(Z_k \geq et) \leq \frac{1}{k^{2l}} \left( \frac{t}{et - l + 1} \right)^l.$$ For $k \geq \frac{1}{2} \log n$, we choose $l = 1$ and get

$$P(Z_k \geq ne^{-k}) \leq \frac{1}{e \left( \frac{1}{2} \log n \right)^2}, \quad k \geq \frac{1}{2} \log n.$$ For $k < \frac{1}{2} \log n$, we have $t = ne^{-k-1} > e^{-1} \sqrt{n}$, so choosing, say $l - 1 = \lfloor e^{-1} \sqrt{n} \rfloor$ yields

$$P(Z_k \geq ne^{-k}) \leq \left( \frac{t}{et - \lfloor e^{-1} \sqrt{n} \rfloor} \right)^l = \left( \frac{1}{e - \lfloor e^{-1} \sqrt{n} \rfloor} \right)^l \leq \left( \frac{1}{e - 1} \right)^l \leq \left( \frac{1}{e - 1} \right)^{e^{-1} \sqrt{n}}, \quad k < \frac{1}{2} \log n.$$
Combining the last two estimates, the union bound gives that the probability of the event $E = \{ \exists k \leq \beta \log n, \ Z_k \geq ne^{-k} \}$ is at most

$$\frac{4}{e} \frac{(\beta - 1/2) \log n + 1}{(\log n)^2} + \frac{1}{2} \frac{\log n}{(e - 1)e^{-1/\sqrt{n}}} < \frac{5\beta}{\log n}$$

(we check that $\frac{1}{2} \frac{\log n}{(e - 1)e^{-1/\sqrt{n}}} < \frac{2}{\log n}$ and simply bound $\frac{4}{e} \frac{(\beta - 1/2) \log n + 1}{(\log n)^2} \leq \frac{4\beta + 2}{\log n}$). To finish, it remains to check that on $E^c$, there are few vertices on the small components. On $E^c$, we have

$$\sum_{k \leq \beta \log n} kZ_k \leq n \sum_{k \leq \beta \log n} ke^{-k} < n \sum_{k=1}^{\infty} ke^{-k} = n \frac{e}{(e - 1)^2} < 0.93n.$$

**Remark 5.** It was shown in [10] that the negative correlation property holds in fact for random vectors with densities of the form $h(\sum f_i(x_i))$, where $h : [0, \infty) \to [0, \infty)$ is a nonincreasing log-concave function ($h = 1_{[0,1]}$ giving uniform densities on GOBs). For such densities, $M$ is finite and can be bounded as for GOBs in terms of certain parameters depending on the functions $f_i$ and $h$.

## 5 Conclusion and Open Questions

We have successfully generalised the results on the regular simplex in [5] to GOBs. The following questions seem most apposite.

**Q1** What we prove in Theorem 2 does not rule out the possibility that in some range of $p$ there is more than one giant component. Can the proof be tightened to rule this out?

**Q2** What is the connectivity or giant component threshold for the intersection of two well-behaved regular simplices?

**Q3** What is the connectivity or giant component threshold for the intersection of a few regular simplices with independent randomly chosen coefficients?

Another, more geometric, direction to pursue is to determine which log-concave random vectors have negatively correlated coordinates. Wojtaszczyk conjectured that this holds for uniform distributions on 1-symmetric convex bodies (that is, bodies symmetric with respect to each hyperplane $x_i = 0$ as well as permuting the coordinates, which form a natural class generalising GOBs). In [7] it was confirmed in the infinitesimal case; see therein for a further discussion of Wojtaszczyk’s conjecture.

**References**
