Far-out Vertices In Weighted Repeated Configuration Model

B. Błaszczyszyn 
INRIA/ENS 
23 avenue d’Italie 
Paris, France 
Bartek.Blaszczyszyn@ens.fr

K. Gaurav 
UPMC/INRIA/ENS 
23 avenue d’Italie 
Paris, France 
kumar.gaurav@inria.fr

ABSTRACT
We consider an edge-weighted uniform random graph with a given degree sequence (Repeated Configuration Model) which is a useful approximation for many real-world networks. It has been observed that the vertices which are separated from the rest of the graph by a distance exceeding certain threshold play an important role in determining some global properties of the graph like diameter, flooding time etc., in spite of being statistically rare. We give a convergence result for the distribution of the number of such far-out vertices. We also make a conjecture about how this convergence result for the distribution of the number of such vertices and the asymptotic distribution of the longest edge etc., in spite of being statistically rare. We give a conjecture about how this relates to the longest edge of the minimal spanning tree on the graph under consideration.

1. INTRODUCTION
Given a degree sequence $d^{(n)} = (d_i^{(n)})^n_{i=1}$ for $n$ vertices labelled $1$ to $n$. Repeated Configuration Model, denoted $RCM[n]$ (or $RCM[(d^{(n)})]$) when talking about multiple models each with different degree sequence), gives a uniform random graph with vertex $i$ having degree $d_i$. Every edge $e$ of $RCM[n]$ is given an i.i.d. edge-length $Y(e)$, with cdf $F$, to get a weighted random graph which we denote by $RCM[n]$.
Given $x_n > 0$, a vertex $i$ in $RCM[n]$ is said to be far-out if the distance of $i$ from its nearest neighbour exceeds $x_n$. That is, if we let $M_n(i) := \min_{j \neq i} Y(e_{ij})$, $i$ is far-out if $M_n(i) > x_n$. In this work, we study the asymptotic distribution of the number of far-out vertices in $RCM[n]$ as $x_n$ is appropriately scaled with $n$. This study is motivated by the following observations:

- In the study of the diameter of $RCM[n]$ in [3], the existence of far-out vertices plays a critical role in determining the difference between the expected distance between two uniformly chosen vertices and the diameter of the graph.
- Penrose shows in [4] that for $n$ points placed uniformly at random in the unit square, the longest edge of the Minimum Spanning Tree (MST) and the longest edge of the Nearest Neighbour Graph (NNG) on these points are asymptotically the same. In other words, the most far-out vertex in the complete graph on these $n$ points determines the longest edge of the MST of this complete graph.

The second observation suggests a conjecture that the longest edge of the MST on $RCM[n]$ would be determined by the most far-out vertex in $RCM[n]$. However in the present abstract, we limit ourselves to presenting the results on the asymptotic distribution of the number of far-out vertices and the asymptotic distribution of the longest edge of the NNG of $RCM[n]$.

2. RESULTS
For a precise statement of the results, we start by assuming a set of conditions on $d^{(n)} = (d_i^{(n)})^n_{i=1}$ given in chapter 7 of [5], which apart from being indispensable for convergence results, enables the construction of $RCM[n]$ for arbitrary large $n$.

CONDITION 1. For each $n$, $d^{(n)} = (d_i^{(n)})^n_{i=1}$ is a sequence of non-negative integers such that $\sum_{i=1}^n d_i^{(n)}$ is even. For $k \in \mathbb{N}$, let $u_k^{(n)} = |\{i : d_i^{(n)} = k\}|$ and $D_n$ be the degree of a uniformly chosen vertex in $RCM[n]$, i.e., $\mathbb{P}(D_n = k) = u_k^{(n)}/n$. Let $D$ be some random variable taking value over strictly positive integers with probability distribution $\{p_r\}_{r \in \mathbb{N}}$. Then the following hold.

1. The degree density condition: $u_k^{(n)}/n \rightarrow p_k$ or equivalently, $D_n \rightarrow D$ in distribution.

2. Convergence of empirical expectations: $\sum_{i=1}^n d_i^{(n)}/n \rightarrow \mathbb{E}[D] \in (0, \infty)$.

Let $d_{\min} := \min\{d : \mathbb{P}(D = d) > 0\}$, $\bar{F} = 1 - F$ and $N_n$ be the number of far-out vertices in $RCM[n]$. In what follows, we assume that $d_{\min} \geq 3$, $d_i^{(n)} \geq d_{\min}$ for all $i$ and $n$ and that $\bar{F}$ is continuous. We have then, Theorem 2. Suppose the sequence $(x_n)_{n \in \mathbb{N}}$ is such that when $n \rightarrow \infty$,

$$n\bar{F}(x_n)^{d_{\min}} \rightarrow x$$

for some fixed $x > 0$. Then $N_n$ converges in the variation distance topology to the Poisson random variable with parameter $xp_{d_{\min}}$ for almost all sequences $\{RCM[n]\}_{n \geq 1}$. Appear in the proceedings of the MAMA workshop, held in conjunction with ACM Sigmetrics/Performance 2012.
COROLLARY 3. Let $M_n := \max_{1 \leq i \leq n} M_n(i)$. Then, for $x > 0$

$$\mathbb{P}(n \hat{F}(M_n)^d_{\min} > x) \to e^{-xp_{d_{\min}}}$$ (2)

Heuristically, given a particular realization of $RCM[n]$, $I^{(n)}_i := 1$ (vertex $i$ is far-out) is a Bernoulli random variable with parameter $\hat{F}(x_n)^d_i$ and $N_n = \sum_{i=1}^n I^{(n)}_i$. Ignoring the weak dependence among $I^{(n)}_i$’s,

$$N_n \overset{d}{\approx} \text{Binomial}(n, \sum_{i=1}^n \hat{F}(x_n)^d_i)$$

Now,

$$\sum_{i=1}^n \hat{F}(x_n)^d_i \approx n\mathbb{E}[\hat{F}(x_n)^d_i] = n\hat{F}(x_n)^d_{\min}\mathbb{E}[\hat{F}(x_n)^d_\min]$$ (3)

where $\hat{D} = D - d_{\min}$, from (1), when $n \to \infty$, $\hat{F}(x_n) \to 0$ and therefore

$$\mathbb{E}[\hat{F}(x_n)^d] \to \mathbb{P}(\hat{D} = 0) = \mathbb{P}(D = d_{\min}) = p_{d_{\min}}$$

This, along with (1) and (3), implies that for large $n$,

$$\sum_{i=1}^n \hat{F}(x_n)^d_i \approx xp_{d_{\min}}$$

and hence,

$$N_n \overset{d}{\approx} \text{Poisson}(xp_{d_{\min}})$$

We make the above steps rigorous using the Stein-Chen method for which we need the following theorem and lemmata. Here, $P_\lambda$ denotes Poisson r.v. with parameter $\lambda$, $\mathcal{L}(\mathbf{V})$ denotes the law of random vector $\mathbf{V}$, and $d_{\text{var}}(\cdot, \cdot)$ denotes the distance in total variation.

THEOREM 4 (STEIN-CHEN). For a finite or countable set $V$, let $(I_j)_{j \in V}$ be a family of not necessarily independent Bernoulli variables with respective parameters $(\pi_j)_{j \in V}$. Let $X = \sum_{j \in V} I_j$ and $\lambda = \sum_{j \in V} \pi_j$. Assume that there exist random variables $(J_{ij})_{i,j \in V,j \neq i}$ defined on the same probability space as $(I_j)_{j \in V}$ and such that for all $i \in V$, the following equality of distributions holds:

$$\mathcal{L}((J_{ij})_{i,j \in V,j \neq i}) = \mathcal{L}((J_{ij})_{j \in V,j \neq i} | I_i = 1)$$ (4)

Then

$$d_{\text{var}}(\mathcal{L}(X), P_\lambda) \leq 2 \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in V} \pi_i (\pi_i + \sum_{j \in V \setminus \{i\}} \mathbb{E}(|I_j - J_{ij}|))$$ (5)

PROOF. see chapter 2, [2]

LEMMA 5. For $\lambda, \mu \geq 0$,

$$d_{\text{var}}(P_\lambda, P_\mu) \leq 2|\lambda - \mu|$$

PROOF. see chapter 2, [2]

LEMMA 6. Given the Condition 1 and the condition on $(x_n)_{n \in \mathbb{N}}$ in Theorem 2, we have that when $n \to \infty$

$$\sum_i \hat{F}(x_n)^d_{\min} \to 0$$ (6)

$$\sum_i d_i^{(n)} \hat{F}(x_n)^d_{\min} + 1 \to 0$$ (7)

PROOF. We have from (1), when $n \to \infty$,

$$\hat{F}(x_n) \to 0$$ (8)

Then,

$$\sum_i \hat{F}(x_n)^d_{\min} \leq n\hat{F}(x_n)^d_{\min} \leq \hat{F}(x_n)^d_{\min}$$

which converges to 0 by (1) and (8).

Similarly,

$$\sum_i d_i^{(n)} \hat{F}(x_n)^d_{\min} + 1 \leq \hat{F}(x_n)n\hat{F}(x_n)^d_{\min} \sum_i d_i^{(n)}$$

which converges to 0 by (1), (8) and Condition 1.

PROOF OF THEOREM 2.

Let $I^{(n)}_i := 1$ (vertex $i$ is far-out in $\hat{RCM}[n]$) be a Bernoulli random variable with parameter $\pi_i := \hat{F}(x_n)^d_{\max}$. Therefore, $N_n = \sum_{i=1}^n I^{(n)}_i$ and let $S = \sum_{i=1}^n \hat{F}(x_n)^d_{\max}$. We will drop the superscript $(n)$, where it is clear from context. Further, let

$$J_{ij} := 1_{k,j,k \neq i} \min (Y(e_{ij}), x_n)$$

From the independence of edge weights, it is clear that (4) is satisfied, given $RCM[n]$. Now,

$$|I_j - J_{ij}| = I_j - J_{ij} = 1_{\exists e_{ij} \in RCM[n], Y(e_{ij}) \leq x_n}$$

and $\min_{k,j,k \neq i} (Y(e_{ij}) > x_n)$

Therefore,

$$\mathbb{E}[|I_j - J_{ij}| | RCM[n]] = 1_{\exists e_{ij} \in RCM[n]}F(x_n)\hat{F}(x_n)^d_{\min}$$

Plugging this into (5), we get

$$d_{\text{var}}(\mathcal{L}(N_n | RCM[n]), P_\lambda) \leq 2 \frac{1 - e^{-\lambda}}{\lambda} \sum_i \pi_i^2$$

$$+ 2 \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \neq j} \pi_i \pi_j$$

$$\leq 2 \min(1, \frac{1}{S}) \sum_i \hat{F}(x_n)^d_{\int}$$

$$+ 2 \min(1, \frac{1}{S}) \sum_i \hat{F}(x_n)^d_{\int} \hat{F}(x_n)^d_{\int}$$

$$\leq 2 \min(1, \frac{1}{S}) \sum_i \hat{F}(x_n)^d_{\int} + \sum_i \hat{F}(x_n)^d_{\int}$$

Thus, by Lemma 6, we have when $n \to \infty$

$$d_{\text{var}}(\mathcal{L}(N_n | RCM[n]), P_\lambda) \to 0$$ (9)
Now,
\[ S = \sum_{i=1}^{n} \tilde{F}(x_n)^{d_i} \sum_{k=d_{\min}}^{\infty} \tilde{F}(x_n)^k u_k \]
\[ \leq n\tilde{F}(x_n)^{d_{\min}} \frac{u_{d_{\min}}}{n} + \tilde{F}(x_n)n\tilde{F}(x_n)^{d_{\min}} \sum_{k=0}^{\infty} \frac{u_k}{n} \]

By Condition 1,
\[ \frac{u_{d_{\min}}}{n} \to p_{d_{\min}} \quad \text{and} \quad \sum_{k=0}^{\infty} \frac{u_k}{n} = \sum_{k=0}^{\infty} \frac{d_k(n)}{n} \to \mathbb{E}[D] \]
which, along with (1) and (8), shows that \( S \to x p_{d_{\min}} \).

Thus, by Lemma 5,
\[ d_{\var}(P_{x p_{d_{\min}}}, P_{S}) \to 0 \quad (10) \]
Moreover by triangle inequality,
\[ d_{\var}(\mathcal{L}(N_{x}[RCM[n]]), P_{x p_{d_{\min}}}) \leq d_{\var}(\mathcal{L}(N_{x}[RCM[n]]), P_{S}) + d_{\var}(P_{x p_{d_{\min}}}, P_{S}) \]
Therefore, from (9) and (10), we have finally that,
\[ d_{\var}(\mathcal{L}(N_{x}[RCM[n]]), P_{x p_{d_{\min}}}) \to 0 \]
This completes the proof.

**Proof of Corollary 3.** It follows from the above theorem simply by taking
\[ x_n = \tilde{F}^{-1} \left( \frac{1}{n} \frac{1}{p_{d_{\min}}} \right) \]
and remarking that
\[ \mathbb{P}(M_n < x_n) = \mathbb{P}(N_n = 0) \to e^{-xp_{d_{\min}}} \]

**Remark 7.** For the case discussed in [3] where edge-weights are i.i.d exponential random variables with parameter 1, i.e., \( \tilde{F}(x) = e^{-x} \), we have by Corollary 3 that for large \( n \),
\[ \mathbb{P}(n e^{-M_{x,d_{\min}}} > O(\log n)) \approx e^{-O(\log n)p_{d_{\min}}} \]
or,
\[ \mathbb{P}(M_n < \frac{1}{d_{\min}} \log n - o(\log n)) \approx o(1) \]
or,
\[ \mathbb{P}(M_n \geq \frac{1}{d_{\min}} \log n - o(1)) \approx 1 - o(1) \]
which is consistent with the results in [3].

**Remark 8.** Elaborating on the conjecture regarding the longest edge of the MST on \( \tilde{RCM}[n] \), let \( \tilde{RCM}[n, x_n] \) denote the graph obtained by keeping only those edges of \( \tilde{RCM}[n] \) whose length is less than \( x_n \). The proof of Theorem 2 suggests that for any \( k \geq 2 \), when \( n \to \infty \) the connected components of size exactly \( k \) disappear in \( \tilde{RCM}[n, x_n] \). We confirm this below for \( k = 2 \). But there is nothing to suggest that there is one and only one unboundedly growing connected component, otherwise we could have concluded that far-out vertices are isolated in \( \tilde{RCM}[n, x_n] \) and thus the longest edge of the MST on \( \tilde{RCM}[n] \) has the most far-out vertex as one of its ends, and its length is same as the nearest-neighbour distance of that vertex.

**Theorem 9.** There are no clusters of size 2 in \( \tilde{RCM}[n, x_n] \) w.h.p.

**Proof.** From the construction of \( RCM[n] \), ([5]):
\[ \mathbb{P}(i, j \text{ are connected in } RCM[n]) = \frac{1}{cn^2} \]
where \( c \) is some constant. Then,
\[ \mathbb{P}(\exists \text{ a connected component of size } 2) \]
\[ \leq n^2 \mathbb{P}(i, j \text{ are connected and they form an isolated component}) \]
\[ \leq n^2 \frac{1}{cn} \tilde{F}(x_n) \tilde{F}(x_n)^{d_{\min} - 1} \]
\[ \leq n \tilde{F}(x_n)^{d_{\min}} \tilde{F}(x_n)^2 \]
which converges to 0 by (1) and (8).

### 3. Impact of Clustering

Taking cue from [1] where (directionally) convex order as a measure of clustering provides a useful tool to compare point processes, we introduce a similar notion for the weighted \( RCM \)'s. Given two random variables \( X \) and \( Y \), we say that \( X \) is convexly smaller than \( Y \), \( X \leq_{cv} Y \), if \( \mathbb{E}(f(X)) \leq \mathbb{E}(f(Y)) \) for all convex functions \( f \). Remark that the comparison necessitates \( \mathbb{E}(X) = \mathbb{E}(Y) \). We will order \( \tilde{RCM}[n] \) according to the order of their limiting degree distributions. This order imposes further constraints on \( RCM \)'s.

**Theorem 10.** Given two \( \tilde{RCM}[n] \), \( \tilde{RCM}[d^1] \) and \( \tilde{RCM}[d^2] \), such that \( D^1 \leq_{cv} D^2 \), we should have one of the following two conditions:
1. \( d_{\min}^1 = d_{\min}^2 = d_{\min} \) and \( \mathbb{P}(D^1 = d_{\min}) \leq \mathbb{P}(D^2 = d_{\min}); \)
2. \( d_{\min}^1 > d_{\min}^2 \).

**Proof.** Suppose \( d_{\min}^1 < d_{\min}^2 \). Then it is possible to construct a convex function \( f \) such that \( f(d_{\min}^1) = 1 \) and \( f(x) = 0 \) for all \( x \geq d_{\min}^1 + 1 \). Then
\[ \mathbb{P}(D^1 = d_{\min}^1) = \mathbb{E}(f(D^1)) > \mathbb{E}(f(D^2)) = 0 \]
which is a contradiction. Now for the case when \( d_{\min}^1 = d_{\min}^2 = d_{\min} \), we have
\[ \mathbb{E}(z^{D^1 - d_{\min}}) \leq \mathbb{E}(z^{D^2 - d_{\min}}) \]
for all \( z \in (0, 1) \). Letting \( z \) approach 0, we have \( \mathbb{E}(z^{D^1 - d_{\min}}) \to \mathbb{P}(D^1 - d_{\min} = 0) \) and therefore,
\[ \mathbb{P}(D^1 = d_{\min}) \leq \mathbb{P}(D^2 = d_{\min}) \]
The above result along with Theorem 2 suggests a connection between convex order on \( \tilde{RCM}[n] \) and the way they "cluster".

### 4. Information Transmission Under Channel-Dependent Emission Capital Constraints

Consider \( n \) servers labelled 1 to \( n \) such that at most \( d_i \) channels of communication can pass through server \( i \), where \( (d_i)^2 \) satisfy the conditions introduced earlier. A server is called closed if the maximum possible channels of communication have been established through it, and is called available otherwise.
Initially, a message is passed to a server $i$ from outside. The server establishes channels with $d_i$ available servers and then tries to transmit the message through each of them by spending an emission capital $C_n$ on each. But the transmission through channel $e_{ij}$ is successful only if $C_n$ exceeds $Y(e_{ij})$, where $\{Y(e_{ij})\}_{i,j}$ are i.i.d. random variables with distribution $F$. From then on, whenever the message reaches server $k$ for the first time, it establishes channels with $d_k - 1$ available servers (excluding the server from where it received the message) and tries to transmit the message through these channels in the same manner as described earlier.

Our results show that for large $n$, $C_n$ should be at least $x_n$ as given by (11) if the message is to reach the entire network w.h.p.

5. REFERENCES

[1] B. Blaszczyszyn and D. Yogeshwaran. Clustering, percolation and directionally convex ordering of point processes. arXiv:1105.4293v1, 2011.

[2] M. Draief and L. Massoulie. *Epidemics and Rumors in Complex Networks*, volume 369 of London Mathematical Society Lecture Notes. Cambridge University Press, 2010.

[3] M. Lelarge and H. Amini. The diameter of weighted random graphs. arXiv:1112.6330v1, 2011.

[4] M. D. Penrose. The longest edge of the random minimal spanning tree. *The Annals of Applied Probability*, 7(2):340–361, 1997.

[5] R. van der Hofstad. *Random graphs and complex networks*. 2009. available at \( http://www.win.tue.nl/~rhofstad/RGCN.html \).