Connectivity Properties and Structure of Random Graphs Obtained by Vertex Percolation on Erdős-Rényi Binomial Random Graph

Dibyendu Saha
dibyendu.0399@gmail.com

Bhaskar Ray
bhaskarray.9916@gmail.com

***Guide*: Antar Bandyopadhyay**
antar@isid.ac.in

June 2020

Abstract

Erdős-Rényi Binomial Random Graph is a Random graph with a fixed number of vertices and each edge of that graph being present with some probability, independent of each other. Vertex Percolation is a process where each vertex in a graph is present with some probability, independent of the other vertices. Needless to say, if a vertex is absent in a graph, all edges emerging from it are also absent. In this project we have carried out simulations to study the behaviour of an Erdős-Rényi Binomial Random Graph after performing Vertex Percolation, in terms of connectivity.

Keywords

Random Graph; Erdős-Rényi; Erdős-Rényi Binomial Random Graph; Vertex Percolation; Connectivity "Threshold".

1 Introduction

Consider the following problem. There is a large state with several cities. There is a road between every pair of cities. Suddenly, the state administration decides to block a few randomly chosen roads for maintenance. Also, some of the cities randomly decide not to allow travel through them, independently of the other cities.

This clearly means that each those cities has blocked all roads that have one end in that city. Now, a traveller standing in one of the remaining cities wants to know the chances that he can travel to all of the remaining cities without having to break any rule.

We are now in a position to frame our objective mathematically. Let G be a random graph obtained by erasing vertices of the complete graph K_n independently with probability $1 - q_n$; where $0 < q_n < 1$ and edges with probability $1 - p_n$; where $0 < p_n < 1$. The main goal of the project is to study the connectivity properties and structure of such random graphs for large n. Note that, p_n and q_n may depend on n and need not be equal. Deletion of edges from K_n gives us an Erdős-Rényi Binomial random graph, while deleting vertices is vertex percolation.

2 Formal Definitions

Let K_n be the complete graph with n vertices. We define

$$\begin{aligned} &\mathcal{V}_n := V(K_n) && \mathcal{E}_n := E(K_n) \\ \implies &\mathcal{V}_n = [n] && \mathcal{E}_n = \{\{u,v\} | u \neq v \; ; \; u, v \in [n]\} \end{aligned}$$

Let

$$\begin{aligned} &(X_v)_{v \in \mathcal{V}_n} \sim \text{i.i.d. Bernoulli}\,(q_n) \\ &(Y_e)_{e \in \mathcal{E}_n} \sim \text{i.i.d. Bernoulli}\,(p_n) \\ &X_v \;\&\; Y_e \text{ are independent } \forall\, v \in \mathcal{V}_n \;\&\; e \in \mathcal{E}_n \end{aligned}$$

Let G be random graph such that

$$\begin{aligned} &V(G) := \{v | v \in \mathcal{V}_n \;\&\; X_v = 1\} \\ &E(G) := \{e := \{u,v\} | e \in \mathcal{E}_n \;\&\; X_v = 1, X_u = 1, Y_e = 1\} \end{aligned}$$

Such a G is our graph of interest. We will denote such a graph as

$$G \sim \mathcal{G}(n, p_n, q_n)$$

3 Methodology

For a graph $G \sim \mathcal{G}(n, p_n, q_n)$, we have the following results. The proofs are provided in Appendix-1.

Result 3.1. $\mathbb{E}\left[|V(G)|\right] = nq_n$

Result 3.2. $\mathbb{P}\left(e \in E(G)\right) = q_n^2 p_n \ \forall e \in \mathcal{E}_n$

Result 3.3. $\mathbb{P}\left(e = \{u, v\} \in E(G) | X_u = 1, X_v = 1\right) = p_n$

Result 3.4. *If vertex percolation is done first, the graph obtained is an Erdős-Rényi Binomial Random Graph on the remaining set of vertices i.e.,* $G|(X_v)_{v\in\mathcal{V}_n} \sim \mathcal{G}\left(|V(G)|, p_n\right)$

3.1 Connectivity in Erdős-Rényi Binomial Random Graph

There are two important results about connectivity in Erdős-Rényi Binomial Random Graph depending on edge probability. One of them gives the idea of Threshold for connectivity property and the other is about probability of connectivity around connectivity threshold.

Connectivity Threshold

Theorem 3.1 (Connectivity Threshold)**.** *Let* $G \sim \mathcal{G}\left(n, p_n\right)$*. Then,*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}(n,p_n)}\left(G \textit{ is connected}\right) = \begin{cases} 0 & ; \dfrac{p_n}{\log(n)/n} \to 0 \\ 1 & ; \dfrac{p_n}{\log(n)/n} \to \infty \end{cases}$$

Critical Window

For an Erdős-Rényi Binomial Random Graph , if edge probability is 'around' threshold then we can state the following about its connectivity,

Theorem 3.2. *Fix* $t \in \mathbb{R}$ *and* $\lambda_n = \log n + t$*;*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}\left(n,\frac{\lambda_n}{n}\right)}(G \textit{ is connected}) = \exp\left\{-\exp\left\{-t\right\}\right\}$$

Corollary 3.2.1. *Fix* $p \in (0, 1)$*;*

$$\mathbb{P}_{\mathcal{G}(n,p)}(G \textit{ is connected}) = \exp\left\{-\exp\left\{-n\left(p - \frac{\log n}{n}\right)\right\}\right\}(1 + o(1))$$

3.2 Connectivity in the graph of our interest

From Result 3.4, we know that if vertex percolation is done ,i.e. all $(X_v)_{v\in\mathcal{V}_n}$ are given, our graph of interest is an Erdős-Rényi Binomial Random Graph with edge probability p_n and number of vertices is $|V(G)|$, which, from Result 3.1 is 'expected' to be nq_n. Also, connectivity is an increasing property for Erdős-Rényi Binomial Random Graphs while it has no such characteristics for Random Graphs obtained by Vertex Percolation. Hence, for now, we shall restrict ourselves to studying the connectivity properties with respect to p for given values of n and q_n.

3.2.1 Idea of a threshold

Keeping the theorem 3.1 and corollary 3.2.1 in mind, for a given n and q_n, if expected number of vertices goes to ∞ as n goes to ∞, for the plot of $\mathbb{P}_{\mathcal{G}(n,p,q_n)}(G\ \textit{is connected})$ vs p, we expect to observe the same behavior as in case of Erdős-Rényi Binomial Random Graph i.e., connectivity probability is 0 up to a certain value of p after which, it rapidly increases to 1.
If the graph were Erdős-Rényi Binomial Random Graph, for a given n, the increase takes place when p is around $\log n/n$, which is the **threshold**. After percolation our graph is also an Erdős-Rényi Binomial Random Graph with $|V(G)|$ vertices, which is (unconditionally) 'expected' to be nq_n. We thus, also want to verify if, for a given n and q_n, the graph of our interest follows a similar trend when p is around $\log(nq_n)/(nq_n)$.

Simulation 1 Our goal is to check whether after percolation, the connectivity of the graph behaves like it has a threshold. For that, we have plotted estimated probability of connectivity, $\hat{p}_{con.}(p)$ vs p for some fixed "large" value of n and different constant values of q_n.

Scheme

- Fix a 'large enough' n.
- Fix a $q_n \in (0,1)$
- Fix a set of values of $p \in (0,1)$. For each of these values, we shall estimate the *probability of connectivity.*
- For each value of p,

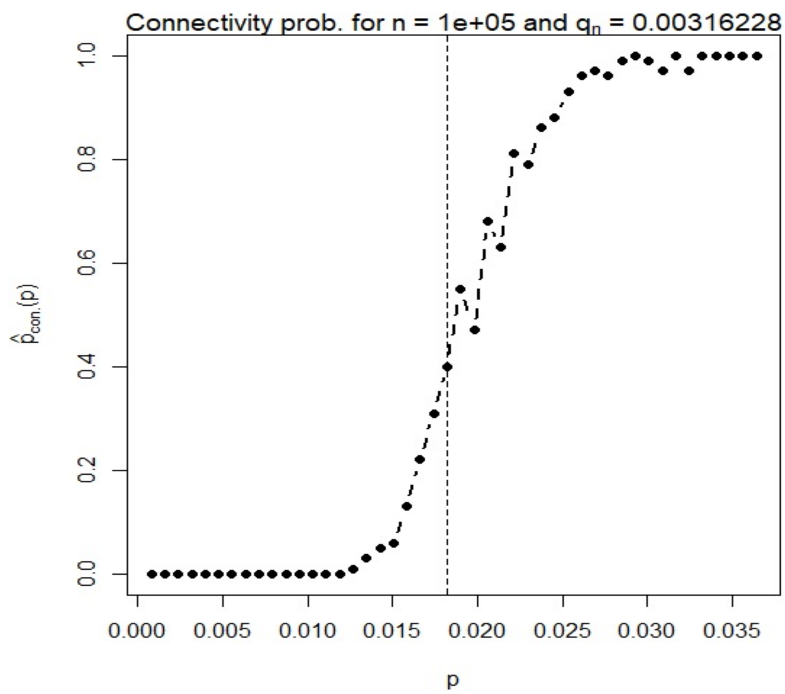
i (Vertex Percolation) Since we are performing vertex percolation first, removing the vertices for which $X_v = 0$ gives us a smaller, yet complete graph with $\sum_{v \in \mathcal{V}_n} X_v$ vertices. This is invariant under permutation of values of X_v. Hence, it is sufficient to simulate the value of $\sum_{v \in \mathcal{V}_n} X_v$, which follows Binomial$(n, q_n)$. So, we draw a random variable from Binomial(n, q_n) ,say m.

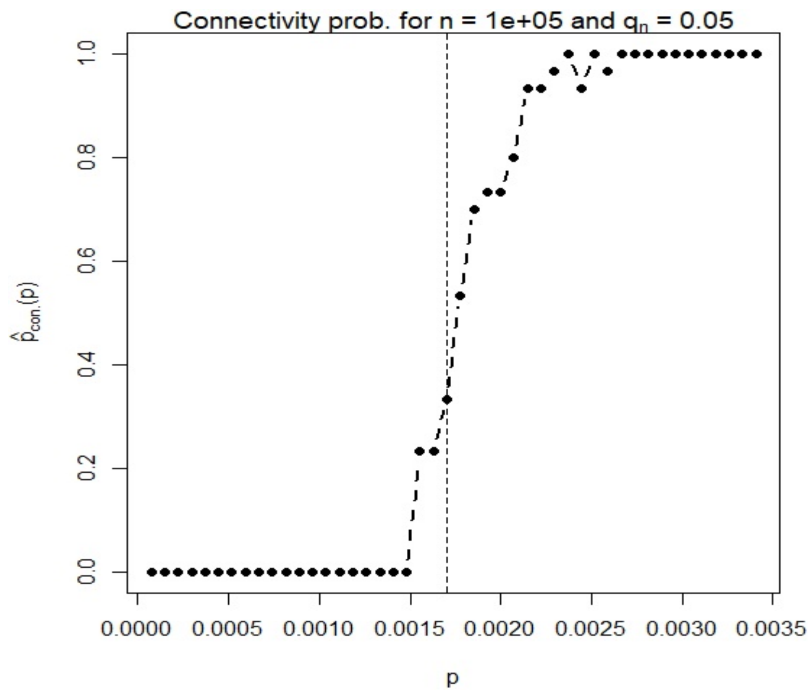
ii We simulate a graph G from $\mathcal{G}(m, p)$.

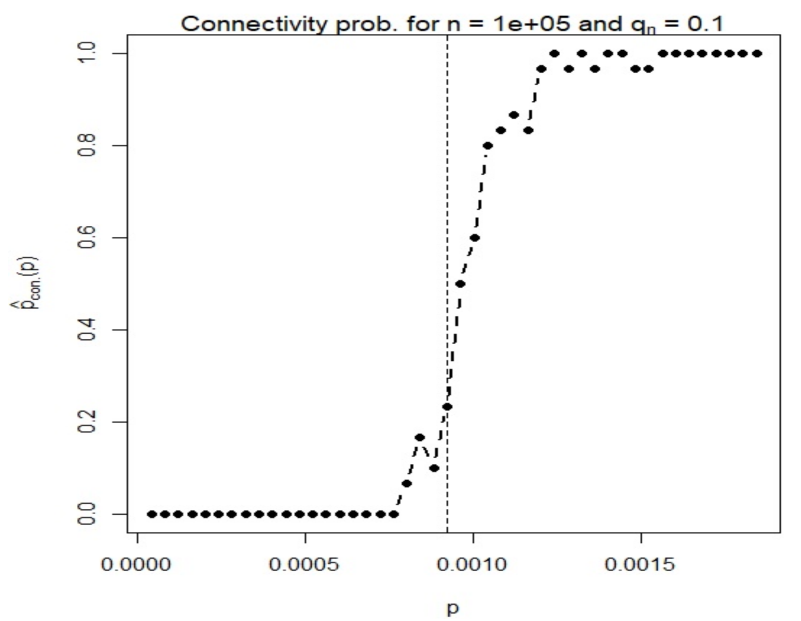
iii Check if G is connected.

iv Repeat steps (i) to (iii) 1000 times and find the proportion of graphs that are connected. This gives the estimated probability of connectivity, $\hat{p}_{con.}(p)$.

- Plot $\hat{p}_{con.}(p)$ against p.




Connectivity prob. for n = 1e+05 and q_n = 0.05
1.0
0.8
0.6
0.4
0.2
0.0
p̂_con.(p)
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0.0030
0.0035
p


Connectivity prob. for n = 1e+05 and q_n = 0.1
1.0
0.8
0.6
0.4
0.2
0.0
p̂_con.(p)
0.0000
0.0005
0.0010
0.0015
p

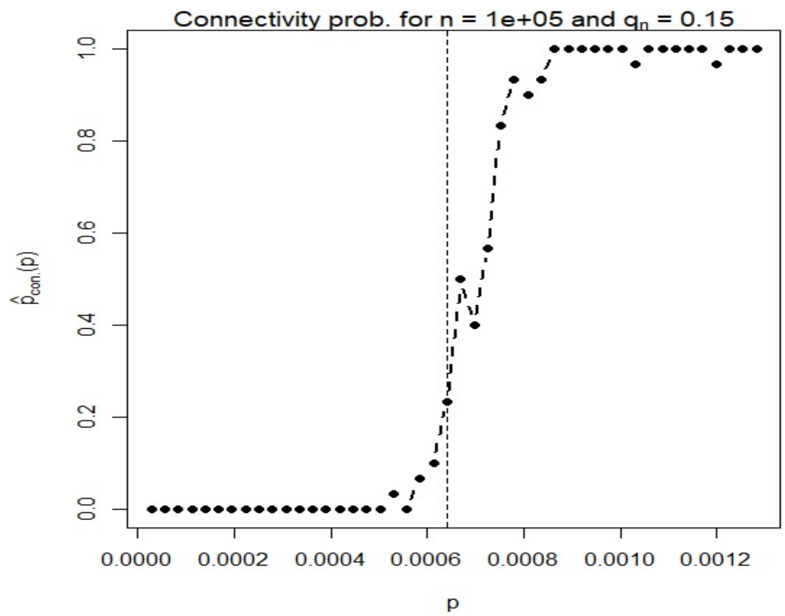


Figure 1: Plots of simulations of connectivity probability, $\hat{p}_{con.}(p)$ vs p for given n and q_n. The vertical line in each graph corresponds to $p = \log(nq_n)/(nq_n)$

Observations

- The connectivity probability apparently stays 0 up to a certain value of p and then increases steadily to 1, for some values of q_n.
- The transition of the connectivity probability from 0 to 1 appears to take place around $\log(nq_n)/(nq_n)$ as expected.

Simulation 2 In order to confirm our observations from Simulation 1, we need to simulate for a wide range of q_n for a fixed n. We want to put all the plots obtained for a different q_n for a fixed value of n in a single plot.

Scheme

- For each value of n, fix a set of values for q_n to simulate.
- For each q_n
  - Follow the scheme for Simulation 1 to obtain values of $\hat{p}_{con.}(p)$ vs p.

- Fit $\hat{p}_{con.}(p) = \exp\{-\exp\{-\delta_n (p - c_n)\}\}$ using IRLS to obtain estimates $\hat{c}_n$ and $\hat{\delta}_n$.
- Plot $\hat{p}_{con.}(p)$ vs $p^* := p - \hat{c}_n$ in the same plot.

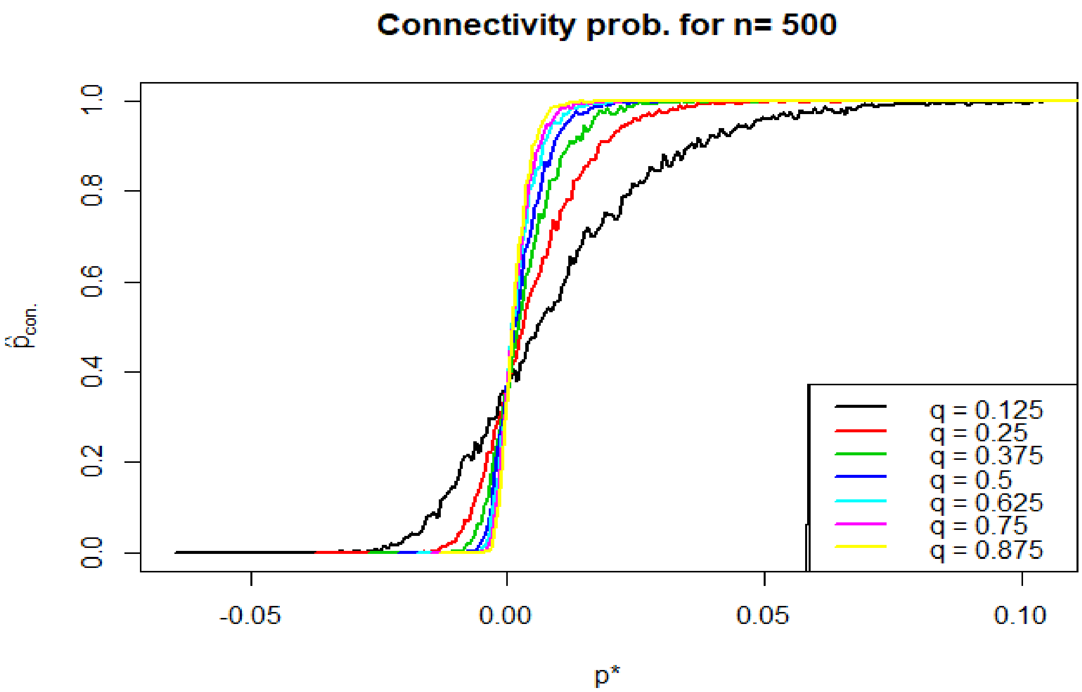


Figure 2: Plots corresponding to various q_n values for $n = 500$

q_n	Multiple Correlation
0.125	0.9997871
0.250	0.9998905
0.375	0.9998995
0.500	0.9999568
0.625	0.9999508
0.750	0.9999827
0.875	0.9999582

Table 1: Multiple Correlation of the fitted model is presented against q_n for $n = 500$

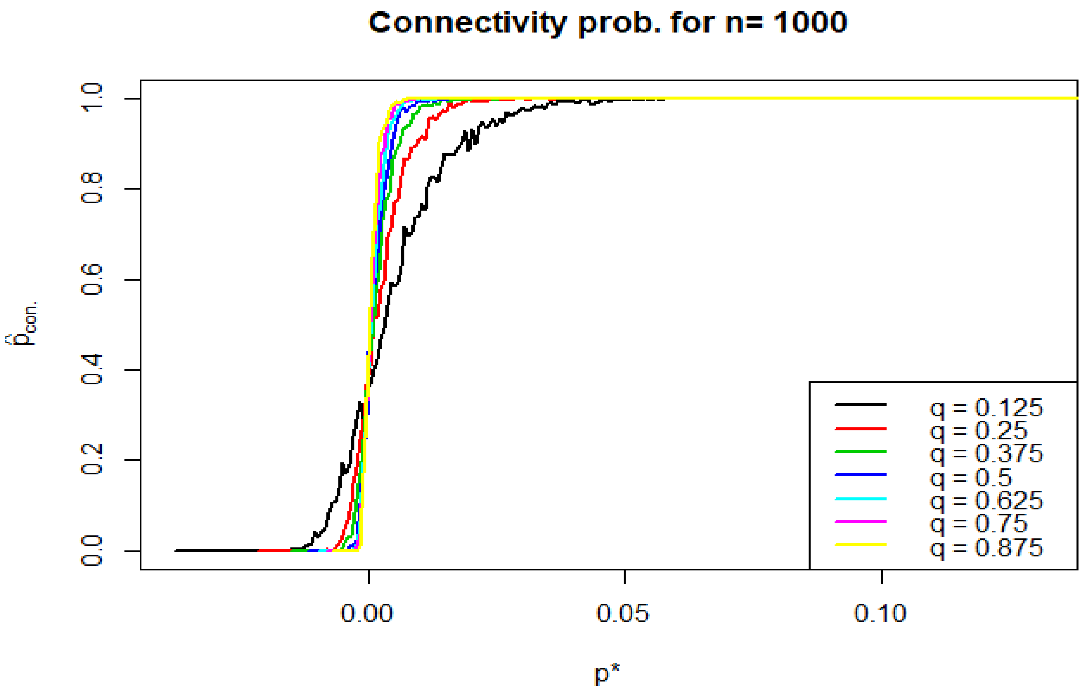


Figure 3: Plots corresponding to various q_n values for $n = 1000$

q_n	Multiple Correlation
0.125	0.9998685
0.250	0.9999298
0.375	0.9999158
0.500	0.9999595
0.625	0.9999591
0.750	0.9999464
0.875	0.9999696

Table 2: Multiple Correlation of the fitted model is presented against q_n for $n = 1000$

Observations

- The connectivity probability stays 0 up to a certain value of p and then increases steadily to 1, confirming observation from Simulation 1.

- The connectivity probability behaves like it has a threshold.
- For increasing value of q_n, the transition from 0 to 1 gets steeper.

3.2.2 Fitting a functional form to the connectivity probability

By now, we have observed that the connectivity probability for our graph of interest behaves much like that for an Erdős-Rényi Binomial Random Graph. We are now interested to see if it conforms to a functional form similar to that stated in Corollary 3.2.1. In view of the previous results and simulations, for $p \in (0, 1)$, we expect it to be

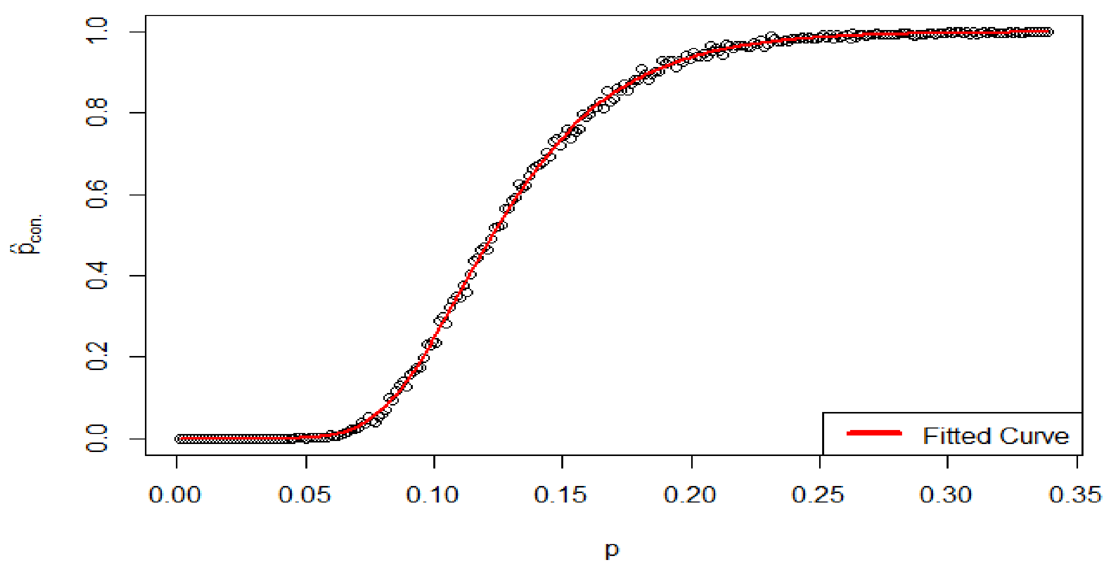
$$\mathbb{P}_{\mathcal{G}(n,p,q_n)}(G \text{ is connected}) = \exp\left\{-\exp\left\{-nq\left(p - \frac{\log nq}{nq}\right)\right\}\right\}(1 + o(1))$$

Simulation 1 We shall estimate the connectivity probability using $\hat{p}_{con.}(p)$ as earlier. We shall then fit it against a model of the form $\exp\{-\exp\{-\delta_n(p - c_n)\}\}$ to see whether it conforms to the functional form.

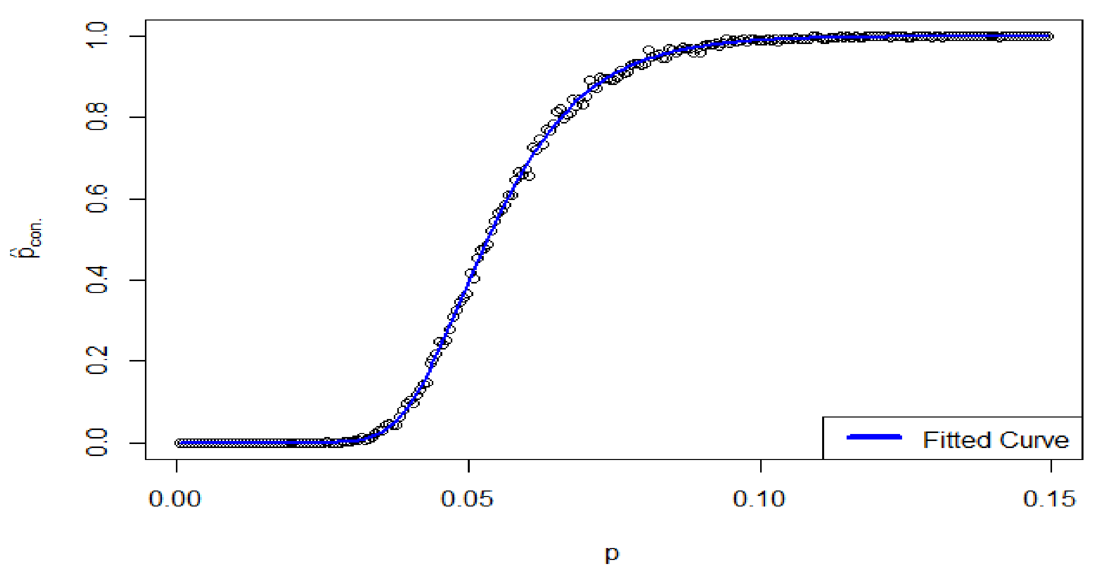
Scheme

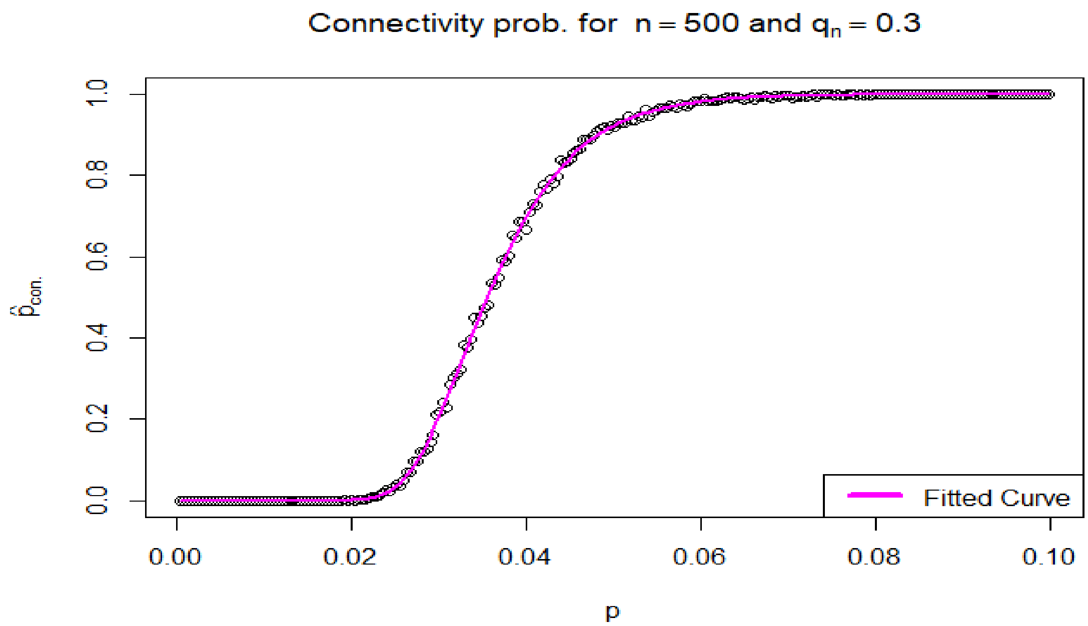
- Fix n and q_n.
- For various values of p, estimate the connectivity probability following the same scheme as in Section 3.2.1 and plot $\hat{p}_{con.}(p)$ against p.
- Fit $\hat{p}_{con.}(p) \approx \exp\{-\exp\{-\delta_n(p - c_n)\}\}$.
- Draw the fitted curve in the same plot.
- For a fixed functional form of q_n, for different values of n, we plot $\hat{p}_{con.}(p)$ against p^* in the same plot.

Connectivity prob. for $n = 100$ and $q_n = 0.3$

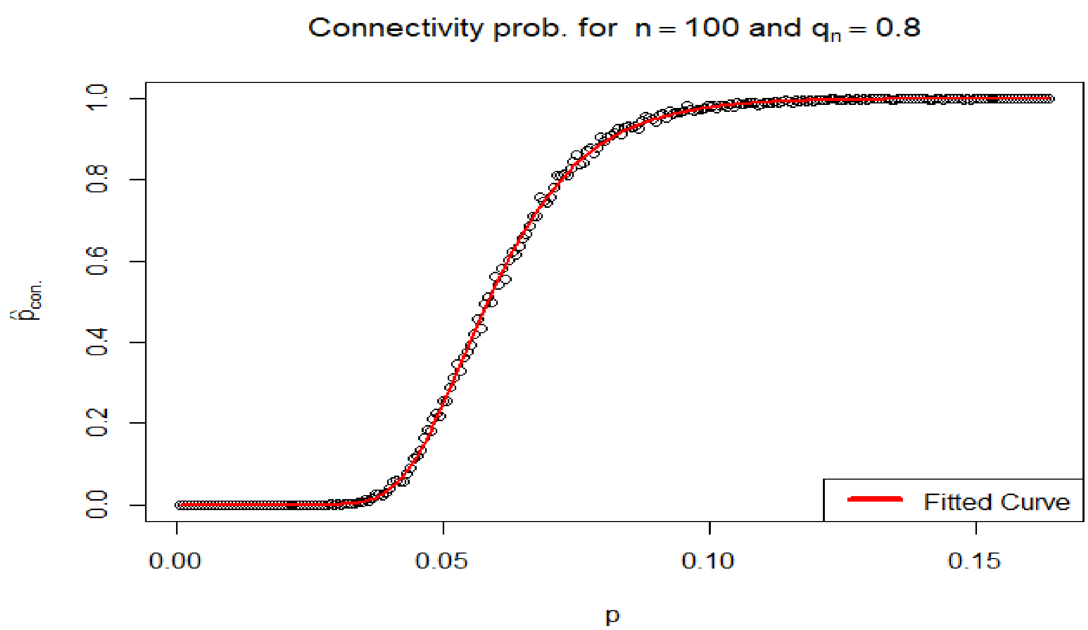


Connectivity prob. for $n = 300$ and $q_n = 0.3$

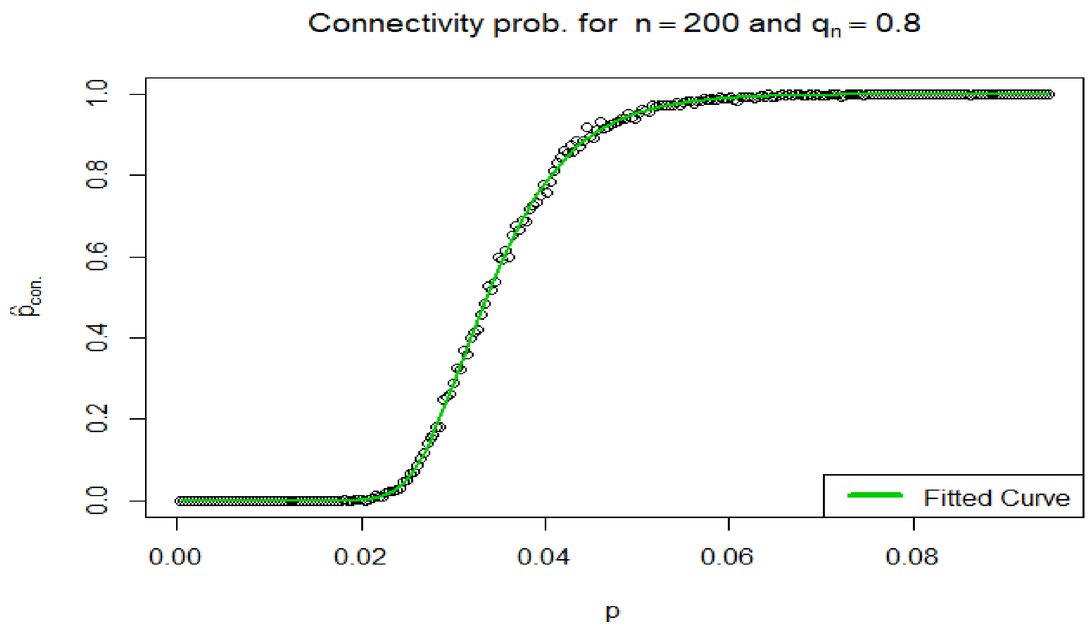


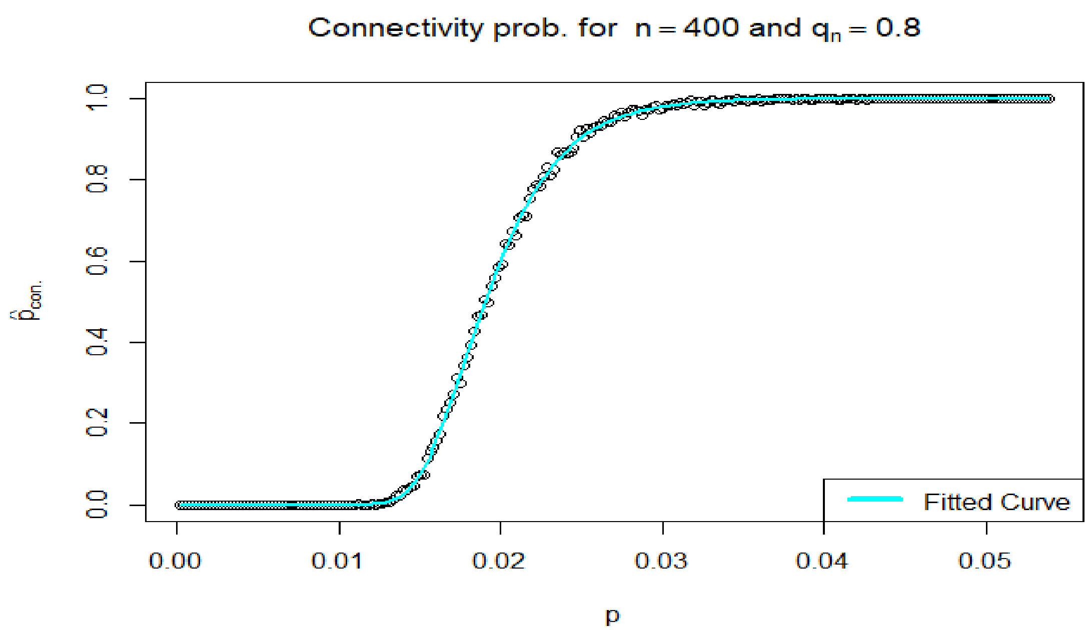


Connectivity prob. for n = 500 and q_n = 0.3
$\hat{p}_{con.}$
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.02
0.04
0.06
0.08
0.10
p
Fitted Curve



Connectivity prob. for n = 100 and q_n = 0.8
$\hat{p}_{con.}$
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
p
Fitted Curve


Connectivity prob. for $n = 200$ and $q_n = 0.8$
$\hat{p}_{con.}$
1.0
0.8
0.6
0.4
0.2
0.0
Fitted Curve
0.00
0.02
0.04
0.06
0.08
p


Connectivity prob. for $n = 400$ and $q_n = 0.8$
$\hat{p}_{con.}$
1.0
0.8
0.6
0.4
0.2
0.0
Fitted Curve
0.00
0.01
0.02
0.03
0.04
0.05
p

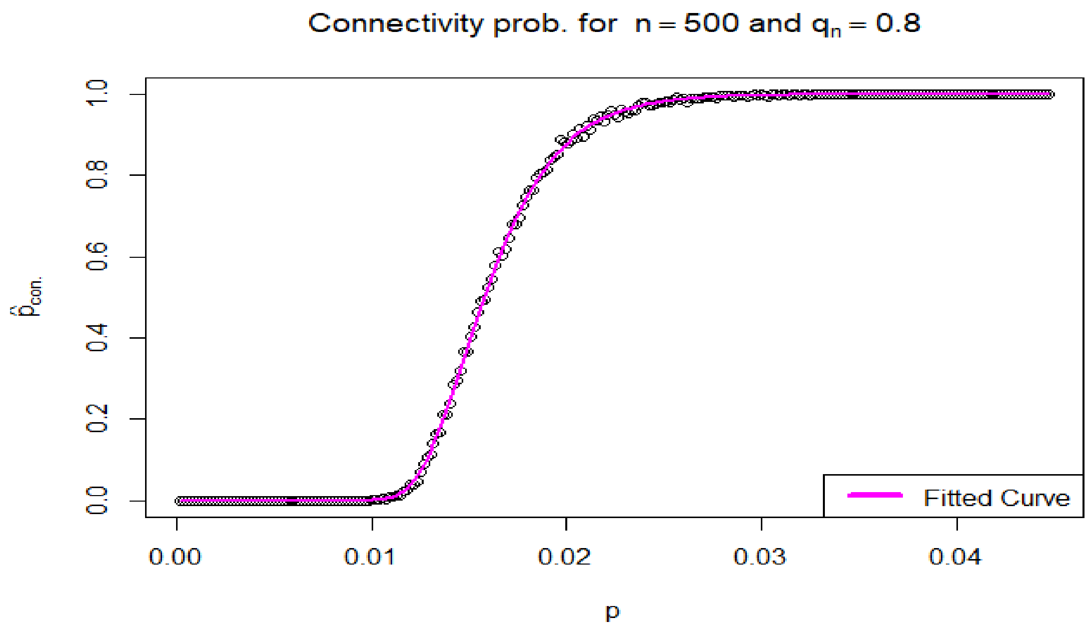
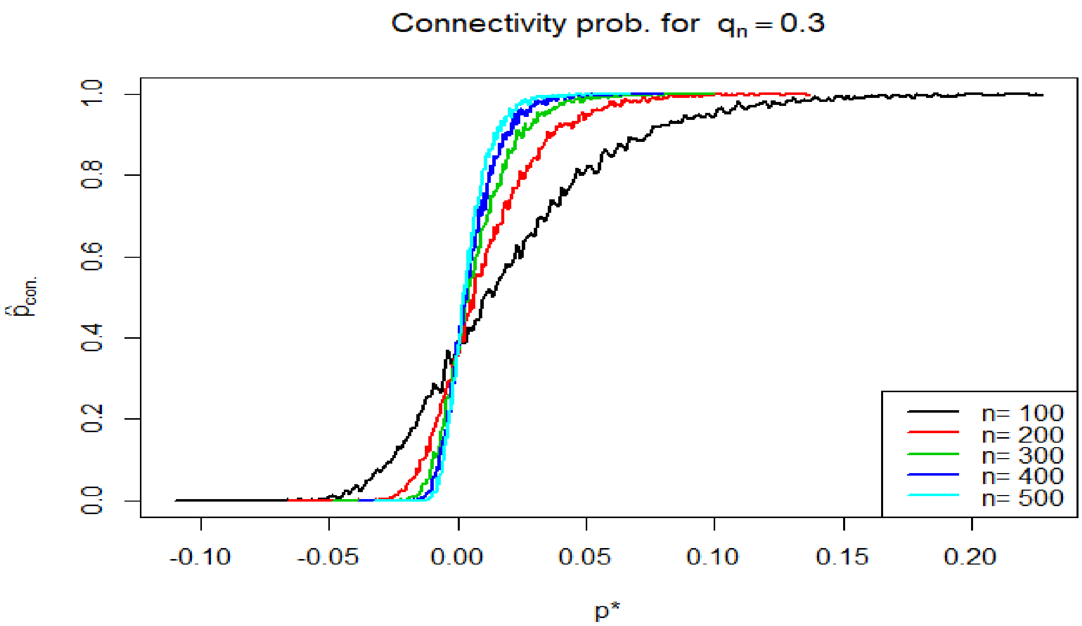
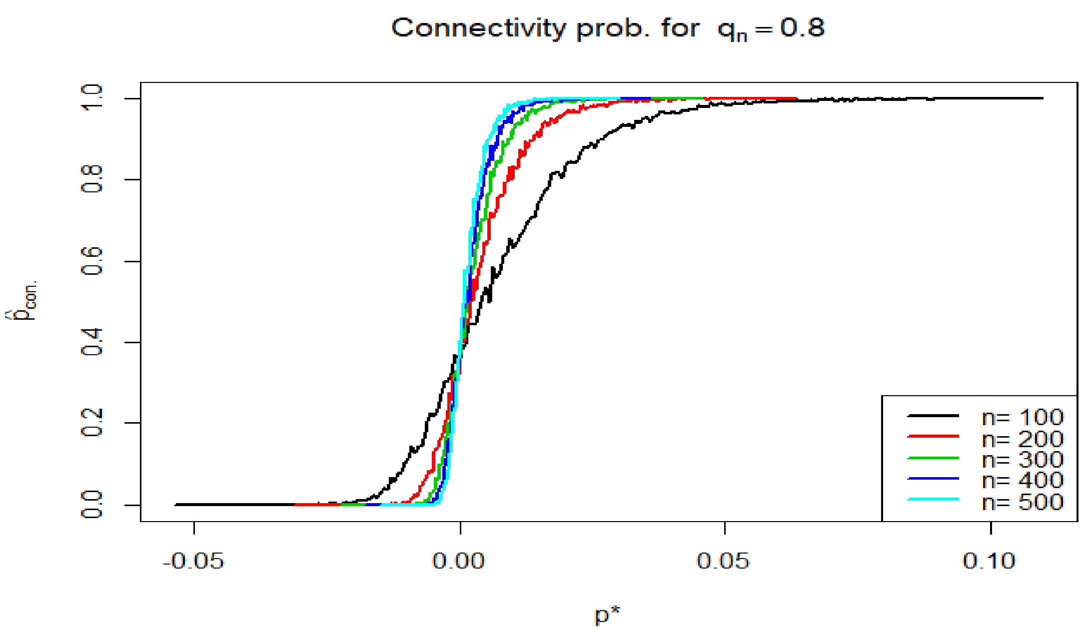


Figure 4: Plot of $\hat{p}_{con.}(p)$ against p for various n and q_n. The fitted curve according to the model are also plotted.


Connectivity prob. for $q_n = 0.3$
$\hat{p}_{con.}$
1.0
0.8
0.6
0.4
0.2
0.0
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
p*
n= 100
n= 200
n= 300
n= 400
n= 500


Connectivity prob. for $q_n = 0.8$
$\hat{p}_{con.}$
1.0
0.8
0.6
0.4
0.2
0.0
-0.05
0.00
0.05
0.10
p*
n= 100
n= 200
n= 300
n= 400
n= 500

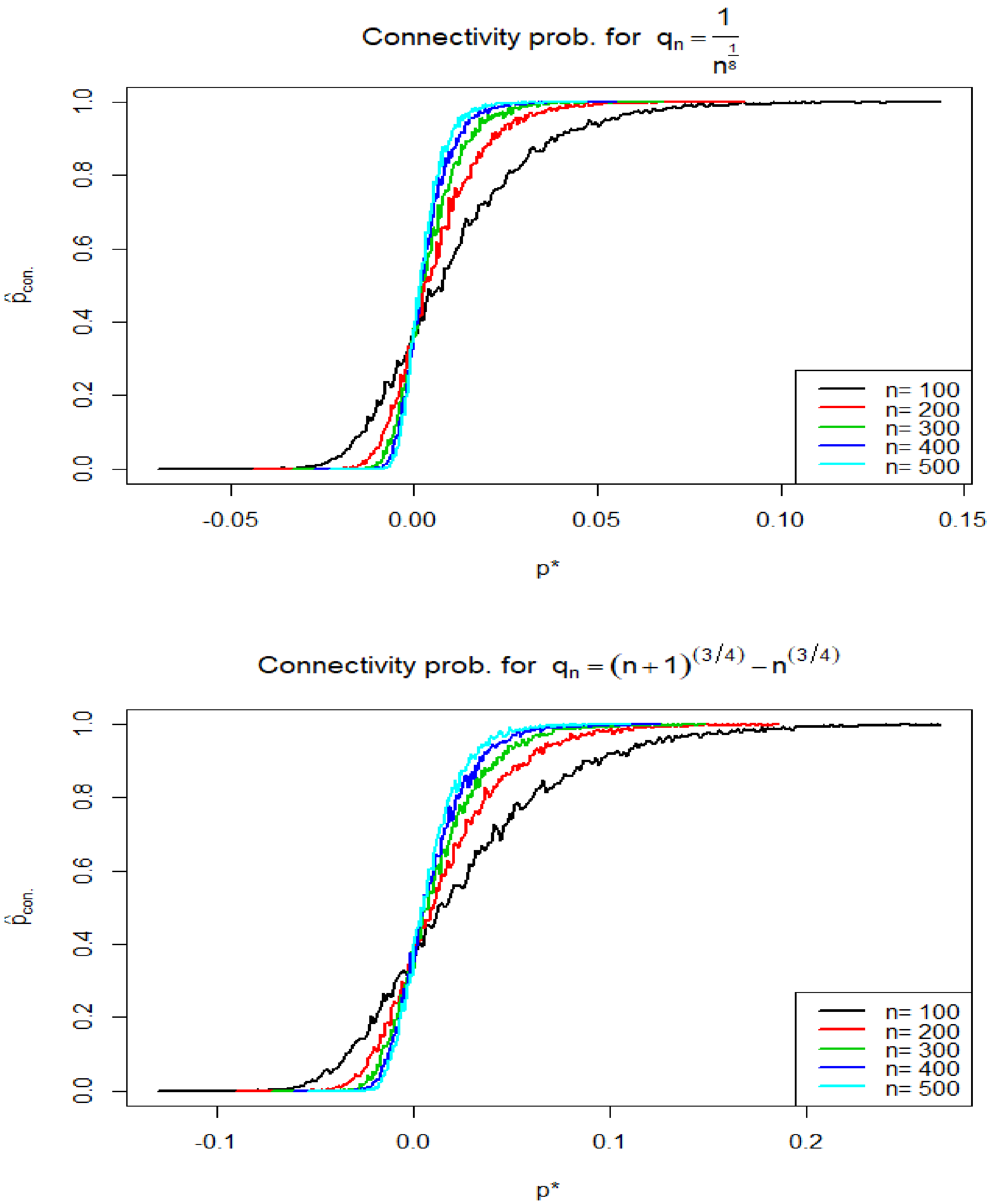


Figure 5: Plot of connectivity probability vs p^* for various values of n for a fixed functional form corresponding to q_n.

Observations

- It is evident from the plots that the fitted model provides a very good estimate of the original data.

- For increasing values of n, the transition from 0 to 1 gets steeper. Together with the previous observations, it seems that the transition gets steeper with increasing value of nq_n, which is the expected number of vertices of G.

Simulation 2 We now need to estimate the parameters, c_n and δ_n in the model. Further we shall check whether $\hat{c}_n \approx \log nq_n/nq_n$ and $\hat{\delta}_n \approx nq_n$.

Scheme

- Fix an $n - q_n$ relationship.

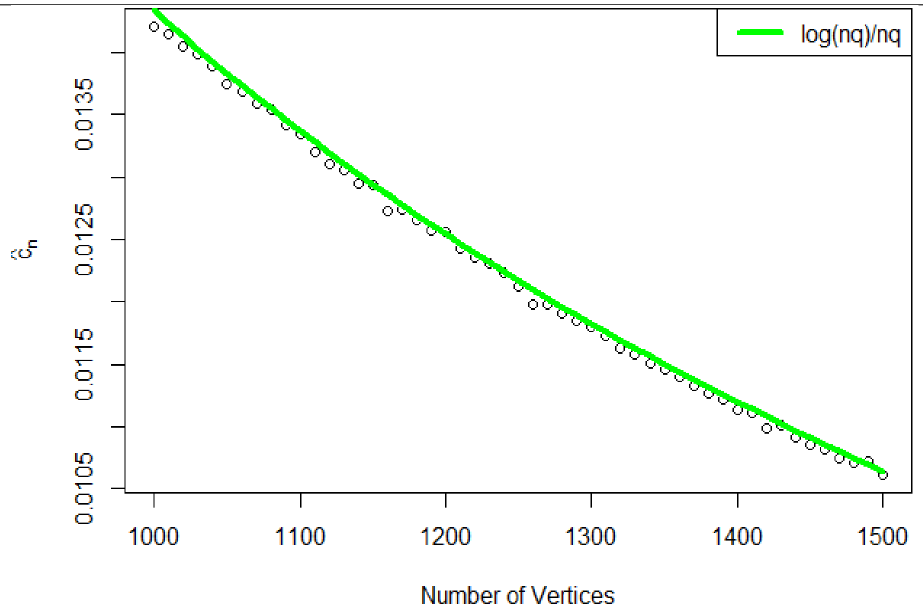
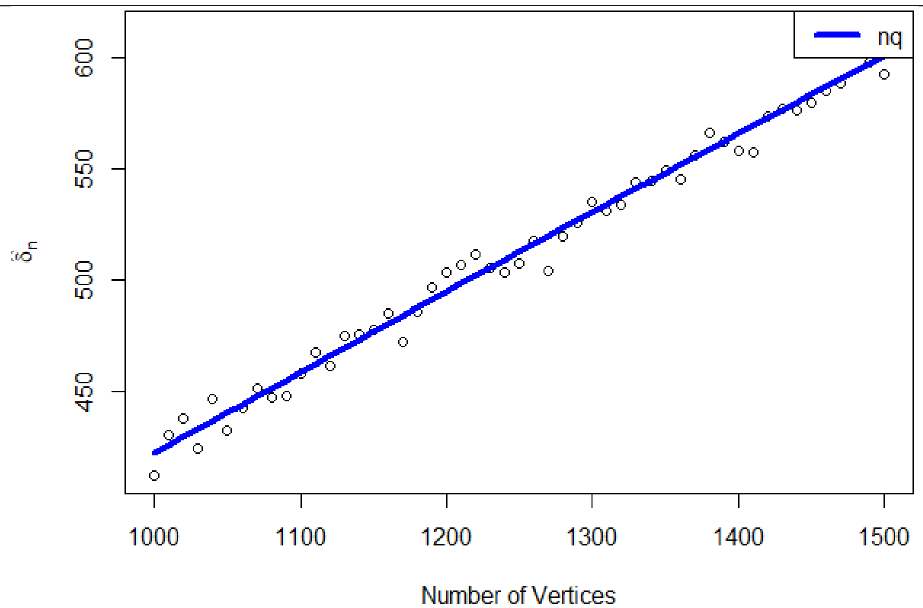
- For $n \in \{1000, 1001, \cdots, 1500\}$,
  - Simulate values for $\hat{p}_{con.}(p)$ vs p for different values of p according to the scheme in Section 3.2.1.
  - Fit $\hat{p}_{con.}(p) \approx \exp\{-\exp\{-\delta_n(p - c_n)\}\}$.
  - Find the estimates $\hat{c}_n$ and $\hat{\delta}_n$

- Plot $n\ vs\ \hat{c}_n$ and $n\ vs\ \hat{\delta}_n$.

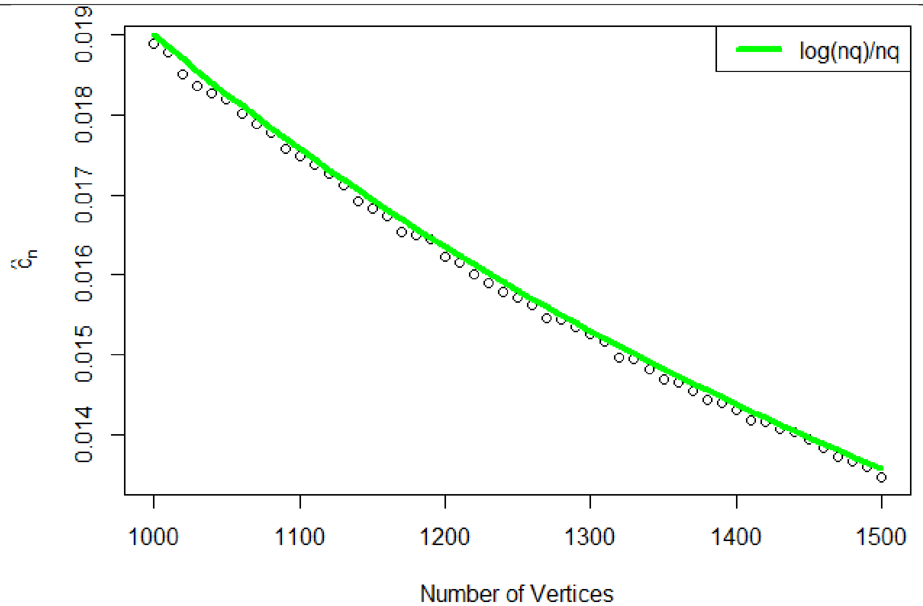
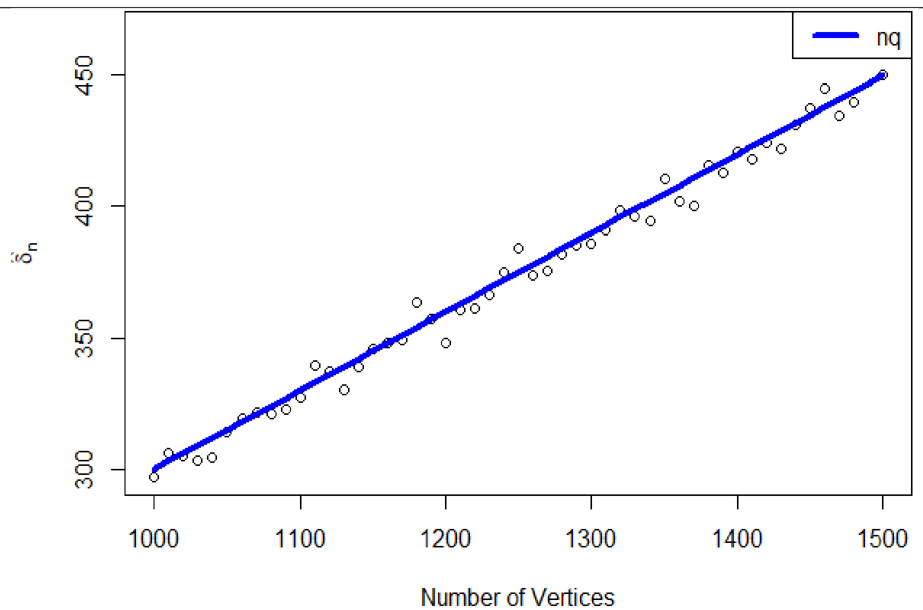
- Calculate $Cor(\{\hat{c}_n\}, \{\log nq_n/nq_n\})$ and $Cor(\{\hat{\delta}_n\}, \{nq_n\})$.

Plots

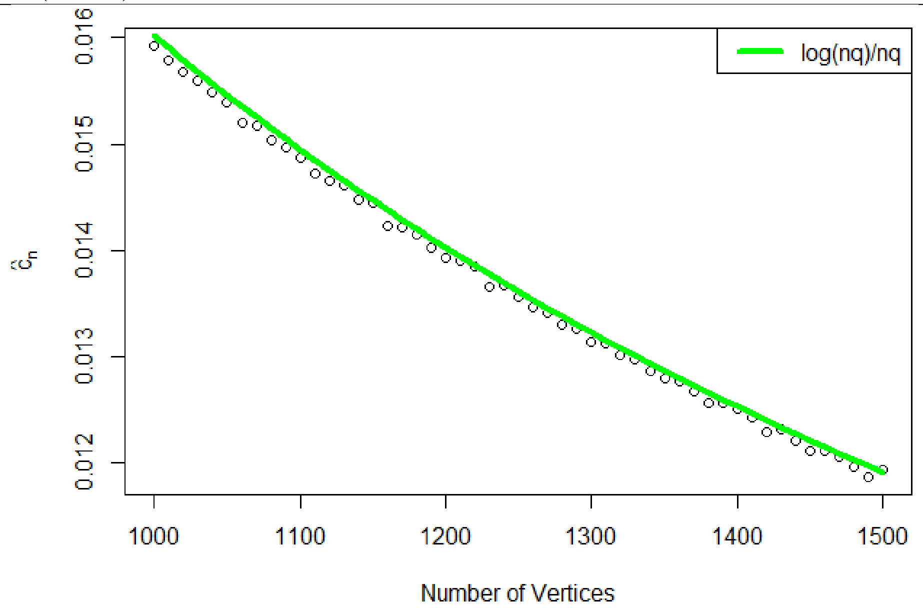
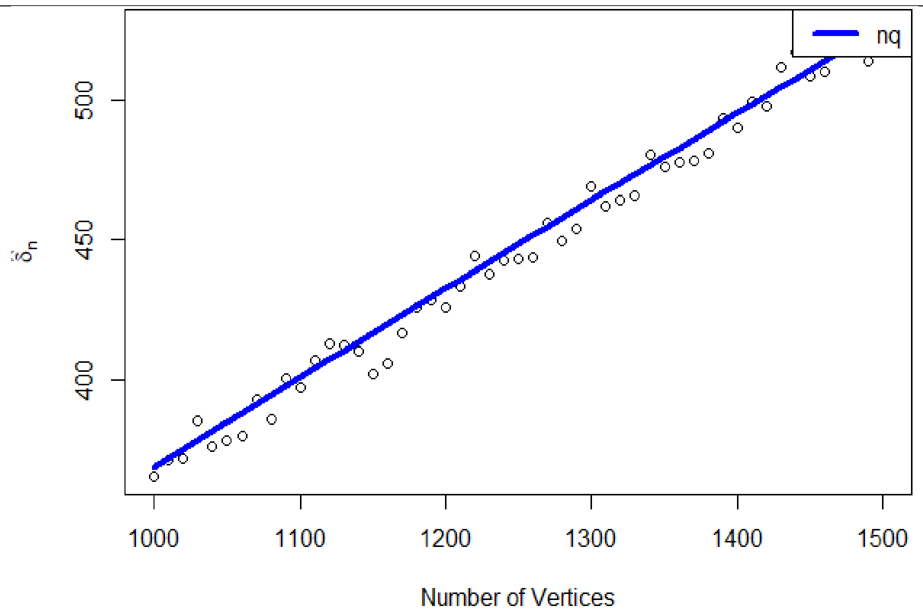
- $n = 1000 - 1500$ and $q_n = n^{-\frac{1}{8}}$

$n\ vs\ \hat{c}_n$ plot	
$Cor\,(\hat{c}_n, \log nq_n/nq_n)$	0.9998316
$n\ vs\ \hat{\delta}_n$ plot	
$Cor\left(\hat{\delta}_n, nq_n\right)$	0.9936192

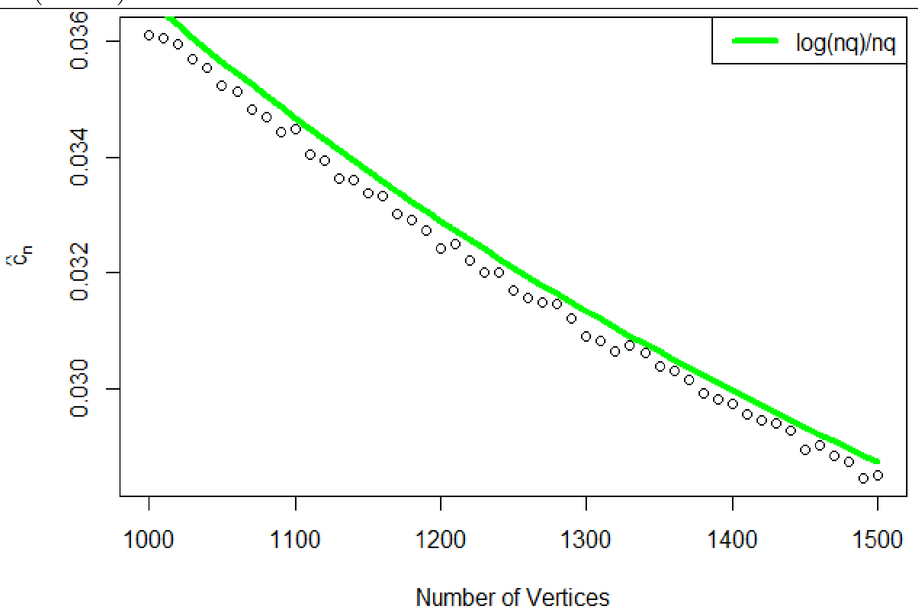
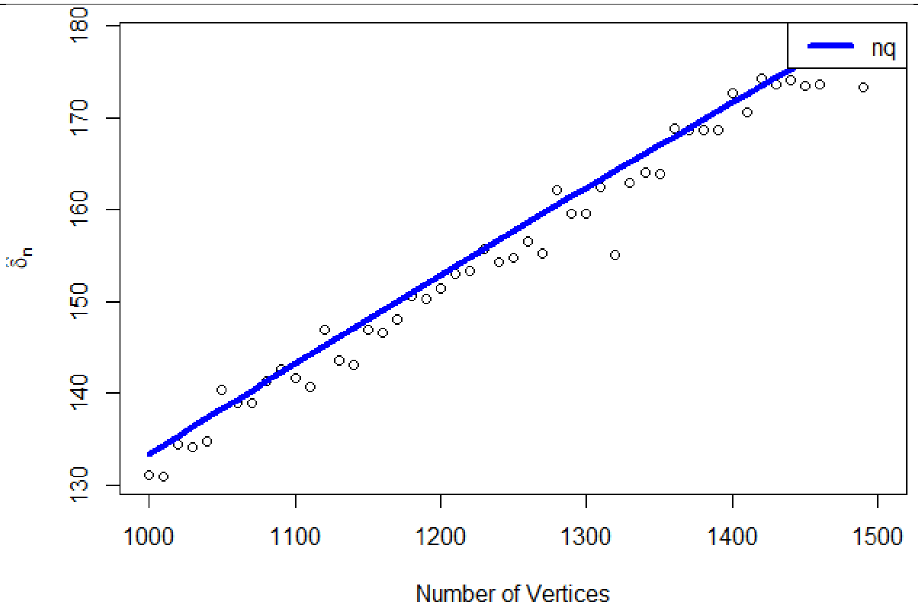
- $n = 1000 - 1500$ and $q_n = 0.3$

$n\ vs\ \hat{c}_n$ plot	
$Cor\left(\hat{c}_n, \log nq_n/nq_n\right)$	0.9996224
$n\ vs\ \hat{\delta}_n$ plot	
$Cor\left(\hat{\delta}_n, nq_n\right)$	0.9905715

- $n = 1000 - 1500$ and $q_n = (n+1)^{(n+1)^{\frac{7}{8}}} - n^{\frac{7}{8}}$

$n \ vs \ \hat{c}_n$ plot	
$Cor\left(\hat{c}_n, \log nq_n / nq_n\right)$	0.9998839
$n \ vs \ \hat{\delta}_n$ plot	
$Cor\left(\hat{\delta}_n, nq_n\right)$	0.9968299

- $n = 1000 - 1500$ and $q_n = (n+1)^{\frac{3}{4}} - n^{\frac{3}{4}}$

<table>
<tr><td>$n\ vs\ \hat{c}_n$ plot</td><td>
</td></tr>
<tr><td>$Cor\left(\hat{c}_n, \log nq_n/nq_n\right)$</td><td>0.9996602</td></tr>
<tr><td>$n\ vs\ \hat{\delta}_n$ plot</td><td>
</td></tr>
<tr><td>$Cor\left(\hat{\delta}_n, nq_n\right)$</td><td>0.9945886</td></tr>
</table>

Observations

- $Cor\left(\hat{c}_n, \log nq_n/nq_n\right)$ and $Cor\left(\hat{\delta}_n, nq_n\right)$ are quite high.
- Linear association of $\hat{\delta}_n, nq_n$ and $\hat{c}_n, \log(nq_n)/(nq_n)$ is positive in all the cases.

4 Discussion and Conclusions

In view of the above simulations we propose the following conjectures under the following set-up

$$
\begin{aligned}
(X_v)_{v\in \mathcal{V}_n} &\sim \text{i.i.d. Bernoulli}\,(q_n) \\
(Y_e)_{e\in \mathcal{E}_n} &\sim \text{i.i.d. Bernoulli}\,(p_n) \\
X_v \;\&\; Y_e &\text{ are independent } \forall v \in \mathcal{V}_n \;\&\; e \in \mathcal{E}_n
\end{aligned}
$$

Conjecture 4.1 (Connectivity "Threshold" after Percolation)**.** *Let G be random graph such that*

$$
\begin{aligned}
V(G) &:= \{v | v \in \mathcal{V}_n \;\&\; X_v = 1\} \\
E(G) &:= \{e := \{u, v\} | e \in \mathcal{E}_n \;\&\; X_v = 1, X_u = 1, Y_e = 1\}
\end{aligned}
$$

and

$$
\lim_{n\to\infty} \mathbb{E}[|V(G)|] = \infty
$$
$$
i.e.\ nq_n \to \infty \ as \ n \to \infty
$$

Then,

$$
\lim_{n\to\infty} \mathbb{P}\left(G \text{ is connected}\right) = \begin{cases} 0 & ; \dfrac{p_n}{\log nq_n/nq_n} \to 0 \\ 1 & ; \dfrac{p_n}{\log nq_n/nq_n} \to \infty \end{cases}
$$

Conjecture 4.2. *Fix $t \in \mathbb{R}$ and $\lambda_n = \log nq_n + t$. Provided that $nq_n \to \infty$ as $n \to \infty$,*

$$
\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}\left(n, \frac{\lambda_n}{nq_n}, q_n\right)}(G \text{ is connected}) = \exp\left\{-\exp\left\{-t\right\}\right\}
$$

Appendix-1

Proof of the Results, Stated in Section 3

Result 3.1. $\mathbb{E}\left[|V(G)|\right] = nq_n$

Proof. We know,

$$\begin{aligned} &(X_v)_{v \in \mathcal{V}_n} \sim \text{i.i.d. Bernoulli}\,(q_n) \\ \implies &\mathbb{E}[X_v] = q_n \;\forall\; v \in \mathcal{V}_n \end{aligned}$$

and,

$$\begin{aligned} V(G) &:= \{v | v \in \mathcal{V}_n \;\&\; X_v = 1\} \\ \therefore |V(G)| &= \sum_{v \in \mathcal{V}_n} X_v \\ \therefore \mathbb{E}\left[|V(G)|\right] &= \mathbb{E}\left[\sum_{v \in \mathcal{V}_n} X_v\right] = \sum_{v \in \mathcal{V}_n} \mathbb{E}\left[X_v\right] = |\mathcal{V}_n| \cdot \mathbb{E}[X_1] = nq_n \end{aligned}$$

□

Result 3.2. $\mathbb{P}\left(e \in E(G)\right) = q_n^2 p_n \;\forall e \in \mathcal{E}_n$

Proof. An edge $e := \{u, v\}$ is present in the set $E(G)$ if

$$X_u = 1,\; X_v = 1 \;\&\; Y_e = 1,\; \text{where } X_u, X_v \sim \text{Bernoulli}(q_n) \;\&\; Y_e \sim \text{Bernoulli}(p_n)$$
$$X_u,\; X_v,\; Y_e \text{ are independent Random Variables.}$$

Therefore,

$$\begin{aligned} \mathbb{P}\left(e \in E(G)\right) &= \mathbb{P}\left(X_u = 1, X_v = 1, Y_e = 1\right) \\ &= \mathbb{P}\left(X_u = 1\right)\mathbb{P}\left(X_v = 1\right)\mathbb{P}\left(Y_e = 1\right) \\ &= q_n \cdot q_n \cdot p_n \\ &= q_n^2 p_n \end{aligned}$$

□

Result 3.3. $\mathbb{P}\left(e=\{u,v\}\in E(G)|X_u=1,X_v=1\right)=p_n$

Proof.

$$\begin{aligned}\mathbb{P}\left(e=\{u,v\}\in E(G)|X_u=1,\ X_v=1\right)&=\mathbb{P}\left(X_u=1,\ X_v=1,\ Y_e=1|X_u=1,\ X_v=1\right)\\&=\mathbb{P}\left(Y_e=1\right)[\because X_u,\ X_v,\ Y_e \text{ are independent}]\\&=p_n\end{aligned}$$

□

Result 3.4. *If vertex percolation is done first, the graph obtained is an Erdős-Rényi Binomial Random Graph on the remaining set of vertices i.e.,* $G|(X_v)_{v\in\mathcal{V}_n}\sim\mathcal{G}\left(|V(G)|,p_n\right)$

Proof. Given, $(X_v)_{v\in\mathcal{V}_n}$, $V(G)$ is a fixed set.

$$\begin{aligned}V(G)&=\{v|X_v=1\}\\|V(G)|&=\sum_{v\in\mathcal{V}_n}X_v\end{aligned}$$

and

$$\begin{aligned}&e:=\{s,t\}\in\mathcal{E}_n\ \textit{such that}\ X_s=0;X_t=0\\\implies&e\notin E(G)\\&i.e.,\ v,\ u\ (v\neq u)\notin V(G)\implies\{u,v\}\notin E(G)\end{aligned}$$

and if $u,\ v\ (u\neq v)\in V(G)$,

$$\begin{aligned}&\mathbb{P}(\{u,v\}\in E(G)|X_u=1,\ X_v=1)=p_n\ \text{[by previous result]}\\&(Y_e)_{e\in\mathcal{E}_n}\text{ are i.i.d. Bernoulli}(p_n)\\\implies&G|(X_v)_{v\in\mathcal{V}_n}\sim\mathcal{G}\left(|V(G)|,p_n\right)\end{aligned}$$

□

Bibliography

1. R. van der Hofstad: *Random Graphs and Complex Networks Volume I & II*
2. `https://en.wikipedia.org/wiki/Percolation_theory`