Variance of the giant component of a random graph

Peter Neal

University of Manchester

13 September 2011

Joint work with Frank Ball (University of Nottingham)

Overview

1. Introduction – random graphs

2. Variance calculations

3. Numerical example

4. Questions – random graphs and epidemics

1. Introduction – random graphs

Configuration model. n vertices.

Vertex i has degree D_i. (D_i half-edges.)

Half-edges connected, uniformly at random, to form edges, and hence, the graph.

Interested in the asymptotic behaviour of the graph as $n \to \infty$.

Let $\bar{\mathbf{D}}^n = (\bar{D}_1, \bar{D}_2, \ldots, \bar{D}_n)$ with $\mathbf{D}^n = (D_1, D_2, \ldots, D_n)$ a random permutation of $\bar{\mathbf{D}}^n$.

Molloy-Reed: $\bar{\mathbf{D}}^n$, a deterministic sequence with $\frac{1}{n}\sum_{i=1}^{n} 1_{\{\bar{D}_i = k\}} \longrightarrow \pi_k$.

Newman-Strogatz-Watts: $\bar{D}_i$ *iid* according to D.

Questions of interest – random graphs (epidemics)

Size of components. Does a giant component exist?

$$\text{Yes, if } \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1.$$

Size (mean) of giant component, R_n: ρn, where

$$\rho = 1 - f(z),$$

and z is the solution in $[0, 1)$

$$z = \frac{1}{\mu} f'(z),$$

where for $0 \leq s \leq 1$, $f(s) = \mathbb{E}[s^D]$ and $f' = \mathbb{E}[Ds^{D-1}]$.

Precisely,

$$\frac{1}{n} R_n \xrightarrow{p} \rho \quad \text{as } n \to \infty.$$

Question: What is the (limiting) distribution of R_n?

Question: What is the (limiting) variance of R_n?

Erdös-Rényi random graph

n vertices.

Probability of an edge between two vertices: $\mu/(n-1)$.

If $\mu > 1$ (supercritical):

$$\sqrt{n}\left(\frac{1}{n}R_n - \rho\right) \xrightarrow{D} N(0, \sigma^2) \quad \text{as } n \to \infty, \tag{1}$$

where ρ is the non-zero solution of $\rho = 1 - \exp(-\mu\rho)$ and $\sigma^2 = \frac{\rho(1-\rho)}{(1-\mu(1-\rho))^2}$.

2. Variance Theorem

For a supercritical Newman-Strogatz-Watts random graph with $\mathbb{E}[D^{12}] < \infty$,

$$var\left(\sqrt{n}\left(\frac{R_n}{n} - \rho\right)\right) \to \sigma^2 \text{ as } n \to \infty, \tag{2}$$

where

$$\sigma^2 = \rho(1-\rho) + \frac{z^2}{1 - f''(z)/\mu}\mu(1+z^2) + \frac{z^4}{(1-f''(z)/\mu)^2}(\mu(\nu-2) + f''(z))$$

and $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$.

For a supercritical Molloy-Reed random graph with $\frac{1}{n}\sum_{i=1}^{n}\bar{D}_i^{12} \to \mathbb{E}[D^{12}] < \infty$, (2) holds with σ_{MR}^2 replaced by

$$\begin{aligned}\sigma_{MR}^2 &= 1-\rho-f(z^2)+\frac{z^2}{1-f''(z)/\mu}\{(1+z^2)\mu-2f'(z^2)\} \\ &\quad +\frac{z^2}{(1-f''(z)/\mu)^2}\{z^2\mu+z^2f''(z)-f'(z^2)-z^2f''(z^2)\}.\end{aligned}$$

Proof.

Let $U_n = n - R_n$, total number of vertices *outside* the giant component.

Let $\mathcal{C}_i^n$ denote the component containing vertex i, and let $C_i^n = |\mathcal{C}_i^n|$.

Fix $0 < \beta < \frac{1}{10}$ and let $\chi_i^n = 1_{\{C_i^n \leq [n^\beta]\}}$.

$$\tilde{U}_n = \sum_{i=1}^{n} \chi_i^n$$

Since the second largest component of a supercritical random graph is of size $O(\log n)$, $|U_n - \tilde{U}_n| \xrightarrow{p} 0$ as $n \to \infty$.

Focus upon the variance of $\tilde{U}_n$.

Coupled branching processes

Condition upon $\bar{\mathbf{D}}^n = (\bar{D}_1^n, \bar{D}_2^n, \ldots, \bar{D}_n^n)$.

Let $\pi_k^n = \frac{1}{n}\sum_{i=1}^n 1_{\{\bar{D}_i = k\}}$, $\mu_n = \frac{1}{n}\sum_{i=1}^n D_i$ and let $\pi_k = \mathbb{P}(D_1 = k)$.

Branching process $\mathcal{B}_1^n$.

Initial ancestor: k offspring with probability π_k^n.

Subsequent individuals: $k-1$ offspring with probability $k\pi_k^n/\mu_n$.

B_1^n, total progeny and E_1^n, event the branching process goes extinct.

Branching process $\mathcal{B}$.

Initial ancestor: k offspring with probability π_k.

Subsequent individuals: $k - 1$ offspring with probability $k\pi_k/\mu$.

B, total progeny and E, event the branching process goes extinct.

$$C_1^n | C_1^n \le [n^\beta] \xrightarrow{D} B|E \quad \text{as } n \to \infty.$$

$$var(\sqrt{n}(\tilde{U}_n/n - (1-\rho)))$$

Exploiting exchangeability of the vertices,

$$\begin{aligned} var(\sqrt{n}(\tilde{U}_n/n - (1-\rho))) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} cov(\chi_i^n, \chi_j^n) \\ &= var(\chi_1^n) + (n-1)cov(\chi_1^n, \chi_2^n). \end{aligned}$$

Now $var(\chi_1^n) \to (1-\rho)\rho$ and

$$(n-1)cov(\chi_1^n, \chi_2^n) = (n-1)\mathbb{E}[cov(\chi_1^n, \chi_2^n|\bar{\mathbf{D}}^n)] + var(\sqrt{n-1}\mathbb{E}[\chi_1^n|\bar{\mathbf{D}}^n]).$$

(The second term is zero for the Molloy-Reed random graph.)

$$(n-1)\mathbb{E}[cov(\chi_1^n, \chi_2^n|\bar{\mathbf{D}}^n)]$$

$$\begin{aligned}
&(n-1)cov(\chi_1^n, \chi_2^n|\bar{\mathbf{D}}^n) \\
= \ &(n-1)\mathbb{E}\left[1_{\{C_1^n \le [n^\beta]\}}\left(\mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n, \mathcal{C}_1^n\right] - \mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]\right)\Big|\bar{\mathbf{D}}^n\right].
\end{aligned}$$

Two parts: $2 \in \mathcal{C}_1^n$ and $2 \notin \mathcal{C}_1^n$.

Note that, if $2 \in \mathcal{C}_1^n$, $\mathbb{E}[1_{\{\mathcal{C}_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n, \mathcal{C}_1^n, 2 \in \mathcal{C}_1^n] = 1_{\{\mathcal{C}_1^n \le [n^\beta]\}}$, whereas $\mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]$ does not depend upon $\mathcal{C}_1^n$.

$$\begin{aligned}
&(n-1)cov(\chi_1^n, \chi_2^n|\bar{\mathbf{D}}^n) \\
= \ &\mathbb{E}\left[1_{\{C_1^n \le [n^\beta]\}}(C_1^n - 1)\left(1_{\{C_1^n \le [n^\beta]\}} - \mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]\right)\Big|\bar{\mathbf{D}}^n\right] \\
+ \ &\mathbb{E}\left[(n - C_1^n)1_{\{C_1^n \le [n^\beta]\}}\left(\mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n, 2 \notin \mathcal{C}_1^n\right] - \mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]\right)\Big|\bar{\mathbf{D}}^n\right].
\end{aligned}$$

$$\mathbb{E}\left[1_{\{C_1^n \le [n^\beta]\}}(C_1^n - 1)\left(1_{\{C_1^n \le [n^\beta]\}} - \mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]\right)\middle|\bar{\mathbf{D}}^n\right]$$

Given that $C_1^n \le [n^\beta]$,

$$\begin{aligned} 1_{\{C_1^n \le [n^\beta]\}} - \mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right] &\xrightarrow{p} 1 - \mathbb{E}[1_{\{E_1\}}] \\ &= 1 - (1-\rho) = \rho. \end{aligned}$$

Also,

$$\mathbb{E}[1_{\{C_1^n \le [n^\beta]\}}(C_1^n - 1)|\bar{\mathbf{D}}^n] \xrightarrow{p} \mathbb{E}[1_{\{E_1\}}(B_1 - 1)].$$

Hence, the limit is

$$\rho\mathbb{E}[1_{\{E_1\}}(B_1 - 1)] = \sigma_1^2.$$

$2 \notin \mathcal{C}_1^n$

Construct two (coupled) realisations of $\mathcal{C}_2^n$ using $\bar{\mathbf{D}}^n$.

Realisation 1: Taking account of $\mathcal{C}_1^n$ and $2 \notin \mathcal{C}_1^n$. Size $\check{C}_2^n$.

Realisation 2: Independent of $\mathcal{C}_1^n$. Size $C_2^{n,I}$.

Let $H_1^n = \sum_{j \in \mathcal{C}_1^n} \bar{D}_j (\approx 2(C_1^n - 1))$.

(Approximate) probability a chosen vertex differs between the two realisations:

Initial vertex: C_1^n/n (choose a vertex in $\mathcal{C}_1^n$)

Subsequent vertices: $H_1^n/n\mu_n$ (choose a half-edge in $\mathcal{C}_1^n$)

1. Replace $\check{C}_2^n$ and $C_2^{n,I}$ by branching process approximations total progenies of branching process approximations $\check{B}_2^n$ and $B_2^{n,I}$.
(Extinction $\check{E}_2^n$ and $E_2^{n,I}$.)

2. Note that

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[(n - C_1^n)1_{\{C_1^n \le [n^\beta]\}}\left(\mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n, 2 \notin \mathcal{C}_1^n\right] - \mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]\right)\middle|\,\bar{\mathbf{D}}^n\right]\right] \\
\approx\ & n\mathbb{E}\left[\mathbb{E}\left[1_{\{C_1^n \le [n^\beta]\}}\left(\mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n, 2 \notin \mathcal{C}_1^n\right] - \mathbb{E}\left[1_{\{C_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]\right)\middle|\,\bar{\mathbf{D}}^n\right]\right] \\
\approx\ & n\mathbb{E}\left[\mathbb{E}\left[1_{\{C_1^n \le [n^\beta]\}}\left(\mathbb{E}\left[1_{\{\check{B}_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n, 2 \notin \mathcal{C}_1^n\right] - \mathbb{E}\left[1_{\{B_2^n \le [n^\beta]\}}|\bar{\mathbf{D}}^n\right]\right)\middle|\,\bar{\mathbf{D}}^n\right]\right] \\
\approx\ & n\mathbb{E}[\mathbb{E}[1_{\{C_1 \le n^\beta\}}(\mathbb{E}[1_{\{\check{E}_2^n\}}|\bar{\mathbf{D}}^n, \mathcal{C}_1^; 2 \notin \mathcal{C}_1^n] - \mathbb{E}[1_{\{E_2^{n,I}\}}|\bar{\mathbf{D}}^n])|\bar{\mathbf{D}}^n]]
\end{aligned}$$

3. Comparing the extinction probabilities between two branching processes with a small, $O(1/n)$, mutation rate. Limit

$$\mathbb{E}\left[\left\{Bf(z) - \sum_{k=1}^{\infty} b_k z^k + \frac{z}{1 - f''(z)/\mu} \sum_{k=1}^{\infty} k b_k (z - z^{k-1})\right\} 1_{\{E_1\}}\right] = \sigma_2^2,$$

where b_k is the total number of vertices of degree k in $\mathcal{C}_1^n$.

$$var(\sqrt{n-1}\mathbb{E}[\chi_1^n|\bar{\mathbf{D}}^n])$$

$$\begin{aligned}\sqrt{n-1}(\mathbb{E}[\chi_1^n|\bar{\mathbf{D}}^n]-(1-\rho)) \;=\; & \sqrt{n-1}(\mathbb{P}(C_1^n \leq [n^\beta]|\bar{\mathbf{D}}^n)-\mathbb{P}(B_1^n \leq [n^\beta]|\bar{\mathbf{D}}^n)) \\ & +\sqrt{n-1}(\mathbb{P}(B_1^n \leq [n^\beta]|\bar{\mathbf{D}}^n)-\mathbb{P}(E_1^n|\bar{\mathbf{D}}^n)) \\ & +\sqrt{n-1}(\mathbb{P}(E_1^n|\bar{\mathbf{D}}^n)-(1-\rho)).\end{aligned}$$

The first two terms converge (in probability) to 0 as $n \to \infty$ and $1-\rho = \mathbb{P}(E_1)$.

This reduces the problem to comparing the extinction probabilities of two branching processes.

$$\sqrt{n-1}(\mathbb{P}(E_1^n|\bar{\mathbf{D}}^n)-\mathbb{P}(E_1)).$$

Let $y_n = \mathbb{P}(E_1^n|\bar{\mathbf{D}}^n)$ and $y = \mathbb{P}(E_1)$. Then

$$y_n = f_n(z_n)$$

$$y = f(z)$$

where

$$z_n = \frac{1}{\mu_n} f_n'(z_n)$$

$$z = \frac{1}{\mu} f'(z)$$

and $f_n(s) = \frac{1}{n}\sum_{i=1}^n s^{\bar{D}_i}$ and $f_n'(s) = \frac{1}{n}\sum_{i=1}^n \bar{D}_i s^{\bar{D}_i - 1}$

First step, using the central limit theorem,

$$\sqrt{n-1}(\mathbb{P}(E_1^n|\bar{\mathbf{D}}^n)-\mathbb{P}(E_1)) \xrightarrow{D} N\left(0, var\left(z^D + \frac{z}{1-f''(z)/\mu}D(z^{D-1}-z)\right)\right).$$

Second step, uniform integrability,

$$\begin{aligned} var(\sqrt{n-1}(\mathbb{E}[\chi_1^n|\bar{\mathbf{D}}^n]-(1-\rho))) &\approx var(\sqrt{n-1}(\mathbb{P}(E_1^n|\bar{\mathbf{D}}^n)-\mathbb{P}(E_1))) \\ &\to var\left(z^D + \frac{z}{1-f''(z)/\mu}D(z^{D-1}-z)\right) \\ &= \sigma_3^2. \end{aligned}$$

Hence, for the Newman-Strogatz-Watts random graph,

$$var\left(\sqrt{n}\left(\frac{\tilde{U}_n}{n}-(1-\rho)\right)\right)\to\rho(1-\rho)+\sigma_1^2+\sigma_2^2+\sigma_3^2$$

and for the Molloy-Reed random graph

$$var\left(\sqrt{n}\left(\frac{\tilde{U}_n}{n}-(1-\rho)\right)\right)\to\rho(1-\rho)+\sigma_1^2+\sigma_2^2.$$

The Theorem is completed by showing that

$$|var(U_n/\sqrt{n})-var(\tilde{U}_n/\sqrt{n})|\to 0$$

and computing σ_1^2, σ_2^2 and σ_3^2.

3. Numerical example

$n = 1000$.

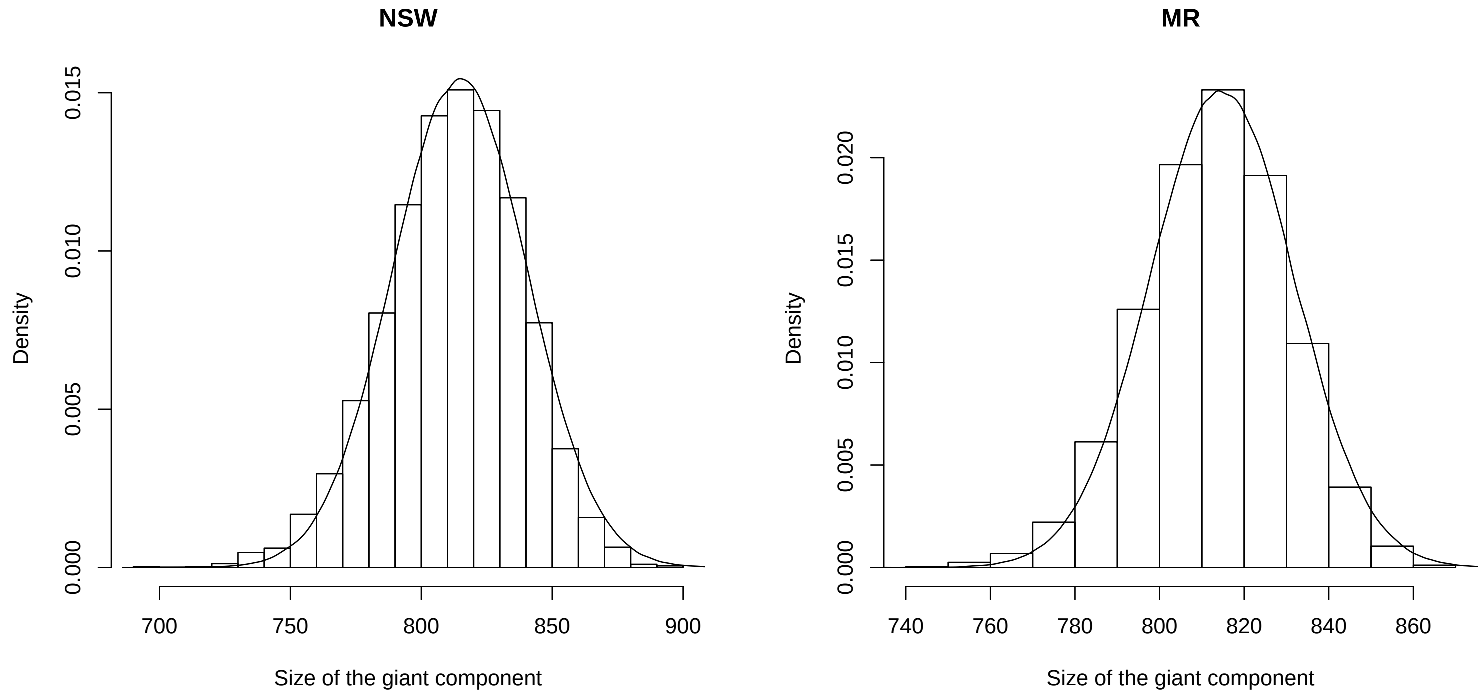
Four choices of degree distribution:-

1. $\mathbb{P}(D = 1) = \mathbb{P}(D = 3) = 1/2$;
2. $\mathbb{P}(D = k) = 1/4 \ (k = 1, 2, 3, 4)$;
3. $D \sim \text{Po}(\mu)$ with $\mu = 2$;
4. $\mathbb{P}(D = k) \propto k^{-m} \ (k = 2, 3, \ldots, n)$ and $\mathbb{P}(D = 1) = 2\mathbb{P}(D = 3)$ with $m = 4$.

10000 Simulations

	NSW random graph		MR random graph	
D	Simulations	Theoretical	Simulations	Theoretical
1	Mean = 813.9	$n\rho = 814.8$	Mean = 815.5	$n\rho = 814.8$
	Var = 664.8	$n\sigma^2 = 644.7$	Var = 271.8	$n\sigma^2_{MR} = 293.6$
2	Mean = 961.6	$n\rho = 961.7$	Mean = 961.2	$n\rho = 961.7$
	Var = 99.4	$n\sigma^2 = 101.4$	Var = 79.2	$n\sigma^2_{MR} = 77.3$
3	Mean = 796.4	$n\rho = 796.8$	Mean = 796.8	$n\rho = 796.8$
	Var = 318.7	$n\sigma^2 = 310.7$	Var = 130.5	$n\sigma^2_{MR} = 136.5$
4	Mean = 890.2	$n\rho = 890.6$	Mean = 887.3	$n\rho = 890.6$
	Var = 659.3	$n\sigma^2 = 660.9$	Var = 361.7	$n\sigma^2_{MR} = 353.0$

CLT?



Histograms of R_n with $n = 1000$, based upon a sample of size 10000, plotted against density plots of $N(n\rho, n\sigma^2)$ with degree distribution 1.

Questions from numerical example

1. Central limit theorem for the size of the giant component.

2. Difference between Newman-Strogatz-Watts random graph with $D \sim \text{Po}(\mu)$ and Erdös-Rényi random graph with degree distribution $\text{Po}(\mu)$.

   For example, $\mu = 2$.

   Newman-Strogatz-Watts random graph.

$$\mathbb{E}\left[\frac{1}{n}R_n\right] \longrightarrow 0.7968, \qquad var\left(\sqrt{n}\left(\frac{1}{n}R_n - \rho\right)\right) \longrightarrow 0.3107$$

   Erdös-Rényi random graph.

$$\mathbb{E}\left[\frac{1}{n}R_n\right] \longrightarrow 0.7968, \qquad var\left(\sqrt{n}\left(\frac{1}{n}R_n - \rho\right)\right) \longrightarrow 0.4595$$

3. Are any of the methods/ideas applicable for subcritical or critical random graphs?

Questions for epidemics on random graphs/networks

1. Epidemics upon random graphs (directed random graphs). Variance of the size of a large epidemic outbreak.

2. Applicability of the above approach to other structures (without exchangeability). Assessing whether or not a particular pair of individuals belong to a giant component/major epidemic.

   Ball and Neal (2008). Reed-Frost epidemic on a network with $D \equiv d \geq 3$. (Distance between individuals/vertices.)

   Ball and Neal (2010). Standard (homogeneously mixing) Reed-Frost epidemic. (Exchangeability.)