Centrality Testing and the Distribution of the Degree Variance in Bernoulli Graphs

Jan Hagberg*

Abstract

Exact and asymptotic distributions of the degree variance are investigated for Bernoulli graphs and uniform random graphs. In particular the range of values of the degree variance and its maximum value are considered. We show that the degree variance is approximately gamma distributed with parameters obtained from the first two moments of the degree variance. Since centrality of a graph can be interpreted as a measure of its heterogeneity in terms of vertex degrees, we can perform a centrality test with a critical value obtained from the gamma distribution.

Key words: Centrality Testing, Bernoulli Graphs, Degree Variance, Gamma Approximation, Uniform Random Graphs.

*Department of Statistics, Stockholm University, S-106 91 Stockholm. E-mail: Jan.Hagberg@stat.su.se. The author would like to thank Ove Frank for very helpful comments. Financial support from the Swedish Council of Research in Humanities and Social Sciences (HSFR), grant No. F0750/96, is gratefully acknowledged.

Contents

1 Introduction

Several measures of graph centrality have been developed over the years, see for example Freeman (1978), Snijders (1981a) and Wasserman&Faust (1994). According to Buckley (1997) the motivation for their development has generally been to find an appropriate way to measure where the "middle" of a certain graph is. One of the reasons why such a rich variety of centrality measures has arisen is that different applications have produced different ideas on what the "middle" should be. Hence, there is no general centrality concept. This paper is concerned with a standard statistical measure; the degree variance in undirected graphs. Thus, centrality of a graph can be interpreted as a measure of its heterogeneity in terms of vertex degrees.

Consider a graph on n vertices and r edges, i.e. a graph of order n and size r, and let x_i be the degree of vertex i, i.e. the number of edges incident to vertex i. Denote the mean degree by $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 2r/n$. The degree variance is defined as

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

How shall we interpret a value s^2 obtained from a given graph? Regular graphs, i.e. graphs for which all degrees are equal, have clearly the property of non-centrality, but, how much must the given graph departure from regularity in order to be regarded as central? There are many reasons why there is no simple answer to the question. One of them is the counter question "couldn't the observed value s^2 be obtained by mere chance?"

In order to make a more precise statement about the centrality, we will assume that no centrality is present if the edges are generated according to either a Bernoulli probability model or a uniform random graph model. The uniform random graph model is equivalent to a Bernoulli model conditional on the number of edges. Random variables will be denoted with uppercase letters and the realized values of a variable will be denoted by the corresponding lowercase letter. Thus, the random variable S^2 can take the value s^2. By deriving an approximate probability distribution of S^2 given that the null hypothesis H_0: *no centrality* is true, we can assess a critical value to test H_0 against alternatives with centrality. The null hypothesis will be rejected if the observed value s^2 is large enough, that is, if the probability value $P\left(S^2 \geqslant s^2 | H_0 \textit{ is true}\right)$ is small enough. Hence, it is of great importance to investigate the upper part of the approximate distribution e.g. the upper 10% tail.

Note that we have to reject the Bernoulli model (and the uniform random graph model) if the null hypothesis is rejected. In fact the null hypothesis of no centrality is modeled by the Bernoulli graph. But, departures from the Bernoulli graph does not imply centrality. Centrality here means, that the degrees vary more than they do in Bernoulli or conditional Bernoulli graphs. If the degrees vary less, as they do for instance in regular graphs, then this is not considered to be a violation of H_0.

The two models of this paper are simple and easy to interpret, but they don't deliver any distribution for the alternative hypothesis H_1*:centrality*. Hence, we can't calculate the power of the test. Tallberg (2000) has analyzed models for the alternative hypothesis. For the alternative hypothesis, Tallberg assumes a modified Bernoulli block model that generates edges with different probabilities within and between different blocks of vertices. See also Frank (2000).

2 Some distributional properties of Bernoulli graphs

We consider a graph on n labeled vertices. The vertices are labeled by integers $1, ..., n$. With probability p, each pair of distinct vertices i and j is connected by an edge. These connections are made independently of each other. Let X_i be the number of edges incident to vertex i and let R be the total number of edges. Since a vertex can have no more than $n-1$ edges it follows that

$$\begin{aligned} X_i &\in Bin(n-1, p) \ , \ i = 1, 2, ..., n \ , \\ R &\in Bin\left(\binom{n}{2}, p\right) \end{aligned} \tag{2.1}$$

and

$$\overline{X} = \frac{2R}{n} \in \frac{2}{n} Bin\left(\binom{n}{2}, p\right) . \tag{2.2}$$

Let $q = 1 - p$. From the properties of the binomial distribution it follows that

$$E\left(X_i\right) = (n-1)\,p \qquad , \quad Var\left(X_i\right) = (n-1)\,pq,$$

and since $\sum X_i = 2R$,we have

$$Var\left(\sum_{i=1}^{n} X_i\right) = 4\binom{n}{2}pq = 2n\,(n-1)\,pq. \tag{2.3}$$

The last variance can also be expanded as $\sum\sum Cov(X_i, X_j) = nVarX_i + n\,(n-1)\,Cov\,(X_i, X_j)$ and it follows that $Cov(X_i, X_j) = pq$ so that

$$Corr(X_i, X_j) = \frac{1}{n-1}. \tag{2.4}$$

Thus, we see that the correlation between X_i and X_j tends to zero for increasing n.

We will estimate the parameter p by its maximum likelihood (ML) estimator

$$\widehat{P} = \frac{R}{\binom{n}{2}}. \tag{2.5}$$

We have

$$E(\widehat{P}) = p \ \text{ and } Var(\widehat{P}) = \frac{2pq}{n(n-1)},$$

and $\widehat{P}$ is approximately normally distributed if $n\,(n-1)\,pq$ is sufficiently large. Hence

$$\widehat{p} \pm 2\sqrt{\frac{\widehat{p}\,(1-\widehat{p})}{\binom{n}{2}}} \tag{2.6}$$

is an approximate 95% confidence interval for p.

Conditional on the number of edges, X_i is hypergeometricly distributed i.e.

$$\begin{aligned} P(X_i &= x_i \mid R = r) = \frac{\binom{n-1}{x_i}\binom{\binom{n}{2}-(n-1)}{r-x_i}}{\binom{\binom{n}{2}}{r}} \\ x_i &= 0, 1, ..., \min\{n-1, r\} \text{ for } i = 1, ..., n. \end{aligned} \tag{2.7}$$

It follows that

$$E(X_i \mid R = r) = \frac{2r}{n} \quad , \quad Var(X_i \mid R = r) = \frac{2r(n^2 - n - 2r)}{n^2(n+1)} \tag{2.8}$$

and since $Var(\sum X_i \mid R = r) = 0$, we have

$$0 = nVar(X_i \mid R = r) + n(n-1)Cov(X_i, X_j \mid R = r)$$

which implies

$$Cov(X_i, X_j \mid R = r) = -\frac{Var(X_i \mid R = r)}{n-1}$$

and

$$Corr(X_i, X_j \mid R = r) = -\frac{1}{n-1} \quad \text{for } i \neq j. \tag{2.9}$$

Thus, the correlation is negative and has the same absolute value as in the unconditional case. We also note that the simultaneous distribution of $(X_1, ..., X_n \mid R = r)$ is multivariate hypergeometric. See, for instance, Johnson, Kotz & Balakrishnan (1997).

3 Moments of the degree variance

The moments of the degree variance S^2 are the building blocks of the results in this paper. Some of them are complicated to derive and these derivations are left in Appendix A. To avoid fractions, we will often use the integer valued random variable $Z = n^2S^2 = n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i<j}(X_i - X_j)^2$. Denote the kth moment of Z by $m_k = E\left(Z^k\right)$. The first three moments are given below,

where the falling factorial $n(n-1)\cdots(n-j+1) = n!/\,(n-j)!$ is denoted by $n_{(j)}$.

$$\begin{aligned}
m_1 &= E\sum_{i<j}(X_i - X_j)^2 \\
&= n\,(n-1)\,E\left(X_i^2 - X_iX_j\right) = n\,(n-1)\,((n-1)\,pq - pq) \\
&= n_{(3)}pq \qquad (3.1)
\end{aligned}$$

$$m_2 = 2\,(n-2)\,n_{(3)}pq + (n-2)\,(n+4)\,n_{(4)}p^2q^2 \qquad (3.2)$$

$$\begin{aligned}
m_3 &= 4n\,(n-1)\,(n-2)^3\,pq \\
&\quad +2n_{(4)}\,(3n-4)\,((n-2)\,(n+6)-8)\,p^2q^2 \\
&\quad +n_{(4)}\left[n^4\,(n+3) - 4\,(3n-4)\,[3\,(n-2)\,(n+6)-(n+4)]\right]p^3q^3
\end{aligned} \qquad (3.3)$$

It follows from m_1 and m_2 that

$$E\left(S^2\right) = \frac{(n-1)\,(n-2)}{n}pq \qquad (3.4)$$

and

$$Var\left(S^2\right) = \frac{2(n-1)(n-2)^2}{n^3}pq\,(1+(n-6)pq)\,. \qquad (3.5)$$

If the X_i are approximated by independent normal variables we get $nS^2/\,(n-1)\,pq$ asymptotic χ^2_{n-1}, and it follows that $((E\,(S^2))\,/n) \to pq$ and $((Var\,(S^2))/n) \to 2p^2q^2$, which agree with (3.4) and (3.5). It can also be shown that

$$Cov\left(S^2, \overline{X}\right) = \frac{2\,(n-1)\,(n-2)\,(q-p)}{n^2}pq \qquad (3.6)$$

and

$$\begin{aligned}
Corr\left(S^2, \overline{X}\right) &= \frac{q-p}{\sqrt{1+(n-6)\,pq}} \qquad (3.7) \\
&= \pm\sqrt{\frac{2}{n}} \text{ if } pq = \frac{1}{6}.
\end{aligned}$$

Further, let

$$
\begin{aligned}
S_i^2 &= \frac{1}{n}\left(X_i - \overline{X}\right)^2, \\
Cov\left(S_i^2, S_j^2\right) &= Cov\left(\left(\frac{1}{n}\left(X_i - \overline{X}\right)^2\right), \left(\frac{1}{n}\left(X_j - \overline{X}\right)^2\right)\right) \quad , i \neq j \text{ and} \\
Corr\left(S_i^2, S_j^2\right) &= \frac{Cov\left(S_i^2, S_j^2\right)}{\sqrt{Var\left(S_i^2\right)}\sqrt{Var\left(S_j^2\right)}} = \frac{Cov\left(S_i^2, S_j^2\right)}{Var\left(S_i^2\right)} \quad , i \neq j.
\end{aligned}
$$

By use of the technique outlined in Appendix A, it can be shown that

$$
\begin{aligned}
Var\left(S_i^2\right) &= \frac{(n-1)(n-2)((n-2)(n-4)+4)}{n^5} pq \\
&+ \frac{2(n-1)(n-2)\left((n-1)_{(3)} + 9(n-3) - 3\right)}{n^5} p^2 q^2
\end{aligned} \tag{3.8}
$$

and

$$
\begin{aligned}
Cov\left(S_i^2, S_j^2\right) &= \frac{(n-2)\left[(n+6)(n-2) - 2n\right]}{n^5} pq \\
&- \frac{4(n-2)\left[(n+6)(n-2) - 6\right]}{n^5} p^2 q^2.
\end{aligned} \tag{3.9}
$$

We have

$$
Var\left(S_i^2\right) \to 2p^2q^2 \quad , \quad n^2 Cov\left(S_i^2, S_j^2\right) \to pq(1 - 4pq)
$$

and

$$
n^2 Corr\left(S_i^2, S_j^2\right) \to \frac{1 - 4pq}{2pq} = 1 \text{ if } pq = \frac{1}{6}. \tag{3.10}
$$

Hence, S_i^2 and S_j^2 are practically uncorrelated and S^2 can be regarded as a sum of almost uncorrelated random variables. For $n = 10$ we find that $Corr\left(S_i^2, S_j^2\right) = \frac{27 - 122pq}{9(282pq + 13)}$ and Figure 1 below shows how this correlation depends on p.

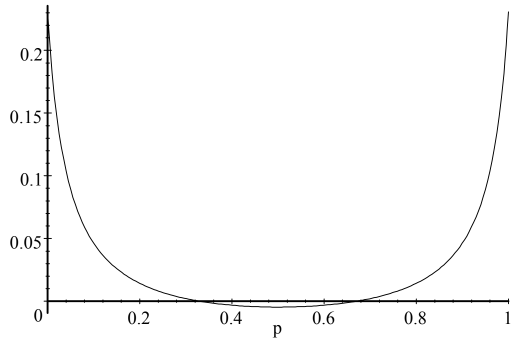

Figure 1. $Corr\left(S_i^2, S_j^2\right)$ for $n = 10$.

Further,

$$Cov\left(S_i^2, \overline{X}\right) = \frac{2(n-1)(n-2)pq(q-p)}{n^3}, \tag{3.11}$$

$$Corr\left(S_i^2, \overline{X}\right) = \left(\frac{2(n-2)(q-p)^2}{2\left((n-1)^2-1\right)(n-1)pq + \left((n-3)^2+3\right)(1-6pq)}\right)^{\frac{1}{2}} \tag{3.12}$$

and if $pq = \frac{1}{6}$ we have that $p = \frac{1}{2} \pm \frac{1}{6}\sqrt{3}$ and

$$Corr\left(S_i^2, \overline{X}\right) = \pm\left(\frac{2}{n(n-1)}\right)^{\frac{1}{2}} = \pm\binom{n}{2}^{-\frac{1}{2}} \text{ if } pq = \frac{1}{6}. \tag{3.13}$$

For the uniform random graph model or the R-conditional Bernoulli graph, we have according to (2.8) that

$$E\left(S^2 \mid R = r\right) = Var\left(X_i \mid R = r\right) = \frac{2r\left(n^2-n-2r\right)}{n^2(n+1)} \tag{3.14}$$

and

$$\begin{aligned} Var\left(S^2 \mid R = r\right) &= \frac{1}{n^2}\sum_i\sum_j Cov\left(X_i^2, X_j^2 \mid R = r\right) \\ &= \frac{8r(r-1)\left(n^2-n-2r\right)\left(n^2-n-2r-2\right)}{n^2(n+1)^2(n+2)\left(n^2-n-4\right)} \end{aligned} \tag{3.15}$$

The derivation of (3.10) can be found in Snijders (1981b) and a brief outline is given in Appendix A.

Finally, the same technique can be used to show that

$$n^3 Var\left(S_i^2 \mid R = r\right) \to 2r \quad , \quad n^4 Cov\left(S_i^2, S_j^2 \mid R = r\right) \to -2r \quad \text{for } i \neq j.$$

and

$$nCorr\left(S_i^2, S_j^2 \mid R = r\right) \to -1.$$

4 Extreme values and other attained values of the degree variance

4.1 The maximum value

The order values of the degree variance for a graph of order n is the set of possible values S^2 can attain such that

$$s_1^2 < s_2^2 < \cdots < s_m^2$$

where s_1^2 is the minimal value and s_m^2 is the maximal value.

The maximum value of the degree variance S^2 has been derived by Snijders (1981b). An alternative proof and a simple formula are given below, showing the exact structure of the graphs that attain the maximal degree variance.. Let $\lfloor x \rfloor$ denote the greatest integer less than or equal to the number x .

Theorem 1 *For Bernoulli graphs of order n, the maximal degree variance is given by*

$$\begin{aligned} s_{\max}^2 &= k\left(n-k\right)\left(\frac{n-k-1}{n}\right)^2 \\ \textit{where } k &= \left\lfloor \frac{n+1}{4} \right\rfloor . \end{aligned}$$

Theorem 2 *A disconnected Bernoulli graph of order n has maximal degree variance if and only if it consists of $\left\lfloor \frac{n+1}{4} \right\rfloor$ isolated vertices and a complete subgraph on the other vertices. A connected Bernoulli graph of order n has maximal degree variance if and only if it consists of $\left\lfloor \frac{n+1}{4} \right\rfloor$ vertices of degree $n-1$ and no other edges.*

Corollary 3

$$\lim_{n\to\infty} \frac{s_{\max}^2}{n^2} = \frac{3^3}{4^4}.$$

Corollary 4 *For Bernoulli graphs of order n the maximal degree variance is attained with probability*

$$\begin{aligned} P\left(S^2 = s_{\max}^2\right) &= \binom{n}{k}\left(p^{\binom{n-k}{2}} q^{\binom{n}{2}-\binom{n-k}{2}} + p^{\binom{n}{2}-\binom{n-k}{2}} q^{\binom{n-k}{2}}\right) \\ \textit{where } k &= \left\lfloor \frac{n+1}{4} \right\rfloor. \end{aligned}$$

For large n we have

$$c_1 (pq)^{7N/2}\left(p^N + q^N\right) \leqslant P\left(S^2 = s_{\max}^2\right) \leqslant c_2 (pq)^{7N/2}\left(p^N + q^N\right)$$

where $c_1 = \left(4 \times 3^{-3/4}\right)^n \sqrt{\frac{8}{3\pi n}} e^{-\frac{1+24n+117n^2}{(12n+1)(9n+1)(3n+1)}}$, $c_2 = \left(4 \times 3^{-3/4}\right)^n \sqrt{\frac{8}{3\pi n}} e^{-13/36n}$ *and* $N = \frac{n^2}{16}$.

Corollary 5 *Let p^* denote the value of p that is $\leqslant \frac{1}{2}$ and yields the highest probability to obtain the maximal degree variance in a Bernoulli graph of order n. Further, let $c_1 = \frac{k(2n-k-1)}{n(n-1)}$ and $c_2 = \frac{n(n-1-4k)+2k(k+1)}{n(n-1-4k)+2k(k+1)-2}$ where $k = \left\lfloor \frac{n+1}{4} \right\rfloor$. It holds that*

$$\min\left(c_1, c_1 c_2\right) < p^* \leqslant \max\left(c_1, c_1 c_2\right)$$

and $p^* = \frac{1}{2}$ for ten values of n only, i.e. for $n = 3, 4, 5, 7, 8, 9, 11, 12, 15, 19$. For increasing n, p^* tends to $7/16$.

Proof: Let G be a graph of order n with degree variance $S^2(G)$. Denote the complement of G by G^c. Since G^c has degrees $X_i^c = n - 1 - X_i$ and average degree $\overline{X}^c = n - 1 - \overline{X}$ we have that a graph and its complement have the same degree variance, i.e.

$$S^2(G) = S^2(G^c). \tag{4.1}$$

Multiply S^2 by n^2 and rewrite the variance formula according to (4.2)

$$n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i<j}\left(X_i - X_j\right)^2 \tag{4.2}$$

Focus on the right hand side of the equality above and assume that we have three distinct degree values, $d_1 < d_2 < d_3$. The sum $(d_2 - d_1)^2 + (d_3 - d_2)^2 + (d_3 - d_1)^2$ will then be included in (4.2). However, for positive $a = d_2 - d_1$ and $b = d_3 - d_2$, the inequality $a^2 + b^2 < (a+b)^2$ states that $(d_2 - d_1)^2 + (d_3 - d_2)^2 < (d_3 - d_1)^2$. Assume that we have c_1 vertices of degree d_1, c_2 vertices of degree d_2 and $n - c_1 - c_2$ vertices of degree d_3. We can increase the degree variance by removing all edges incident to the vertices of degree d_3 and d_2 and then make a complete subgraph of order $n - c_1$. The c_1 vertices previously of degree d_1 shall now be of degree $d_1' \leqslant d_1$. That is, exactly two distinct degree values is a necessary condition for maximum. Further, from (4.2), we see that one solution is to put $d_1 = 0$ and $d_2 > 0$. Let k be the number of vertices of degree 0, which implies that the other $n - k$ vertices must be of maximal degree $d_2 = n - k - 1$. This yields the degree variance

$$\begin{aligned} s_k^2 &= \frac{k\overline{x}}{n} + \frac{(n-k)\,(n-k-1-\overline{x})^2}{n} \\ \text{where } \overline{x} &= \frac{(n-k)\,(n-k-1)}{n} \end{aligned} \tag{4.3}$$

that is

$$s_k^2 = k\,(n-k)\left(\frac{n-k-1}{n}\right)^2$$

and it follows that $s_{\max}^2 = \max\limits_k s_k^2$. By writing

$$s_k^2 = f_n\,(k) = k\,(n-k)\left(1 - \frac{k+1}{n}\right)^2$$

it follows that

$$f_{n-1}\left(k\right) \leqslant f_n\left(k\right) \text{ for all } k$$

and

$$f_n\left(k\right) = 0 \text{ for } k = 0,\ n-1,\ n.$$

Further, denote by k_n any value of k for which

$$f_n\left(k \pm 1\right) \leqslant f_n\left(k\right) \ .$$

That is, k should satisfy the inequalities A and B below.

$$\begin{gathered} \mathbf{A}\ :\ 0 \leqslant k(n-k)(n-k-1)^2 - (k+1)(n-k-1)(n-k-2)^2 \\ 4k^2 - \left(5n-8\right)k + \left(n-2\right)^2 \leqslant 0 \\ a_n \leqslant k \leqslant A_n \end{gathered} \tag{4.4}$$

where

$$a_n = \frac{5n - 8 - \sqrt{9n^2 - 16n}}{8} \quad , \quad A_n = \frac{5n - 8 + \sqrt{9n^2 - 16n}}{8}. \tag{4.5}$$

$$\begin{gathered} \mathbf{B}\ :\ 0 \leqslant k(n-k)(n-k-1)^2 - (k-1)(n-k+1)(n-k)^2 \\ 4k^2 - 5nk + n\left(n+1\right) \geqslant 0 \\ k \leqslant b_n \ \text{ or } k \geqslant B_n \end{gathered} \tag{4.6}$$

where

$$b_n = \frac{5n - \sqrt{9n^2 - 16n}}{8} \quad , \quad B_n = \frac{5n + \sqrt{9n^2 - 16n}}{8}. \tag{4.7}$$

That is, k should belong to the interval $[a_n, A_n]$ but not to the interval (b_n, B_n). Since $b_n = a_n{+}1$ and $B_n = A_n{+}1$, this means that $b_n{-}1 \leqslant k_n \leqslant b_n$.

If b_n is an integer both $b_n - 1$ and b_n are possible values for k_n. Otherwise there is a unique $k_n = \lfloor b_n \rfloor$.

If we rewrite b_n as

$$\begin{aligned} b_n &= \frac{5n - \sqrt{9n^2 - 16n}}{8} = \frac{n}{4} + \frac{3n}{8}\left(1 - \sqrt{1 - \frac{16}{9n}}\right) \\ &= \frac{n}{4} + \gamma_n. \end{aligned} \tag{4.8}$$

we see that $\gamma_2 = \frac{1}{2}$. Using a generalization of the binomial theorem, it can be shown that

$$\begin{aligned} \gamma_n &= \frac{3n}{8}\left(1 - \left(1 - 2\sum_{j=1}^{\infty} \frac{1}{j}\binom{2j-2}{j-1}\left(\frac{4}{9n}\right)^j\right)\right) \\ &= \frac{1}{3}\sum_{j=1}^{\infty} \frac{1}{j}\binom{2j-2}{j-1}\left(\frac{4}{9n}\right)^{j-1}. \end{aligned} \tag{4.9}$$

Hence, for $n > 2$ we have that $\frac{1}{3} < \gamma_n < \frac{1}{2}$, i.e. b_n is an integer if and only if $n = 2$. Thus, $k_n = \left\lfloor \frac{n}{4} + \gamma_n \right\rfloor$ is unique for $n > 2$ and

$$\frac{n}{4} + \frac{1}{3} < b_n < \frac{n}{4} + \frac{1}{2} \quad , n > 2. \tag{4.10}$$

Thus, since the cases $n = 1$ and $n = 2$ are trivial, we have $k_n = \left\lfloor \frac{n}{4} + \frac{1}{4} \right\rfloor = \left\lfloor \frac{n+1}{4} \right\rfloor$ for all n. ■

A rather lengthy proof of the formula for the maximal degree variance in a Bernoulli graph conditional on size $R = r$, that is $s^2_{\max}$ for uniformly random graphs of order n and size r, can be found in Snijders (1981b). An alternative proof similar to the previous proof is given here.

Let $G(r)$ be a graph of order n and size r consisting of

$$\begin{aligned} & k \text{ vertices of degree } 0 \\ & 1 \text{ vertex of degree } r_0 \\ & r_0 \text{ vertices of degree } n-k-1 \\ & n-k-1-r_0 \text{ vertices of degree } n-k-2 \\ \text{where } r_0 \ &= \ r - \binom{n-k-1}{2} \text{ and } k \text{ is given by} \end{aligned} \tag{4.11}$$

$$\binom{n-k-1}{2} < r \leqslant \binom{n-k}{2}. \tag{4.12}$$

This inequality implies that for $r = 1, 2, 3, ...$, the values of $n-k-1$ should be given as $1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, ...$, that is by $\left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor$ (Sloane & Plouffe (1995)). Thus

$$k = n - 1 - \left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor \text{ for } r > 0. \tag{4.13}$$

For such graphs we have the degree variance

$$\begin{aligned}
s^2(r) &= \frac{1}{n}\left[r_0^2 + r_0 (n-k-1)^2 + (n-k-1-r_0)(n-k-2)^2\right] - \left(\frac{2r}{n}\right)^2 \\
&= \frac{1}{4n}\left[(n-k)(n-k-1)(n-k-2)(n-k-3)\right] \\
&\quad + \frac{1}{n}\left[r\left(5(n-k-1) + r - (n-k)^2\right)\right] - \left(\frac{2r}{n}\right)^2 \\
&= \frac{(n-k)_{(4)} + 4r\left(5(n-k-1) + r - (n-k)^2\right)}{4n} - \left(\frac{2r}{n}\right)^2 \qquad (4.14)
\end{aligned}$$

where k is given by (4.13).

Theorem 6 *For graphs of order n and size r the maximal degree variance is given by the largest of the two values $s^2(r)$ and $s^2\left(\binom{n}{2} - r\right)$. Except for $n = 7$, $r = 11$*

$$\max\left\{s^2(r), s^2\left(\binom{n}{2} - r\right)\right\} = \begin{cases} s^2(r) & \text{if } r \geqslant \binom{n}{2}/2 \\ s^2\left(\binom{n}{2} - r\right) & \text{if } r < \binom{n}{2}/2. \end{cases}$$

For $n = 7$, $r = 11$ $s^2(10) > s^2(11)$.

Proof. The complement of $G(r)$ has $\binom{n}{2} - r$ edges and has the same degree variance. However, $s^2\left(\binom{n}{2} - r\right)$ might be larger or smaller than $s^2(r)$ and the largest of these two numbers is the maximal degree variance. To see this, consider the maximal degree variance without specifying the number of edges. According to Theorem 1 the optimal graph then has

$$r = \binom{n - \left\lfloor \frac{n+1}{4} \right\rfloor}{2}$$

edges. This r is larger than or equal to $\binom{n}{2}/2$ except when $n=3$ or $n=7$ since

$$\binom{n-k^*}{2}-\left(\binom{n}{2}-\binom{n-k^*}{2}\right)=\begin{cases}\frac{(n-1)(n+3)}{16} & \text{for} \quad n=1,5,9,... \\ \frac{n(n+8)-4}{16} & \text{for} \quad n=2,6,10,... \\ \frac{n(n-10)+5}{16} & \text{for} \quad n=3,7,11,... \\ \frac{n(n-4)}{16} & \text{for} \quad n=4,8,12,...\end{cases}$$

where $k^*=\left\lfloor\frac{n+1}{4}\right\rfloor$. These numbers are all non-negative except for $n=3$ and $n=7$. For $n=3$, $s^2(1)=s^2(2)$ but for $n=7$, $s^2(10)>s^2(11)$. ■

If $r^*=\max\left(r,\binom{n}{2}-r\right)$ then the optimal graph should be $G(r^*)$ and $s^2_{\max}=s^2(r^*)$ with the exception mentioned above.

The expressions for $s^2_{\max}$ in some particular cases are given by

$$s^2_{\max}=\frac{r(r+1)}{n}-\left(\frac{2r}{n}\right)^2 \quad \text{for } r\leqslant n-1 \tag{4.15}$$

and

$$\begin{aligned} s^2_{\max} &= \frac{n^2-5n+6}{n} \\ &= \frac{(n-2)(n-3)}{n} \quad \text{for } r=n>2. \end{aligned} \tag{4.16}$$

For $r=n>2$ the optimal graph is a star with one extra edge and it is possible to give an simple expression for its probability of occurance:

$$\begin{aligned} & P\left(S^2=s^2_{\max}\mid R=n\right) \\ &= \frac{n(n-1)(n-2)}{2\binom{\binom{n}{2}}{n}}. \end{aligned} \tag{4.17}$$

4.2 The minimum value

Theorem 7 *The minimum degree variance for a graph G of order n and size r is equal to*

$$s^2_{\min}=\theta(1-\theta)$$

$$\textit{where } \theta = \begin{cases} \frac{2r}{n} - \left\lfloor \frac{2r}{n} \right\rfloor & \textit{if} \quad r \leqslant \binom{n}{2}/2 \\ \frac{n(n-1)-2r}{n} - \left\lfloor \frac{n(n-1)-2r}{n} \right\rfloor & \textit{if} \quad r > \binom{n}{2}/2. \end{cases}$$

Proof. If $\frac{2r}{n} = m$ where m is an integer, then $\theta = 0$ and G should be regular with all degrees equal to m, i.e. an m-regular graph. Otherwise G should have $n(1-\theta)$ vertices of degree $\left\lfloor \frac{2r}{n} \right\rfloor = m$ and $n\theta$ vertices of degree $m+1$. Such graphs exist, since if n is even it is possible to construct m-regular graphs for $m = 0, 1, 2, ..., n-1$ and if n is odd it is possible to construct m-regular graphs for $m = 0, 2, 4, ..., n-1$. See, for example, Chartrand and Lesniak (1996). Hence,

$$\begin{aligned} s^2_{\min} &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \\ &= \frac{n(1-\theta)(m-\overline{x})^2 + n\theta(m+1-\overline{x})^2}{n} \end{aligned}$$

where $\overline{x} = \frac{2r}{n} = m + \theta$. It follows that

$$s^2_{\min} = (1-\theta)\theta^2 + \theta(1-\theta)^2 = \theta(1-\theta).$$

■

4.3 Other attained values

If we multiply s^2 by n^2 we get $z = n\sum x_i^2 - 4r^2$ and by ordering the distinct values of z, we get the integer sequence $z_1 < z_2 < \cdots < z_m$ where z_1 is the minimal and z_m is the maximal value.

Theorem 8 *For $4 \leqslant r \leqslant n-2$, z attains*

$$m = \begin{cases} 4 + \binom{r-1}{2} & \textit{if} \quad 4 \leqslant r \leqslant \frac{n}{2} \\ n + \frac{(r-2)(r-5)}{2} & \textit{if} \quad \frac{n}{2} \leqslant r \leqslant n-2 \end{cases} \tag{4.18}$$

distinct values $z_1 < z_2 < \cdots < z_m$ given according to

$$z_j = \begin{cases} 2r(n-2r) + 2n(j-1) & \textit{if} \quad 4 \leqslant r \leqslant \frac{n}{2} \\ 2(2r-n)(n-r) + 2n(j-1) & \textit{if} \quad \frac{n}{2} \leqslant r \leqslant n-2 \end{cases} \tag{4.19}$$

for $j = 1, ..., m-1$ *and*

$$\begin{aligned} z_m &= nr(r+1) - 4r^2 \\ &= z_{m-1} + 2n(r-3). \end{aligned} \qquad (4.20)$$

Proof. Consider $\sum x_i^2$ which attains the values

$$2r, 2(r+1), 2(r+2), ..., 2(r+m-2), r(r+1) \text{ if } 4 \leqslant r \leqslant \frac{n}{2}$$

for any sequence of graphs $G_1, ..., G_m$, satisfying the following requirements:

G_1 has $2r$ vertices of degree 1 and $n-2r$ vertices of degree 0,

G_2 has 1 vertex of degree 2, $2(r-1)$ vertices of degree 1 and $n-2r+1$ vertices of degree 0,

G_3 has 2 vertices of degree 2, $2(r-2)$ vertices of degree 1 and $n-2r+2$ vertices of degree 0,

...

G_{r+1} has r vertices of degree 2, and $n-r$ vertices of degree 0,

G_{r+2} has 1 vertex of degree 4, $r-5$ vertices of degree 2, 6 vertices of degree 1 and $n-r-2$ vertices of degree 0. If the graph is of size 4, G_{r+2} has 1 vertex of degree 3, 2 vertices of degree 2, 1 vertex of degree 1 and $n-4$ vertices of degree 0,

G_{r+3} has 1 vertex of degree 4, $r-4$ vertices of degree 2, 4 vertices of degree 1, and $n-r-1$ vertices of degree 0,

...

G_{m-3} has 1 vertex of degree $r-1$, $r+1$ vertices of degree 1 and $n-r-2$ vertices of degree 0,

G_{m-2} has 1 vertex of degree $r-1$, 1 vertex of degree 2, $r-1$ vertices of degree 1 and $n-r-1$ vertices of degree 0,

G_{m-1} has 1 vertex of degree $r-1$, 2 vertices of degree 2, $r-3$ vertices of degree 1 and $n-r$ vertices of degree 0,

G_m has 1 vertex of degree r, r vertices of degree 1 and $n-r-1$ vertices of degree 0.

Figure 2 illustrates such a sequence of graphs for $n=10$ and $r=5$.

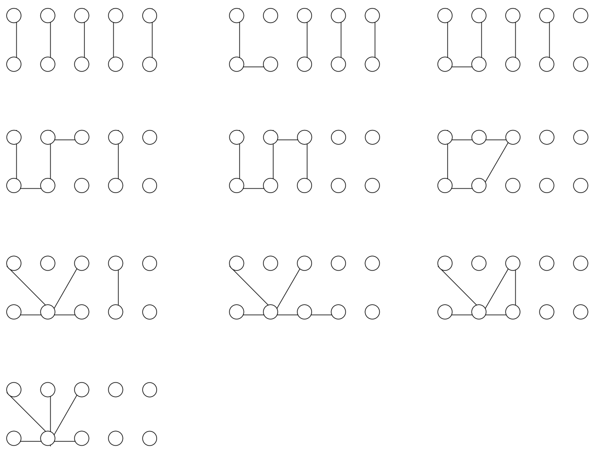

Figure 2. The graphs $G_1, ..., G_{10}$ for $n=10$ and $r=5$.

According to Theorem 7,

$$\begin{aligned} z_1 &= n^2 \left(\frac{2r}{n} - \left\lfloor \frac{2r}{n} \right\rfloor \right) \left(1 - \frac{2r}{n} + \left\lfloor \frac{2r}{n} \right\rfloor \right) \\ &= 2r\,(n-2r) + n^2 \left\lfloor \frac{2r}{n} \right\rfloor \left(\frac{4r-n}{n} - \left\lfloor \frac{2r}{n} \right\rfloor \right). \end{aligned} \tag{4.21}$$

Thus, for $r \leqslant n-2$ the minimum value of $\sum x_i^2$ is given by

$$\frac{z_1 + 4r^2}{n} = \begin{cases} 2r & \text{if} \quad 4 \leqslant r \leqslant \frac{n}{2} \\ \\ 2\,(3r - n) & \text{if} \quad \frac{n}{2} \leqslant r \leqslant n - 2 \end{cases}$$

For $4 \leqslant r \leqslant \frac{n}{2}$ the graphs G_1 and G_m have minimum and maximum values of $\sum x_i^2$. Every possible intermediate value is attained by the other graphs in the sequence. For $\frac{n}{2} \leqslant r \leqslant n-2$ some of the initial graphs in the sequence can not be constructed and the sequence starts with the graph G_{2r-n+1}. The same argument applies to this sequence. Theorem 8 follows. ■

For graphs of size $r \leqslant 3$ we have

$$z_1\,(n, r) = \begin{cases} 0 & \text{if} \quad r = 0 \\ 2\,(n-2) & \text{if} \quad r = 1 \\ 4\,(n-4) & \text{if} \quad r = 2 \\ 6\,(n-6) & \text{if} \quad r = 3 \end{cases}$$

$$m = \begin{cases} 1 & \text{if} \quad r = 0 \\ 1 & \text{if} \quad r = 1 \\ 2 & \text{if} \quad r = 2 \\ 4 & \text{if} \quad r = 3 \end{cases}$$

and the difference between any two successive values is $2n$. Thus, Theorem 8 is valid for $r \leqslant n-2$. Theorem 8 is in general not valid for $r \geqslant n-1$. For $r \geqslant n-1$ there are some exceptions to the recurrence relation

$$z_j = 2n + z_{j-1} \text{ for } j = 2, ..., m-1 \tag{4.22}$$

where z_1 is given by (4.21) and for some j in the right tail $z_j - z_{j-1} \geqslant 2n$. It is not known for all n and all r where in the right tail $z_j - z_{j-1} > 2n$. All possible degree variances for graphs of order 7 are given in Table 1.

r	$z = n^2s^2$
0	0
1	10
2	12,26
3	6,20,34,48
4	6,20,34,48,62,76
5	12,26,40,54,68,82,110
6	10,24,38,52,66,80,94,108,150
7	0,14,28,42,56,70,84,98,112,140
8	10,24,38,52,66,80,94,108,122,136
9	12,26,40,54,68,82,110,124,138
10	6,20,34,48,62,76,90,104,118,132,146,160

Table 1. The possible values of $z = n^2s^2$ for graphs of order $n = 7$ and size $r = 0, ..., 10$.

Turning to graphs with no restriction on size, the recurrence relation is given by

$$z_j = 2n + z_{j-\lambda} \text{ for } j = 2, ..., m-1 \tag{4.23}$$

where λ is a integer valued lag length of the recurrence relation and $z_j - z_{j-\lambda} \geqslant 2n$ in the tails. It is not known for all n where in the right tail $z_j - z_{j-\lambda} > 2n$, but all values in the left tail can be derived by starting from (4.21). In particular, the first three values of z_j are

$$\begin{aligned} z_1 &= 0 \ , \ z_2 = 2(n-2) \ , \ z_3 = 2n \text{ if } n \text{ is even, and} \\ z_1 &= 0 \ , \ z_2 = n-1 \ , \ z_3 = 2(n-2) \text{ if } n \text{ is odd .} \end{aligned}$$

The parameter λ i.e. the lag length, tells us in how many subsequences $z_1, ..., z_m$ can be separated so that the difference between any two succesive terms, except for the right tail, is $2n$ in every subsequence. From Table 1 where $n = 7$, we can see that $\lambda = 4$,

$$z_j = 14 + z_{j-4} \ , \ z_1 = 0 \ , \ z_2 = 6 \ , \ z_3 = 10 \ , \ z_4 = 12$$

and in this case λ is equal to the size of the set of initial values. In general, λ and the initial values are obtained in the following way: Calculate the minimum value of z for every $r \leqslant \binom{n}{2}/2$ and order the distinct minimum values $a_1 < a_2 < \cdots < a_m$. Denote the set $\{a_1, ..., a_m\}$ by A. Remove those a_j that are equal to $a_i + 2kn = a_j$ for some i and some k. The remaining set is denoted Λ. The number of elements in Λ equals the lag length λ of the recurrence relation (4.23). Further, $\Lambda \subseteq I$, the set of initial values, and I consists of all possible z-values within the range of Λ. It should be noted that I might contain elements not in A. The reason is that the minimum value z_1 for fixed r might be larger than z_2 for another r. For example, for $n = 15$ we have

$$\begin{aligned}
A &= \{0, 14, 26, 36, 44, 50, 54, 56\}, \\
\Lambda &= \{0, 14, 26, 36, 50, 54\}, \\
I &= \{0, 14, 26, 30, 36, 44, 50, 54\}, \\
z_j &= 30 + z_{j-6} \text{ for } j = 9, 10, 11, \ldots \text{ so that} \\
z_9 &= 56, \ z_{10} = 60, \ z_{11} = 66, \ldots
\end{aligned}$$

The lag length of the recurrence relation (4.23) and the initial values for $n = 7, ..., 20$ are listed in Table 2. For $n = 4, 12, 20, 28, ...$ it is possible to replace (4.23) by an alternative recursion with lag $\lambda/2$ so that fewer initial values are needed. This recursion is $z_j = n + z_{j-\lambda/2}$.

n	λ	The initial values of $z = n^2 s^2$
7	4	0,6,10,12
8	2	0,12,16
9	4	0,8,14,18
10	3	0,16,20,24
11	6	0,10,18,22,24,28,30
12	4	0,20,24,32,36
13	7	0,12,22,26,30,36,38,40,42,48
14	4	0,24,28,40,48
15	6	0,14,26,30,36,44,50,54
16	3	0,28,32,48
17	9	0,16,30,34,42,50,52,60,66,68,70,72
18	4	0,32,36,56,72,80
19	10	0,18,34,38,48,56,60,70,72,76,78,84,86,88,90
20	6	0,36,40,64,76,80,84,96

Table 2: The lag length and initial values of $z_j = 2n + z_{j-\lambda}$.

All possible values of $z = n^2 s^2$ for graphs of order $n = 3, ..., 10$ are listed in Appendix C.

5 The distribution of the degree variance

5.1 The exact distribution

The distribution of the degree variance S^2 is related to the distribution of the ordered degree sequence, but is even more complicated to determine. Without restriction the vertices can be labeled so that $(x_1 \leqslant x_2 \leqslant, ..., \leqslant x_n)$. Any given ordered degree sequence and its ordered complement, $(n-1-x_n, ..., n-1-x_1)$ have the same degree variance, s^2, but that value is not necessarily unique among different degree sequences. No effective method of degree sequence enumeration is known, and the number of graphs and the number of distinct values of the degree variances increase rapidly with n, as indicated by Table 3. For methods of graphical enumeration see, for example, Harary (1969) , Deo (1973) or Sloane & Plouffe (1995).

n	Number of unlabeled graphs	Number of distinct S^2values
3	4	2
4	11	4
5	34	11
6	156	14
7	1 044	43
8	12 346	34
9	274 668	102
10	12 005 168	110

Table 3: The number of distinct S^2 values and the number of unlabeled graphs of order n for $3 \leqslant n \leqslant 10$.

Hence, it is very time consuming to find the exact distribution of S^2 if $n \geqslant 7$. Even the task of determining the possible values of S^2 is cumbersome for modest values of n, as shown in the previous section. If we let $g_n(t) = Et^{Z_n}$ be the probability generating function of Z_n and $M_n(t) = Ee^{tZ_n}$ be the corresponding moment generating function, then for $n = 3$ and $n = 4$ we have

$$g_3(t) = \left[3pqt^2 + (1-3pq)\right] \ , \ M_3(t) = \left[3pqe^{2t} + (1-3pq)\right], \qquad (5.1)$$

$$g_4(t) = \left[2pqt^4 + (1-2pq)\right]^3 \ , \ M_4(t) = \left[2pqe^{4t} + (1-2pq)\right]^3 . \qquad (5.2)$$

Due to the structure of the recurrence relation (4.23), the irregularities in the tails and the rapidly increasing number of graphs, it is much harder or in practice impossible, to derive the corresponding functions or the distribution function for Bernoulli graphs of high order. Thus, there is a need for approximate methods.

The exact distributions of the degree variance for Bernoulli graphs and conditional Bernoulli graphs of order n for $3 \leqslant n \leqslant 6$ are given in Appendix *B*.

5.2 Gamma approximations

A random variable Y has a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$,denoted by $Y \in Gamma\,(\alpha, \beta)$, if its density function is given by

$$g_{\alpha,\beta}(y) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}} y^{\alpha-1} e^{-y/\beta} \ , \ 0 \leqslant y < \infty.$$

For $Y \in Gamma\,(\alpha, \beta)$ it holds that

$$E(Y^k) = \frac{\beta^k}{\Gamma(\alpha)} \Gamma(\alpha + k) = \beta^k \prod_{j=0}^{k-1} (\alpha + j) \ , \ k \geqslant 0. \qquad (5.3)$$

In particular the first two moments yield

$$E(Y) = \alpha\beta \ \text{ and } Var(Y) = \alpha\beta^2. \qquad (5.4)$$

Let $U_i, i = 1, 2, ..., n$ be a sequence of n *independent identically distributed* normal random variables with mean μ and variance σ^2, that is, $U_1, ..., U_n$ are *iid* $N(\mu, \sigma^2)$. Let $W = \frac{1}{n}\sum_{i=1}^{n}\left(U_i - \overline{U}\right)^2$where $\overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$. Then, according to known results (Johnson & Kotz 1970), W is gamma distributed i.e.

$$W \in Gamma\left(\frac{n-1}{2}, \frac{2\sigma^2}{n}\right) \qquad (5.5)$$

and according to (5.3)

$$E(W) = \frac{n-1}{n}\sigma^2 \text{ and } Var(W) = 2\left(\frac{n-1}{n^2}\sigma^4\right). \tag{5.6}$$

The degrees of the vertices in a Bernoulli graph are binomially distributed with $\mu = (n-1)\,p$ and $\sigma^2 = (n-1)\,pq$; and binomially distributed random variables are approximately normally distributed if their variances are sufficiently large. Thus, neglecting the weak pairewise dependence between the vertex degrees, we can argue that S^2 is approximately

$$Gamma\left(\frac{n-1}{2}, \frac{2\,(n-1)\,pq}{n}\right). \tag{5.7}$$

However, due to the dependence between the vertex degrees, this gamma distribution does not have the correct mean and variance. A gamma distribution with the correct mean and variance can be obtained by choosing the gamma distribution parameters α and β so that $\alpha\beta = E(S^2)$ and $\alpha\beta^2 = Var(S^2)$, where $E(S^2)$ and $Var(S^2)$ are given by (3.10) and (3.11). This leads to

$$\alpha = \frac{n(n-1)}{2\,[1+(n-6)pq]}pq \text{ and } \beta = \frac{2(n-2)}{n^2}\,[1+(n-6)pq]\,. \tag{5.8}$$

The α-parameters given by (5.7) and (5.8) are equal when $pq = 1/6$.

Table 4 shows the first three moments of S^2 derived under independence assumptions (unadjusted gamma) and adjusted for the dependence (adjusted gamma).

Moment	Unadjusted gamma	Adjusted gamma
1	$\frac{(n-1)^2}{n}pq$	$\frac{(n-1)(n-2)}{n}pq$
2	$\frac{(n+1)(n-1)^3}{n^2}p^2q^2$	$\frac{(n-1)(n-2)^2pq((n+4)(n-3)pq+2)}{n^3}$
3	$\frac{(n+3)(n+1)(n-1)^4}{n^3}p^3q^3$	$\frac{8(n-1)(n-2)^3}{n^5}pq + \frac{2(n-1)(n-2)^3\left(3n^2+5n-48\right)}{n^5}p^2q^2 + \frac{(n-1)(n-2)^3(n-3)(n+4)\left(n^2+3(n-8)\right)}{n^5}p^3q^3$

Table 4. The first three approximate moments of S^2.

The exact first two moments of S^2 equal the adjusted gamma moments, and the exact third moment obtained from (3.3), is equal to

$$\begin{aligned} E\left(\left(S^2\right)^3\right) &= \frac{4\left(n-1\right)\left(n-2\right)^3}{n^5}pq \\ &+\frac{2\left(n-1\right)_{(3)}\left(3n-4\right)\left[\left(n-2\right)\left(n+6\right)-8\right]}{n^5}p^2q^2 \\ &+\frac{\left(n-1\right)_{(3)}\left[n^4\left(n+3\right)-4\left(3n-4\right)\left[3\left(n-2\right)\left(n+6\right)-\left(n+4\right)\right]\right]}{n^5}p^3q^3. \end{aligned} \tag{5.9}$$

Let D_n denote the difference between the exact third moment of S^2 and the third adjusted gamma moment. We have that

$$\begin{aligned} D_n &= -\frac{4\left(n-1\right)\left(n-2\right)^3}{n^5}pq+\frac{4\left(n-1\right)\left(n-2\right)\left(3n^3-22n^2+48n-24\right)}{n^5}p^2q^2 \\ &-\frac{64\left(n-1\right)^2\left(n-2\right)\left(n-3\right)\left(n-4\right)}{n^5}p^3q^3. \end{aligned}$$

For fixed p, $|D_n|$ increase with n and the stationary points of D_n tends to $p=\frac{1}{2}$ and $p=\frac{1}{2}\pm\frac{1}{4}\sqrt{2}$. Thus

$$\lim_{n\to\infty} D_n = 4p^2q^2\left(3-16pq\right)\ ,\ \max D_n = \frac{1}{16} \text{ and } \min D_n = -\frac{1}{4}. \tag{5.10}$$

The corresponding differences between the exact moments and the moments of the unadjusted gamma distributed variable W are tending to the following limits:

$$\begin{aligned} \left[E\left(S^2\right)-E\left(W\right)\right] &\to -pq\ , \\ \frac{E\left(\left(S^2\right)^2\right)-E\left(W^2\right)}{n} &\to -2p^2q^2 \text{ and} \\ \frac{E\left(\left(S^2\right)^3\right)-E\left(W^3\right)}{n^2} &\to -3p^3q^3. \end{aligned} \tag{5.11}$$

Thus, the unadjusted gamma approximation gives a bias to the mean and increasing biases to higher moments. The adjusted gamma approximation with correct first two moments has a bias for the third moment which is bounded by $-1/4$ and $1/16$.

Since the distribution of S^2 is discrete and the gamma distribution is continuous, we can improve the approximation by the use of a *continuity correction.* Let $Z = n^2 S^2$, denote the ordered distinct values of Z by $z_1 < z_2 < \cdots < z_m$ and let $s_j^2 = z_j/n^2$ for $j = 1, ..., m$. We have

$$\begin{aligned}
P(S^2 &= s_j^2) = P\left(z_j - \frac{z_j - z_{j-1}}{2} < Z < z_j + \frac{z_{j+1} - z_j}{2}\right) \\
&= P\left(z_j - a < Z < z_j + b\right) = P\left(s_j^2 - \frac{a}{n^2} < S^2 < s_j^2 + \frac{b}{n^2}\right) \\
&\approx G_{\alpha,\beta}\left(s_j^2 + \frac{b}{n^2}\right) - G_{\alpha,\beta}\left(s_j^2 - \frac{a}{n^2}\right) \\
&\approx \frac{a+b}{n^2} g_{\alpha,\beta}\left(s_j^2\right) = \frac{s_{j+1}^2 - s_{j-1}^2}{2} g_{\alpha,\beta}\left(s_j^2\right), \qquad (5.12)
\end{aligned}$$

$$P(S^2 \leqslant s_j^2) = P(S^2 \leqslant s_j^2 + \frac{b}{n^2}) \approx G_{\alpha,\beta}\left(s_j^2 + \frac{b}{n^2}\right) \qquad (5.13)$$

and

$$P(S^2 \geqslant s_j^2) = P(S^2 \geqslant s_j^2 - \frac{a}{n^2}) \approx 1 - G_{\alpha,\beta}\left(s_j^2 - \frac{a}{n^2}\right). \qquad (5.14)$$

where $a = \frac{z_j - z_{j-1}}{2}$ and $b = \frac{z_{j+1} - z_j}{2}$.

The values $z_4, z_5, ..., z_{m-1}$ are not all known for $n > 10$. However by writing

$$z = 2n\left[r + \sum_{i=1}^{n} \binom{x_i}{2}\right] - 4r^2, \qquad (5.15)$$

where r is the number of edges, we see that z is even and z is also divisible by 4 if n is even.

Theorem 9 *For $n > 2$ it holds that*

$$\begin{aligned}
\min_{1 \leqslant j < m} (z_{j+1} - z_j) &= 2 \text{ if } n \text{ is odd,} \\
\min_{1 \leqslant j < m} (z_{j+1} - z_j) &= 4 \text{ if } n \text{ is even.}
\end{aligned}$$

Proof. Let G be a graph of order n, size r and let $z_{j+1} = nc - 4r^2$ where $c = \sum x_i^2$. Add one edge to G and let
$z_j = n(c + \Delta) - 4(r+1)^2$. We have that

$$z_{j+1} - z_j = \begin{cases} 2 & \text{if} \quad r = \frac{n\Delta - 2}{8} \\ 4 & \text{if} \quad r = \frac{n\Delta}{8}. \end{cases}$$

i) For $n = 3, 7, 11, ...,$ and $r = \frac{n\Delta-2}{8}$: Let the new edge connect two vertices previously of degree one, i.e. $\Delta = 6$ and $z_{j+1} - z_j = 2$.
ii) For $n = 5, 9, 13, ...,$and $r = \frac{n\Delta-2}{8}$: Let the new edge connect two vertices previously of degree zero, i.e. $\Delta = 2$ and $z_{j+1} - z_j = 2$.
iii) For $n = 4, 6, 8, ...,$and $r = \frac{n\Delta}{8}$: Let the new edge connect a vertex previously of degree zero to a vertex previously of degree one, i.e. $\Delta = 4$ and $z_{j+1} - z_j = 4$. ∎

That is, if we observe a value s_j^2 and don't know the values of s_{j-1}^2 and s_{j+1}^2 we can use

$$a = b = \begin{cases} 1 & \text{if } n \text{ is odd} \\ 2 & \text{if } n \text{ is even} \end{cases} \tag{5.17}$$

in (5.12) - (5.14).

Since hypergeometric random variables can be approximated by the binomial distribution it follows that gamma approximation is valid for S^2 in Bernoulli graphs conditional on the size. Thus, S^2 in the R-conditional Bernoulli graph is approximately $Gamma\,(\alpha\ ,\ \beta)$ where

$$\alpha = \frac{r(n+2)\left[n(n-1)-4\right]\left[n(n-1)-2r\right]}{2n^2(r-1)\left[n(n-1)-2(r+1)\right]} \tag{5.18}$$

and

$$\beta = \frac{4(r-1)\left[n(n-1)-2(r+1)\right]}{(n+2)(n+1)\left[n(n-1)-4\right]}. \tag{5.19}$$

As above for unconditional Bernoulli graphs, we can improve the approximation by the use of a *continuity correction.* For the degree variance S^2 in R-conditional Bernoulli graphs we use

$$a = b = n. \tag{5.20}$$

in (5.12) - (5.14).

5.3 Simulation results

In the figures of this section $F(s^2)$ means the simulated distribution function of S^2 and $G(s^2)$ means the gamma distribution function in the adjusted approximation. Results based on the exact distribution of S^2 for $n = 6$ and results based on the approximate distribution of S^2 obtained from 10^7 simulated graphs for each $n = 7, 8, ..., 15, 20, 30$ and 10^6 simulated graphs for $n = 100$ show that the adjusted gamma approximation to the distribution function of S^2 in Bernoulli graphs works well, especially when $P(S^2 \geqslant s^2) \leqslant 0.10$. For $n = 8, ..., 12$ the adjusted approximation is very good when p is close to 0.2. For graphs of higher order the approximation is better when the variance is higher i.e. when p tends to 0.5. When p is fixed the accuracy of the approximation for graphs of order n is better than the accuracy for graphs of order $n+1$ if n is even. The latter is due to the relative smoothness of the distribution when n is even and it is reflected by a lower lag length of the recurrence relation (4.23). Figure 3 and 4 show the difference in smoothness of the distribution for $n = 7$ and $n = 10$. Table 5 shows the differences between the exact distribution of S^2 and the adjusted and unadjusted gamma approximations respectively for $n = 6$, $p = 0.1$ and $p = 0.5$. From Table 6 it can be seen that the unadjusted gamma approximation is bad, which agree with (5.11). Details of the differences between the simulated and the approximated distribution function are shown in Figure 5 and 6.

Further, to investigate the accuracy of the adjusted gamma approximation to the distribution function of S^2 conditional on R, 10^6 graphs were simulated for $n = 7, ..., 12, 15$ and various values of r. The results in Table 7 show that the gamma approximation to the distribution function of S^2 in uniform random graphs is better than the corresponding approximation in Bernoulli graphs. This is explained by the smoothness of the distribution in uniform random graphs, which agree with the recurrence relation (4.22) and can also be seen from Figure 7.

$n = 6$, $p = 0.1$						
z	$P(Z = z)$	Adj. diff.	Unadj. diff.	$P(Z \leqslant z)$	Adj. diff.	Unadj. diff.
0	.210155	-.036779	.125352	.210155	-.036779	.125352
8	.467669	.133990	.145285	.677824	.097211	.270637
12	.051256	-.147356	-.228063	.729080	-.050145	.042573
20	.171050	.063438	.005814	.900130	.013293	.048388
24	.051431	-.004643	-.031431	.951561	.008651	.016957
32	.015618	-.012975	-.022267	.967179	-.004324	-.005311
36	.023013	.008633	.006686	.990192	.004308	.001376
44	.007374	.000210	.000620	.997566	.004519	.001996
48	.000314	-.003230	-.002396	.997880	.001289	-.000400
56	.001913	.000170	.000850	.999793	.001459	.000450
60	.000140	-.000715	-.000270	.999933	.000744	.000180
68	.000017	-.000400	-.000138	.999950	.000344	.000042
72	.000029	-.000174	-.000029	.999980	.000170	.000013
80	.000021	-.000078	-.000001	1.000000	.000093	.000013
$n = 6$, $p = 0.5$						
z	$P(Z = z)$	Adj. diff.	Unadj. diff.	$P(Z \leqslant z)$	Adj. diff.	Unadj. diff.
0	.005249	.002218	-.003899	.005249	.002218	-.003899
8	.097961	.047829	.038574	.103211	.050047	.034675
12	.076904	-.048359	-.024155	.180115	.001688	.010520
20	.202148	.033779	.082007	.382263	.035467	.092526
24	.104370	-.065564	-.017482	.486633	-.030098	.075045
32	.181274	.035667	.068299	.667908	.005570	.143344
36	.070801	-.041438	-.028068	.738709	-.035868	.115276
44	.106201	.025849	.023094	.844910	-.010020	.138370
48	.049439	-.005033	-.018376	.894348	-.015052	.119994
56	.056763	.021352	.002674	.951111	.006300	.122668
60	.021973	-.000295	-.020396	.973084	.006005	.102273
68	.021973	.008342	-.010732	.995056	.014348	.091541
72	.004578	-.003582	-.020363	.999634	.010766	.071178
80	.000366	-.004427	-.018460	1.000000	.006338	.052718

Table 5. The differences between the.exact distribution of $Z = n^2S^2$ and the adjusted and unadjusted gamma approximations respectively for $n = 6$, $p = 0.1$ and $p = 0.5$.

Table 6 below shows the Kolmogorov distance, i.e.the greatest absolute deviation between the simulated distribution function of S^2 and the adjusted and the unadjusted gamma approximation respectively for various values of n and p. The corresponding distance in the upper 10% tail is given within parenthesis. Table 7 on page 35 shows the corresponding values for the adjusted gamma approximation to S^2 conditional on R.

n		$p = 0.1$	$p = 0.2$	$p = 0.4$	$p = 0.5$
7	Adj.	.1132 (.0252)	.0609 (.0112)	.0455 (.0077)	.0474 (.0088)
	Unadj.	.2854 (.0547)	.1586 (.0629)	.1319 (.1234)	.1334 (.1064)
8	Adj.	.0348 (.0063)	.0121 (.0017)	.0197 (.0047)	.0202 (.0048)
	Unadj.	.1557 (.0104)	.1088 (.0661)	.1118 (.0973)	.1100 (.0946)
9	Adj.	.0873 (.0162)	.0193 (.0053)	.0241 (.0042)	.0236 (.0051)
	Unadj.	.1617 (.0332)	.1157 (.0573)	.1053 (.0867)	.1076 (.0936)
10	Adj.	.0370 (.0019)	.0057 (.0014)	.0114 (.0028)	.0121 (.0033)
	Unadj.	.1387 (.0113)	.0981 (.0552)	.0939 (.0781)	.0944 (.0801)
12	Adj.	.0245 (.0054)	.0081 (.0012)	.0101 (.0025)	.0106 (.0028)
	Unadj.	.1366 (.0164)	.0949 (.0508)	.0883 (.0698)	.0883 (.0734)
15	Adj.	.0257 (.0087)	.0135 (.0034)	.0108(.0028)	.0107 (.0027)
	Unadj.	.1205 (.0213)	.0888 (.0466)	.0822 (.0595)	.0816 (.0614)
20	Adj.	.0136 (.0029)	.0056 (.0013)	.0046 (.0014)	.0049 (.0015)
	Unadj.	.0417 (.0727)	.0715 (.0352)	.0668 (.0478)	.0668 (.0486)
30	Adj.	.0111 (.0018)	.0045 (.0013)	.0028 (.0012)	.0039 (.0017)
	Unadj.	.0716 (.0110)	.0548 (.0281)	.0546 (.0357)	.0544 (.0367)
100	Adj.	.0020 (.0011)	.0033 (.0015)	.0011 (.0006)	.0009 (.0004)
	Unadj.	.0301 (.0075)	.0287 (.0131)	.0287 (.0154)	.0289 (.0162)

Table 6: The Kolmogorov distances for S^2. The distances in the 10% quantiles are given within parenthesis.

Figure 3 and 4 on page 32 show the distribution of Z i.e. n^2S^2 for $p = 0.5$ and $n = 7$ and 10 respectively. Note that the different lag lengths of the recurrence relation (4.23) are reflected in the figures. For $n = 7$ the lag length is 4 and the lag length is 3 for $n = 10$.

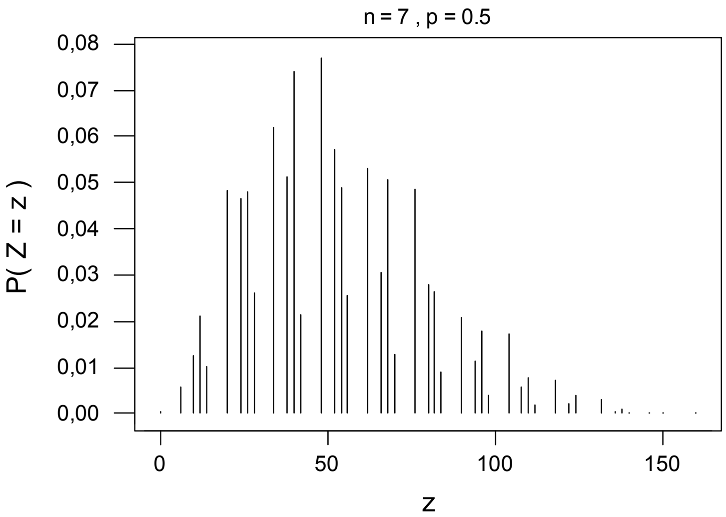

Figure 3 The simulated distribution of Z for $n = 7$ and $p = 0.5$.

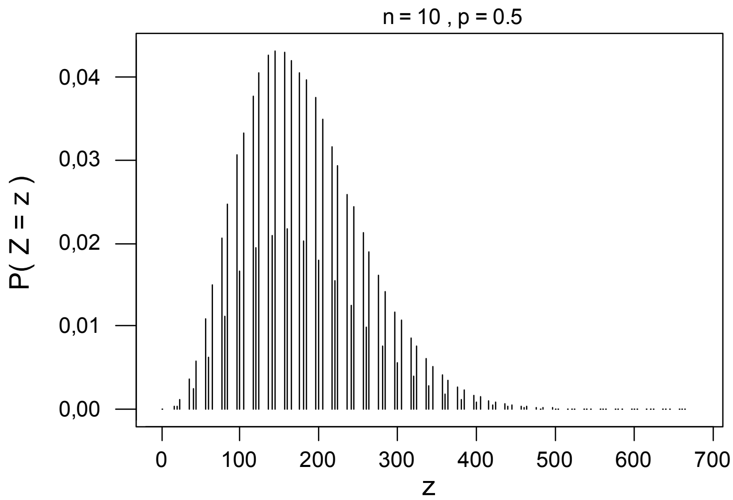

Figure 4 The simulated distribution of Z for $n = 10$ and $p = 0.5$.

Figure 5 below shows the differences $F(s^2) - G(s^2)$ plotted against $F(s^2)$ for $n = 15$. The dependence structure of the differences varies for different values of n and p and can hardly be modeled without knowledge of the true distribution. In Figure 6 the differences between the simulated distribution function of S^2 and the adjusted gamma approximation are plotted against z for $n = 15$ and $p = 0.5$. From Figure 6 we get an inkling of the structure of the differences.

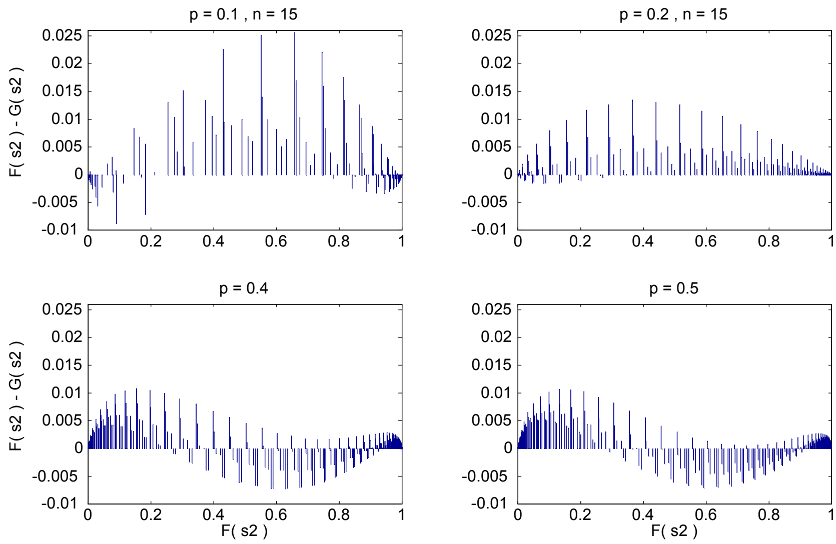

Figure 5. $F\left(s^2\right) - G(s^2)$ plotted against $F\left(s^2\right)$ for $n = 15$.

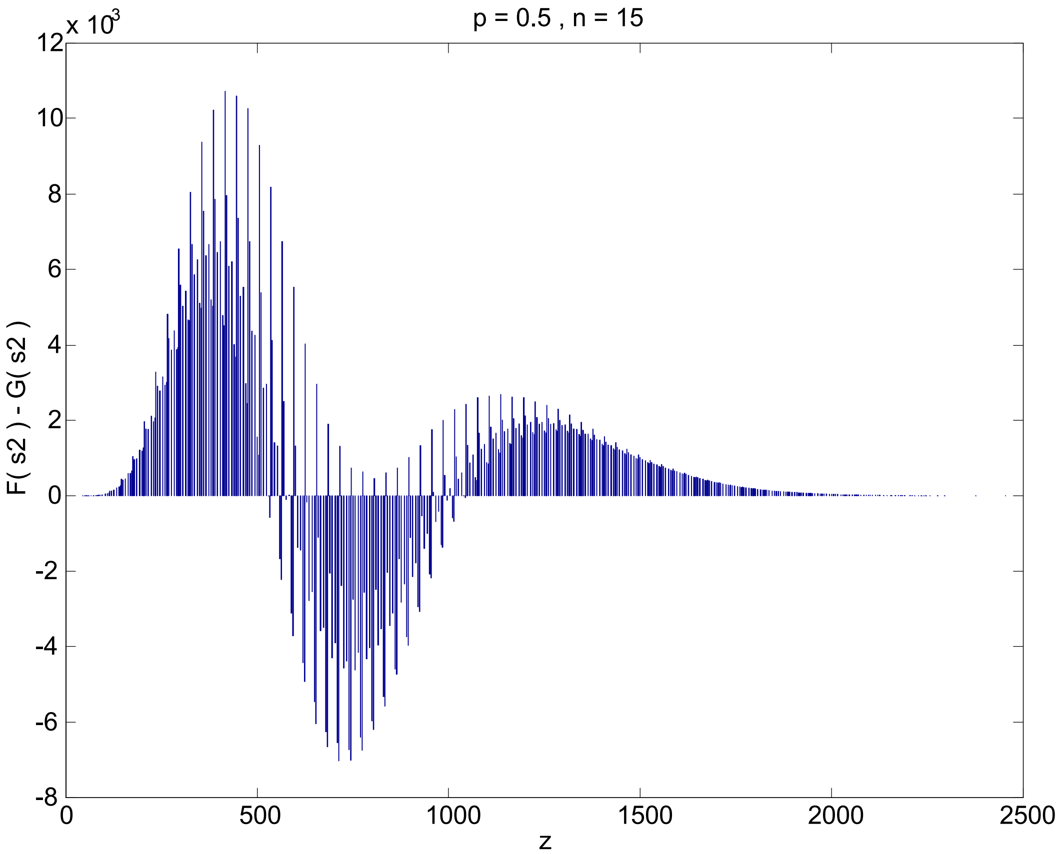

Figure 6. The differences between the simulated distribution function of S^2 and the adjusted gamma approximation plotted against z for $n = 15$ and $p = 0.5$.

n					
7		$r = 5$	$r = 7$	$r = 9$	$r = 10$
		.0301 (.0074)	.0421 (.0071)	.0355 (.0083)	.0282 (.0084)
10		$r = 5$	$r = 10$	$r = 17$	$r = 22$
		.0103 (.0103)	.0125 (.0030)	.0116 (.0036)	.0117 (.0035)
12		$r = 5$	$r = 10$	$r = 15$	$r = 30$
		.0118 (.0064)	.0052 (.0038)	.0077 (.0028)	.0111 (.0034)
15		$r = 5$	$r = 10$	$r = 20$	$r = 50$
		.0201 (.0075)	.0041 (.0029)	.0034 (.0028)	.0062 (.0024)

Table 7. The Kolmogorov distances for S^2conditional on $R = r$. The distances in the upper 10% tails are given within parenthesis.

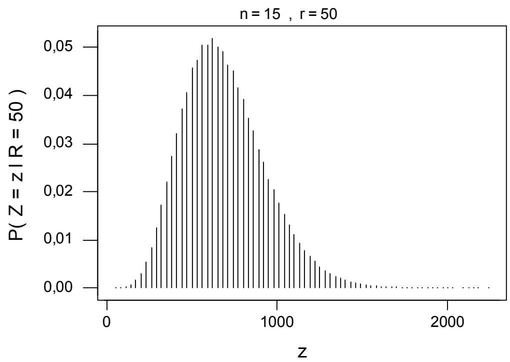

Figure 7. The simulated distribution of Z conditional on $R = 50$.

6 Application to Padgett's Florentine families

One part of the network data compiled by Padgett, (Padgett & Ansell (1993)) consist of marriage relations among 16 families in the 15th century Florence, Italy. In Figure 8 we have drawn an edge between a pair of vertices i.e. a pair of families if a member of one family marries a member of the other. A more detailed description of the network can be found in Wasserman & Faust (1994).

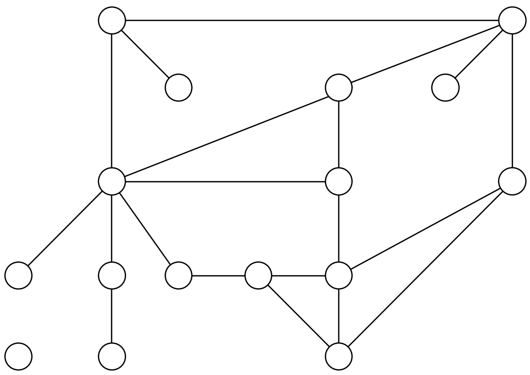

Figure 8. Marital relations between Padgett's Florentine families.

The statistics of the network are:

$$n = 16 \ , \ r = 20 \ , \ \widehat{p} = \frac{1}{6} \text{ and } s^2 = 2.125.$$

If we assume that the edges are generated according to a Bernoulli model, S^2 is approximately $Gamma(6.977 \ , \ 0.261)$ and $P\,(S^2 \geqslant 2.125) = 0.2588$. Conditional on $R = 20$, S^2 is approximately $Gamma(8.827 \ , \ 0.208)$ and $P\,(S^2 \geqslant 2.125) = 0.2443$. Thus, there is no strong evidence against the hypothesis that there is no centrality. Wasserman and Faust (1994) come to the same conclusion in some of their investigations of these data. However, they also demonstrated some other findings obtained with other models.

References

[1] Buckley,F.(1997). *Central Embeddings and Iterated Centrality.* Graph Theory Notes of New York XXXlll, 40-43 (1997), N.Y. Academy Sciences.

[2] Chartrand,C. & Lesniak, L. (1996). *Graphs & Digraphs, 3rd ed.* London: Chapman & Hall.

[3] Deo, N. (1974). *Graph Theory With Applications to Engineering and Computer Science*. New Jersey: Prentice-Hall.

[4] Frank, O. (1979). *Moment Properties of Subgraph Counts in Stochastic Graphs.* Ann. N.Y. Academy Sciences, 319: 207-218.

[5] Frank, O. (2000) *Using centrality modeling in network surveys.* Stockholm University, Department of Statistics.

[6] Freeman, L. (1978). *Centrality in Social Networks: Conceptual Clarification.* Social Networks, 1: 215-239.

[7] Harary, F. (1969). *Graph Theory.* Reading: Perseus Books.

[8] Johnson, N. & Kotz, S. (1970). *Continuous Univariate Distributions-1.* Boston: Houghton Mifflin.

[9] Johnson, N., Kotz, S. & Balakrishnan, N. (1997). *Discrete Multivariate Distributions.* New York: Wiley.

[10] Johnson, N., Kotz, S. & Kemp, A. (1992). *Univariate Discrete Distributions, 2nd ed.* New York: Wiley.

[11] Padgett, J.F. & Ansell, C.K. (1993). *Robust action and the rise of the Medici, 1400-1434.* American Journal of Sociology. 98, 1259 -1319.

[12] Sloane, N. & Plouffe, S. (1995). *The Encyclopedia of Integer Sequences.* San Diego: Academic Press.

[13] Snijders, T. (1981a). *The Degree Variance: An Index of Graph Heterogeneity.* Social Networks, 3: 163 -174

[14] Snijders, T. (1981b). *Maximum Value and Null Moments of the Degree Variance.* TW-report 229. Department of Mathematics, University of Groningen.

[15] Tallberg, C. (2000). *Centrality and Random Graphs.* Research Report 2000:7, Stockholm University, Department of Statistics.

[16] Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* New York Cambridge University Press.

A Product moments and moments of the degree variance

If $U \in Bin(n-t,p)$, then is $E\left(U^k\right) = \sum_{j=1}^{k} \frac{S(k,j)(n-t)!}{(n-t-j)!}p^j$, where $S(k,j) = \frac{1}{j!}\sum_{i=0}^{j}(-1)^i\binom{j}{i}(j-i)^k$ are the Stirling numbers of the second kind (Johnson, Kotz & Kemp, 1992). The product moment $EX_1^{w_1}X_2^{w_2}\cdots X_t^{w_t}$ is denoted by $A_{w_1w_2\cdots w_t}$. Using conditional expectation we can, for instance, obtain A_{22} in the following way. Let $X_{1,2}=1$ if there is an edge connecting the vertices 1 and 2, and let $X_{1,2}=0$ otherwise.

$$\begin{aligned}
EX_1^2X_2^2 &= A_{22} = pE\left(X_1^2|X_{1,2}=1\right)E\left(X_2^2|X_{1,2}=1\right) \\
&\quad +qE\left(X_1^2|X_{1,2}=0\right)E\left(X_2^2|X_{1,2}=0\right) \\
&= p\left(E(U+1)^2\right)^2 + q\left(EU^2\right)^2, \ \ (U \in Bin(n-2,p)) \\
&= \left[(n-1)p + (n-1)(n-2)p^2\right]^2 + \left[2(n-2)p+1\right]^2 pq
\end{aligned}$$

In general, for $X_i \in Bin(n-1,p)$, $i=1,...,n$ and $U_i \in Bin(n-t,p)$, $i=1,2,...,t$, $t=1,...,n$ we have

$A_{w_1w_2\cdots w_t} = \sum_{i=1}^{2^{\binom{t}{2}}} a_i$, where

$$\begin{aligned}
a_1 &= q^{\binom{t}{2}}EU_1^{w_1}EU_2^{w_2}\cdots EU_{t-1}^{w_{t-1}}EU_t^{w_t} \\
a_2 &= pq^{\binom{t}{2}-1}EU_1^{w_1}EU_2^{w_2}\cdots E\left(U_{t-1}+1\right)^{w_{t-1}}E\left(U_t+1\right)^{w_t} \\
&\ \ \vdots \\
a_{2^{\binom{t}{2}}-1} &= p^{\binom{t}{2}-1}qE\left(U_1+t-1\right)^{w_1}\cdots E\left(U_{t-1}+t-2\right)^{w_{t-1}}E\left(U_t+t-2\right)^{w_t} \\
a_{2^{\binom{t}{2}}} &= p^{\binom{t}{2}}E\left(U_1+t-1\right)^{w_1}\cdots E\left(U_{t-1}+t-1\right)^{w_{t-1}}E\left(U_t+t-1\right)^{w_t}
\end{aligned} \tag{A.1}$$

For example,

$$
\begin{aligned}
A_{321} &= \sum_{i=1}^{8} a_i \quad \text{where} \\
a_1 &= q^3 EU_1 EU_2^2 EU_3^3 \\
a_2 &= pq^2 EU_1 E\left(U_2+1\right)^2 E\left(U_3+1\right)^3 \\
a_3 &= pq^2 E\left(U_1+1\right) EU_2^2 E\left(U_3+1\right)^3 \\
a_4 &= pq^2 E\left(U_1+1\right) E\left(U_2+1\right)^2 EU_3^3 \\
a_5 &= p^2 q E\left(U_1+1\right) E\left(U_2+1\right)^2 E\left(U_3+2\right)^3 \\
a_6 &= p^2 q E\left(U_1+1\right) E\left(U_2+2\right)^2 E\left(U_3+1\right)^3 \\
a_7 &= p^2 q E\left(U_1+2\right) E\left(U_2+1\right)^2 E\left(U_3+1\right)^3 \\
a_8 &= p^3 E\left(U_1+2\right) E\left(U_2+2\right)^2 E\left(U_3+2\right)^3
\end{aligned}
$$

Although the calculations based on (A.1) are straight forward they are somewhat cumbersome. An alternative approach is given by Frank (1979). Below follows some general formulas and formulas for special cases deduced from (A.1), that are integral parts of the moments of Section 3.

$$
\begin{aligned}
EX_1 X_2 \cdots X_t &= A_{[t]} \\
&= \sum_{c=0}^{\left\lfloor \frac{t}{2} \right\rfloor} \binom{t}{2c} \binom{2c}{c} \frac{c!}{2^c} \left(n-1\right)^{t-2c} p^{t-c} q^c \\
&= \left(n-1\right)^t p^t + t! \sum_{c=1}^{\left\lfloor \frac{t}{2} \right\rfloor} \frac{2^{-c}\left(n-1\right)^{t-2c}}{c!\left(t-2c\right)!} p^{t-c} q^c
\end{aligned}
$$

$$
\begin{aligned}
EX_1^w X_2 &= A_{w1} \\
&= \left(n-1\right) p \sum_{j=1}^{w} \frac{S\left(w,j\right)\left(n-1\right)!}{\left(n-1-j\right)!} p^j + q \sum_{j=1}^{w} \frac{jS\left(w,j\right)\left(n-2\right)!}{\left(n-1-j\right)!} p^j \\
&= \sum_{j=1}^{w} S\left(w,j\right)\left(\left(n-1\right)^2 p + jq\right) \frac{\left(n-2\right)!}{\left(n-j-1\right)!} p^j
\end{aligned}
$$

$$
\begin{aligned}
EX_1^w X_2^w &= A_{ww} \\
&= p\left(\sum_{k=1}^{w}\sum_{j=1}^{k}\binom{k}{j}\frac{S(k,j)\,(n-2)!}{(n-2-j)!}p^j+1\right)^2+q\left(\sum_{j=1}^{w}\frac{S(w,j)\,(n-2)!}{(n-2-j)!}p^j\right)^2
\end{aligned}
$$

$$
A_{211}=(n-1)^4\,p^3+(3n-1)\,p^2q+(n-3)\left(2\,(n-1)-n^2\,(n-2)\right)p^3q
$$

$$
\begin{aligned}
A_{42} &= p+(n+16)\,(n-2)\,p^2 \\
&\quad +2\,(n-2)\left(4n^2+15n-59\right)p^3 \\
&\quad +(n-2)\,(n-3)\left(13n^2+21n-114\right)p^4 \\
&\quad +(n-2)\,(n-3)\left(7n^3-22n^2-49n+136\right)p^5 \\
&\quad (n-2)^2\,(n-3)\,(n-4)\left(n^2-2n-7\right)p^6
\end{aligned}
$$

$$
\begin{aligned}
A_{411} &= (3n+11)\,p^2+\left(n^3+32n^2-69n-64\right)p^3 \\
&\quad +\left(7n^4+7n^3-212n^2+362n+66\right)p^4 \\
&\quad +\left(6n^5-39n^4+12n^3+339n^2-570n+36\right)p^5 \\
&\quad +\left(n^6-12n^5+47n^4-40n^3-144n^2+268n-48\right)p^6
\end{aligned}
$$

$$
\begin{aligned}
A_{321} &= (3n+5)\,p^2+\left(n^3+17n^2-30n-40\right)p^3 \\
&\quad +(n-3)\left(4n^3+23n^2-59n-26\right)p^4 \\
&\quad +2\,(n-3)\left(2n^4-3n^3-28n^2+51n+5\right)p^5 \\
&\quad +(n-2)\,(n-3)\left(n^4-5n^3-2n^2+26n-2\right)p^6
\end{aligned}
$$

$$
\begin{aligned}
A_{3111} &= 3p^2+\left(6n^2+24n-48\right)p^3 \\
&\quad +\left(n^4+23n^3-81n^2-46n+142\right)p^4 \\
&\quad +\left(3n^5-6n^4-69n^3+210n^2-42n-132\right)p^5 \\
&\quad +\left(n^6-9n^5+20n^4+26n^3-120n^2+58n+36\right)p^6
\end{aligned}
$$

$$
\begin{aligned}
A_{222} &= 3(n+1)p^2 + \left(n^3 + 12n^2 - 21n - 20\right)p^3 \\
&\quad +3(n-3)\left(n^3 + 6n^2 - 10n - 6\right)p^4 \\
&\quad +3(n-3)\left(n^4 - 17n^2 + 20n + 4\right)p^5 \\
&\quad +n(n-3)^2\left(n^3 - 3n^2 + 12 - 6n\right)p^6
\end{aligned}
$$

$$
\begin{aligned}
A_{2211} &= 3p^2 + \left(6n^2 + 12n - 28\right)p^3 \\
&\quad + \left(n^4 + 14n^3 - 40n^2 - 58n + 98\right)p^4 \\
&\quad + \left(2n^5 + n^4 - 68n^3 + 145n^2 + 12n - 104\right)p^5 \\
&\quad + \left(n^6 - 8n^5 + 13n^4 + 34n^3 - 96n^2 + 28n + 32\right)p^6
\end{aligned}
$$

$$
\begin{aligned}
A_{21111} &= 3(5n-1)p^3 + \left(10n^3 + 9n^2 - 105n + 50\right)p^4 \\
&\quad + \left(n^5 + 9n^4 - 70n^3 + 68n^2 + 102n - 74\right)p^5 \\
&\quad + \left(n^6 - 7n^5 + 6n^4 + 40n^3 - 62n^2 - 18n + 28\right)p^6
\end{aligned}
$$

From the product moments we have

$$
\begin{aligned}
m_2 &= E\left(Z^2\right) = E\left(n\sum_{i=1}^{n}(X_i - \overline{X})^2\right)^2 \\
&= n^2\left[nEX^4 + n(n-1)A_{22}\right] + n^4E\overline{X}^4 \\
&\quad -2n^2\left[EX^4 + 2(n-1)A_{31} + (n-1)A_{22} + (n-1)(n-2)A_{211}\right] \\
&= (n-2)EX^4 - 4(n-1)A_{31} + (n-1)(n-2)A_{22} \\
&\quad -2(n-1)(n-2)A_{211} + n^4E\overline{X}^4 \\
&= 2n(n-1)(n-2)^2pq + n_{(4)}(n-2)(n+4)p^2q^2 \ ,
\end{aligned}
$$

$$
\begin{aligned}
m_3 &= E\left(Z^3\right) = E\left(n\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^3 \\
&= n^6EX_i^6 - 3n^4EX_i^4\left(\sum_{i=1}^{n}X_i\right)^2 + 3n^2EX_i^2\left(\sum_{i=1}^{n}X_i\right)^4 - E\left(\sum_{i=1}^{n}X_i\right)^6 \\
&= n^2\left(3+n\left(n-3\right)\right)EX^6 - 6n\left(n_{(3)}\right)A_{51} + 3n^2\left(n-1\right)\left(n\left(n-3\right)+7\right)A_{42} \\
&\quad -3n\left(n_{(3)}\right)\left(n-6\right)A_{411} - 6n\left(n_{(3)}\right)A_{33} - 12n\left(n_{(3)}\right)\left(n-4\right)A_{321} \\
&\quad +12n\left(n_{(4)}\right)A_{3111} + n\left(n_{(3)}\right)\left(n\left(n-3\right)+9\right)A_{222} \\
&\quad -3n\left(n_{(4)}\right)\left(n-6\right)A_{2211} + 3n\left(n_{(5)}\right)A_{21111} - n^6E\overline{X}^6 \\
&= 4n_{(3)}\left(n-2\right)^2pq + 2n_{(4)}\left(3n-4\right)\left[\left(n-2\right)\left(n+6\right)-8\right]p^2q^2 \\
&\quad +n_{(4)}\left[n^4\left(n+3\right)-4\left(3n-4\right)\left[3\left(n-2\right)\left(n+6\right)-\left(n+4\right)\right]\right]p^3q^3
\end{aligned}
$$

and

$$
\begin{aligned}
n^5E\left(\left(S_i^2\right)^2\right) &= n^5E\left(\left(\frac{1}{n}\left(X_i-\overline{X}\right)^2\right)^2\right) \\
&= \left(n-2\right)\left(\left(n-1\right)^2+1\right)EX^4 - 4\left(n-1\right)\left(\left(n-2\right)^2+n\right)A_{31} \\
&\quad +6\left(n-1\right)\left(n-2\right)A_{22} + 6\left(n-1\right)\left(n-2\right)\left(n-4\right)A_{211} \\
&\quad -4\left(n-1\right)_{(3)}A_{[4]} + n^3E\overline{X}^4 \\
&= \left(n-1\right)\left(n-2\right)\left(n\left(n-6\right)+12\right)pq + 3\left(n-1\right)_{(3)}\left(n\left(n-2\right)+8\right)p^2q^2.
\end{aligned}
$$

For R-conditional Bernoulli graphs we have (Johnson, Kotz & Kemp, 1992)

$$
EX_{(k)} = \frac{r!\left(n-1\right)!\left(\frac{n(n-1)}{2}-k\right)!}{\left(r-k\right)!\left(n-1-k\right)!\left(\frac{n(n-1)}{2}\right)!}.
$$

Lengthy computations yield

$$EX_1 = \frac{2r}{n} \quad , \quad EX_1^2 = \frac{2r}{n} + \frac{4r\,(r-1)}{(n+1)\,n},$$

$$EX_1^3 = \frac{2r}{n} + \frac{12r\,(r-1)}{(n+1)\,n} + \frac{8r\,(r-1)\,(r-2)\,(n-3)}{(n+1)\,n\,(n^2-n-4)},$$

$$\begin{aligned} EX_1^4 &= \frac{2r}{n} + \frac{28r\,(r-1)}{(n+1)\,n} + \frac{48r\,(r-1)\,(r-2)\,(n-3)}{(n+1)\,n\,(n^2-n-4)} \\ &\quad + \frac{16r\,(r-1)\,(r-2)\,(r-3)}{(n+2)\,(n+1)\,n\,(n^2-n-4)} \end{aligned}$$

and

$$\begin{aligned} EX_1^2X_2^2 &= \frac{2r}{n\,(n-1)} + \frac{4r\,(r-1)\,(4+n)}{(n+1)\,n\,(n-1)} + \frac{16r\,(r-1)\,(r-2)\,(n^2-5)}{(n+1)\,n\,(n-1)\,(n^2-n-4)} \\ &\quad + \frac{16r\,(r-1)\,(r-2)\,(r-3)\,(n-2)}{(n+2)\,n\,(n-1)\,(n^2-n-4)}. \end{aligned}$$

Finally we get

$$\begin{aligned} E\left(S^2 \mid R=r\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{2r}{n}\right)^2\right) = EX_1^2 - (EX_1)^2 \\ &= \frac{2r\,(n^2-n-2r)}{n^2\,(n+1)}, \end{aligned}$$

$$E\left(\left(S^2 \mid R=r\right)^2\right) = \frac{1}{n}EX_1^4 + \frac{n-1}{n}EX_1^2X_2^2 - \frac{8r^2}{n^2}EX_1^2 + \left(\frac{2r}{n}\right)^4$$

and

$$\begin{aligned} Var\left(S^2 \mid R=r\right) &= \frac{1}{n}EX_1^4 + \frac{n-1}{n}EX_1^2X_2^2 - \left(EX_1^2\right)^2 \\ &= \frac{8r\,(r-1)\,(n^2-n-2r)\,(n^2-n-2r-2)}{n^2\,(n+1)^2\,(n+2)\,(n^2-n-4)}. \end{aligned}$$

B The exact distribution of the degree variance for 3⩽n⩽6

Table B1 below contains the values of the parameters needed to determine the following probabilities:

$$
\begin{aligned}
P\left(S^2 = \frac{z}{n^2}\right) &= \sum_{i=1}^{m} a_i p^{r_i} q^{\binom{n}{2}-r_i} + \sum_{i=1}^{m} a_i p^{\binom{n}{2}-r_i} q^{r_i} \\
P\left(S^2 = \frac{z}{n^2} \mid R = r_i\right) &= \frac{a_i}{\binom{\binom{n}{2}}{r_i}}
\end{aligned}
$$

where $m = m\,(n, z)$, $a_i = a_i\,(n, z)$, $r_i = r_i\,(n, z)$ for $i = 1, ..., m$.

n	z	m	$a_1, ..., a_m$	$r_1, ..., r_m$
3	0	1	1	0
3	2	1	3	1
4	0	2	$1, 3$	$0, 2$
4	4	2	$6, 6$	$1, 3$
4	8	1	12	2
4	12	1	4	3
5	0	2	$1, 6$	$0, 5$
5	4	2	$15, 30$	$2, 3$
5	6	2	$10, 70$	$1, 4$
5	10	1	60	5
5	14	2	$30, 60$	$2, 3$
5	16	1	75	4
5	20	1	30	5
5	24	1	30	3
5	26	1	60	4
5	30	1	30	5
5	36	1	5	4
6	0	3	$1, 15, 70$	$0, 3, 6$
6	8	5	$15, 45, 270, 465, 810$	$1, 2, 4, 5, 7$
6	12	2	$180, 1080$	$3, 6$
6	20	4	$60, 480, 972, 1800$	$2, 4, 5, 7$
6	24	2	$180, 1530$	$3, 6$
6	32	3	$405, 810, 1755$	$4, 5, 7$
6	36	2	$80, 1080$	$3, 6$
6	44	3	$180, 480, 1080$	$4, 5, 7$
6	48	1	810	6
6	56	3	$30, 270, 630$	$4, 5, 7$
6	60	1	360	6
6	68	1	360	7
6	72	1	75	6
6	80	1	6	5

Table B1. The exact distribution of the degree variance for $3 \leqslant n \leqslant 6$.

C The possible values of the degree variance times $\mathbf{n^2}$ for $\mathbf{3 \leqslant n \leqslant 10}$

$n = 3:$
0 2

$n = 4:$
0 4 8 12

$n = 5:$
0 4 6 10 12 14 16 20 24 26 30 36

$n = 6:$
0 8 12 20 24 32 36 44 48 56 60 68 72 80

$n = 7:$
0 6 10 12 14 20 24 26 28 34 38 40 42 48 52 54 56 62 66 68 70 76 80 82 84 90 94 96 98 104 108 110 112 118 122 124 132 136 138 140 146 150 160

$n = 8:$
0 12 16 28 32 44 48 60 64 76 80 92 96 108 112 124 128 140 144 156 160 172 176 188 192 204 208 220 224 236 240 252 272 300

$n = 9:$
0 8 14 18 20 26 32 36 38 44 50 54 56 62 68 72 74 80 86 90 92 98 104 108 110 116 122 126 128 134 140 144 146 152 158 162 164 170 176 180 182 188 194 198 200 206 210 216 218 224 230 234 236 242 248 252 254 260 266 270 272 278 284 288 290 296 302 306 308 314 320 324 326 332 338 342 344 350 356 360 362 368 374 378 380 386 392 396 398 404 410 414 416 422 428 432 434 446 450 458 470 504

$n = 10:$
0 16 20 24 36 40 44 56 60 64 76 80 84 96 100 104 116 120 124 136 140 144 156 160 164 176 180 184 196 200 204 216 220 224 236 240 244 256 260 264 276 280 284 296 300 304 316 320 324 336 340 344 356 360 364 376 380 384 396 400 404 416 420 424 436 440 444 456 460 464 476 480 484 496 500 504 516 520 524 536 540 544 556 560 564 576 580 584 596 600 604 616 620 624 636 640 644 656 660 664 676 680 684 696 700 704 716 744 756 784