Research Report

Department of Statistics

No. 2003:7

General Moments of Degrees in Random Graphs

Jan Hagberg

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

General Moments of Degrees in Random Graphs

Jan Hagberg

Abstract

Degree moments and functions of degree moments are investigated for three random graph models. The degree of vertex i in a graph is the number of edges incident to vertex i. Exact and asymptotic formulas are given for various degree statistics, in particular the degree variance S^2.

Key words: Uniform Random Graphs, Degree Sequences, Bernoulli Graphs, Degree Moments, Degree Statistics, Degree Variance.

1 Introduction

In statistical applications of random graph models the degree moments, and functions of the degree moments, have been found useful both as summary statistics and for inference on particular random graph models. See, for example, Hagberg (2000, 2003), Snijders (1981a), and Wasserman & Faust (1994). Section 2 introduces the three random graph models of this paper. The moment notation and formulas for basic statistical measures are given in Section 3. Sections 4 and 5 give particular and general formulas for various mixed moments of the type $EX_1^{m_1} \cdots X_t^{m_t}$, where $X_1, ..., X_t$ are the degrees of t distinct vertices in a random graph of order n. Here t is the dimension of the moment and $m = m_1 + ... + m_t$ is the total order of the moment. Section 6 gives some applications to degree statistics with focus on the degree variance.

2 Degree distributions for some graph models

2.1 Uniform graphs of order n and size r

Let $\mathcal{G}(n,r)$ be the class of all graphs with r edges on n fixed labeled vertices, $|\mathcal{G}(n,r)| = \binom{N}{r}$ where $N = \binom{n}{2}$. If the r pairs of distinct vertices i and j connected by an edge, are chosen uniformly at random without replacement among the N pairs, we get a random graph $\mathbf{G}$ that is uniformly distributed on $\mathcal{G}(n,r)$. This is denoted by $\mathbf{G} \sim Uniform(n,r)$ or $\mathbf{G} \sim Uniform(\mathcal{G}(n,r))$ and we have

$$P(\mathbf{G} = G) = p(G) = \frac{1}{\binom{N}{r}} \text{ for each } G \in \mathcal{G}(n,r).$$

Let $\mathbf{X}$ be the random adjacency matrix of $\mathbf{G}$. The entries X_{ij} in the $n \times n$ zero-one matrix $\mathbf{X}$ indicate which pairs of vertices are adjacent. If vertex i and j are adjacent, then $X_{ij} = 1$, and if vertex i and j are not adjacent, then $X_{ij} = 0$. Obviously $X_{ij} = X_{ji}$ and by definition $X_{ii} = 0$. The N elements X_{ij} for $i < j$ are

$$X_{ij} \sim Bernoulli\left(\frac{r}{N}\right),$$

but they are not independent.

Let X_i be the number of edges incident to vertex i, i.e. the degree of vertex i. We have $X_i = \sum_{j=1}^{n} X_{ij}$ for $i = 1, ..., n$. It follows that X_i is hypergeometrically distributed with selection size r and subgroup sizes $n-1$ and $N-n+1$. This is denoted by $X_i \sim Hypg.(r; n-1, N-n+1)$ and means that

$$P(X_i = x) = p_x = \frac{\binom{n-1}{x}\binom{N-n+1}{r-x}}{\binom{N}{r}},$$
$$N = \binom{n}{2}, 0 \le r \le N \text{ and}$$
$$\max\left\{0, r - \binom{n-1}{2}\right\} \le x \le \min\{n-1, r\}. \quad (2.1)$$

Example: for $n = 5$, $r = 8$ we have that $2 \leq x \leq 4$, $p_2 = \frac{2}{15}$, $p_3 = \frac{8}{15}$ and $p_4 = \frac{5}{15}$.

Note that neither X_{ij} for all $i < j$ nor X_i for all $i = 1, ..., n$ are independent. For instance, let

$$\begin{aligned} X_1 &= X_{12} + U_1 \text{ and} \\ X_2 &= X_{12} + U_2, \end{aligned} \tag{2.2}$$

where $U_1 = \sum_{j=3}^{n} X_{1j}$ and $U_2 = \sum_{j=3}^{n} X_{2j}$. The multivariate distribution of (X_{12}, U_1, U_2) is hypergeometrically distributed with selection size r and subgroup sizes $1, n-2, n-2, N-1-2(n-2)$ which is denoted

$$(X_{12}, U_1, U_2) \sim Mult.Hypg.\left(r; 1, n-2, n-2, N-1-2(n-2)\right).$$

For details on the multivariate hypergeometric distribution, see, for instance, Johnson, Kotz & Balakrishnan (1997).We have that

$$P\left(X_1 = x, X_2 = y\right) = p_{xy} = P\left(\left(X_{12}, U_1, U_2\right) = (0, x, y) \text{ or } (1, x-1, y-1)\right)$$

$$= \frac{\binom{1}{0}\binom{n-2}{x}\binom{n-2}{y}\binom{N-1-2(n-2)}{r-0-x-y} + \binom{1}{1}\binom{n-2}{x-1}\binom{n-2}{y-1}\binom{N-1-2(n-2)}{r-1-x+1-y+1}}{\binom{N}{r}} \neq p_x p_y. \tag{2.3}$$

2.2 Uniform graphs of order n

Let $\mathcal{G}(n) = \bigcup_{r=0}^{N} \mathcal{G}(n, r)$. A random graph that is uniformly distributed on $\mathcal{G}(n)$ is denoted $\mathbf{G} \sim Uniform\left(\mathcal{G}(n)\right)$ or $\mathbf{G} \sim Uniform\,(n)$. In this model, all 2^N labeled graphs of order n have the same probability to occur, so that

$$P\left(\mathbf{G} = G\right) = p\left(G\right) = \frac{1}{2^N} \text{ for each } G \in \mathcal{G}(n).$$

The adjacency matrix $\mathbf{X}$ of $\mathbf{G}$ has elements

$$X_{ij} \sim Bernoulli\left(\frac{1}{2}\right), \; i \neq j.$$

The vertex degrees are

$$X_i \sim Bin(n-1, \frac{1}{2}) \ , \ i = 1, 2, ..., n$$

and the size of $\mathbf{G}$ is

$$R = \frac{1}{2}\sum_{i=1}^{n} X_i \sim Bin(N, \frac{1}{2}).$$

Note that here X_{ij} for all $i < j$ are independent but X_i for $i = 1, ..., n$ are not independent. If we use (2.2) again we have that X_{12}, U_1, U_2 are independent $Bernoulli\left(\frac{1}{2}\right), Bin\left(n-2, \frac{1}{2}\right), Bin\left(n-2, \frac{1}{2}\right)$ so that

$$\begin{aligned}
p_{xy} &= P\left(X_1 = x, X_2 = y\right) \\
&= P\left((X_{12}, U_1, U_2) = (0, x, y) \ \text{ or } \ (1, x-1, y-1)\right) \\
&= \frac{1}{2}\binom{n-2}{x}\left(\frac{1}{2}\right)^{n-2}\binom{n-2}{y}\left(\frac{1}{2}\right)^{n-2} \\
&\quad +\frac{1}{2}\binom{n-2}{x-1}\left(\frac{1}{2}\right)^{n-2}\binom{n-2}{y-1}\left(\frac{1}{2}\right)^{n-2} \\
&= p_x p_y 2\left[\frac{x}{n-1}\frac{y}{n-1} + \left(1 - \frac{x}{n-1}\right)\left(1 - \frac{y}{n-1}\right)\right] \\
&\neq p_x p_y. \qquad (2.4)
\end{aligned}$$

2.3 Bernoulli graphs of order n and edge probability p

The Bernoulli model of order n and edge probability p, is a generalization of the uniform model of order n. With probability p, $0 < p < 1$, each pair of distinct vertices i and j is connected by an edge. Hence, $\mathbf{G} \sim Bernoulli\,(n, p)$ means that $P\left(\mathbf{G} = G\right) = p^r q^{N-r}$ for each $G \in \mathcal{G}\left(n, r\right)$ where $r = 0, ..., N$ and $q = 1 - p$. Now the adjacency indicators, degrees and size of $\mathbf{G}$ have the following distributions

$$\begin{aligned}
X_{ij} &\sim Bernoulli(p) \ , i \neq j, \\
X_i &\sim Bin(n-1, p) \ , \ i = 1, 2, ..., n \text{ and} \qquad (2.5) \\
R &\sim Bin\left(N, p\right).
\end{aligned}$$

Here X_{ij} for all $i < j$ are independent but X_i for $i = 1, ..., n$ are not independent. For instance, using (2.2),

$$\begin{aligned} p_{xy} &= P(X_1 = x, X_2 = y) \\ &= P((X_{12}, U_1, U_2) = (0, x, y) \text{ or } (1, x-1, y-1)) \\ &= q\binom{n-2}{x} p^x q^{n-2-x} \binom{n-2}{y} p^y q^{n-2-y} \\ &\quad + p\binom{n-2}{x-1} p^{x-1} q^{n-1-x} \binom{n-2}{y-1} p^{y-1} q^{n-1-y} \\ &= p_x p_y \frac{1}{q}\left[\frac{x}{n-1}\frac{y}{n-1} + \left(1 - \frac{x}{n-1}\right)\left(1 - \frac{y}{n-1}\right)\right] \\ &\neq p_x p_y. \end{aligned} \tag{2.6}$$

3 Degree moments

3.1 Degrees in a graph

Consider a graph G on n vertices and r edges, i.e. a graph of order n and size r. The number of edges incident to vertex i, the degree of vertex i, is denoted by x_i, and $(x_1, ..., x_n)$ is the degree sequence of G. The sum of the degrees of the vertices of the graph is equal to twice the number of its edges, that is $\sum_{i=1}^{n} x_i = 2r$. The ordered degree sequence of G is the sequence $(d_1, d_2, ..., d_n)$ obtained by listing the vertex degrees $(x_1, x_{2,} ..., x_n)$ in non-increasing order $(d_1 \geq d_2 \geq ... \geq d_n)$. For example, with $n = 4$ and $r = 2$ we have 15 labeled graphs, 13 degree sequences $(x_1, ..., x_4)$ and 2 ordered degree sequences $(d_1, ..., d_4)$. One degree sequence, $(1, 1, 1, 1)$ is common to 3 labeled graphs, but the other 12 labeled graphs have unique degree sequences corresponding to the 12 permutations of the ordered degree sequence $(2, 1, 1, 0)$.

3.2 Exchangeability

The degrees of the vertices in the models of this paper are exchangeable random variables denoted $X_1, ..., X_n$. In particular exchangeability means

$$\begin{aligned} EX_i &= EX_1 , \\ EX_i^2 &= EX_1^2 , \\ EX_iX_j &= EX_1X_2 \ , i \neq j, \\ EX_i^2X_j &= EX_iX_j^2 = EX_1^2X_2 \ , i \neq j, \\ EX_i^2X_j^2 &= EX_1^2X_2^2 \ , i \neq j \ \text{ and so on.} \end{aligned}$$

3.3 Moment notation

We use the following notation for single moments of order m

$$A_m = EX_1^m,$$

and

$$A_{m_1 \cdots m_t} = EX_1^{m_1} \cdots X_t^{m_t},$$

for mixed moments of dimension t, separate orders $(m_1, ..., m_t)$ and total order $m = m_1 + ... + m_t$. Due to exchangeability it is no restriction to assume $m_1 \geq m_2 \geq ... \geq m_t$. For example, the moments of order 4 are $A_4, A_{31}, A_{22}, A_{211}$,and A_{1111}.

The factorial moment $E(X_1(X_1-1)(X_1-2)\cdots(X_1-m+1))$ is denoted

$$B_m = EX_1^{(m)}.$$

It is well known (See, for instance, Graham, Knuth and Patashnik (1994).) that powers and factorials are related according to

$$X_1^m = \sum_{k=0}^{m} S_{mk} \frac{X_1!}{(X_1-k)!} = \sum_{k=0}^{m} S_{mk} X_1^{(k)} \tag{3.1}$$

where S_{mk} are the Stirling numbers of the second kind. Therefore we have that

$$A_m = \sum_{k=1}^{m} S_{mk} B_k \tag{3.2}$$

where S_{mk}, the Stirling numbers of the second kind, can be explicitly given by

$$S_{mk} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^m . \tag{3.3}$$

The first few of these numbers are given in Table 1.

	k					
	1	2	3	4	5	6
m						
1	1					
2	1	1				
3	1	3	1			
4	1	7	6	1		
5	1	15	25	10	1	
6	1	31	90	65	15	1

Table 1. Stirling numbers of the second kind, S_{mk}

3.4 Means, variances, covariances and correlations

If we use the notation above, we obtain the following expressions for the means, variances, covariances and correlations.

$$\begin{aligned} EX_1 &= A_1, \\ VarX_1 &= A_2 - A_1^2, \\ Cov(X_1, X_2) &= A_{11} - A_1^2 , \\ Corr(X_1, X_2) &= \frac{A_{11} - A_1^2}{A_2 - A_1^2}, \end{aligned}$$

where $A_m = \sum_x x^m p_x$ and $A_{11} = \sum_x \sum_y xyp_{xy}$, and the p_x and p_{xy} are given in the previous section for the three models considered. This leads to the formulas given in Table 2.

		Uniform (n,r)	Uniform (n)	Bernoulli (n,p)
$E(X_1)$		$\frac{2r}{n}$	$\frac{n-1}{2}$	$(n-1)p$
$Var(X_1)$		$\frac{2r\left(n^2-n-2r\right)}{n^2(n+1)}$	$\frac{n-1}{4}$	$(n-1)pq$
$Cov(X_1,X_2)$		$-\frac{2r\left(n^2-n-2r\right)}{n^2(n+1)(n-1)}$	$\frac{1}{4}$	pq
$Corr(X_1,X_2)$		$-\frac{1}{n-1}$	$\frac{1}{n-1}$	$\frac{1}{n-1}$

Table 2. Means, variances, covariances and correlations.

We see that the correlations have the same absolute value for all the models and tend to zero for increasing n. In the sequel we will not treat the uniform graph model of order n separately since it is a particular case of the Bernoulli (n,p) model with $p=\frac{1}{2}$.

4 Degree moments in uniform graphs of order n and size r

Theorem 1 *For the Uniform(n,r)-graph the degrees X_i have moments*

$$A_m = EX_1^m = \sum_{k=1}^{m} S_{mk} \frac{k!\binom{r}{k}\binom{n-1}{k}}{\binom{N}{k}} \tag{4.1}$$

for $m=1,2,\ldots$ and $N=\binom{n}{2}$. Mixed moments are given by

$$A_{m_1m_2} = \frac{r}{n}\sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_2}\binom{m_1}{k_1}\binom{m_2}{k_2}\mu^*_{k_1k_2} + \left(1-\frac{r}{n}\right)\mu_{m_1m_2} \tag{4.2}$$

where

$$\mu_{m_1m_2} = \frac{\sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2} S_{m_1k_1}S_{m_2k_2}k_1!k_2!\binom{n-2}{k_1}\binom{n-2}{k_2}\binom{r}{k_1+k_2}}{\binom{N-1}{k_1+k_2}}$$

and μ^* is obtained from μ by replacing r by $r-1$.

Proof. According to (2.1) X_1 has a hypergeometric distribution. Its factorial moments are given by Johnson, Kotz & Kemp (1992), and they are

$$B_m = EX_1^{(m)} = \frac{m!\binom{r}{m}\binom{n-1}{m}}{\binom{N}{m}}.$$

Hence, from (3.2)

$$A_m = \sum_{k=1}^{m} S_{mk} B_k = \sum_{k=1}^{m} S_{mk} \frac{k!\binom{r}{k}\binom{n-1}{k}}{\binom{N}{k}}.$$

The mixed moments

$$\begin{aligned} A_{m_1 m_2} &= EX_1^{m_1} X_2^{m_2} = E\left(X_{12}+U_1\right)^{m_1}\left(X_{12}+U_2\right)^{m_2} \\ &= \frac{r}{n} E\left[\left(1+U_1\right)^{m_1}\left(1+U_2\right)^{m_2} \mid X_{12}=1\right] + \left(1-\frac{r}{N}\right) E\left[U_1^{m_1} U_2^{m_2} \mid X_{12}=0\right] \\ &= \frac{r}{n} \sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} \binom{m_1}{k_1}\binom{m_2}{k_2} E\left[U_1^{k_1} U_2^{k_2} \mid X_{12}=1\right] + \left(1-\frac{r}{N}\right) E\left[U_1^{m_1} U_2^{m_2} \mid X_{12}=0\right] \end{aligned}$$

where

$$\left(U_1, U_2 \mid X_{12}=0\right) \sim Mult.Hypg.\left(r, n-2, n-2, N-1-2(n-2)\right)$$

and

$$\begin{aligned} E\left(U_1 U_2 \mid X_{12}=0\right) &= \frac{\sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} S_{m_1 k_1} S_{m_2 k_2} k_1! k_2! \binom{n-2}{k_1}\binom{n-2}{k_2}\binom{r}{k_1+k_2}}{\binom{N-1}{k_1+k_2}} \\ &= \mu_{m_1 m_2}(r) \text{ say.} \end{aligned}$$

Hence

$$A_{m_1 m_2} = \frac{r}{n} \sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} \binom{m_1}{k_1}\binom{m_2}{k_2} \mu_{k_1 k_2}(r-1) + \left(1-\frac{r}{n}\right) \mu_{m_1 m_2}(r).$$

■

Some particular formulas obtained from Theorem 1 are given below.

$$\begin{aligned} A_1 &= B_1, \\ A_2 &= B_1 + B_2, \\ A_3 &= B_1 + 3B_2 + B_3, \\ A_4 &= B_1 + 7B_2 + 6B_3 + B_4, \end{aligned}$$

and so on, where

$$\begin{aligned}
B_1 &= \frac{2r}{n}, \\
B_2 &= \frac{4r(r-1)}{(n+1)n}, \\
B_3 &= \frac{8(n-3)r(r-1)(r-2)}{(n+1)n(n^2-n-4)}, \\
B_4 &= \frac{16(n-4)r(r-1)(r-2)(r-3)}{(n+2)(n+1)n(n^2-n-4)}.
\end{aligned}$$

$$\begin{aligned}
A_{11} &= \frac{2r}{n(n-1)} + \frac{4r(r-1)}{(n+1)(n-1)}, \\
A_{21} &= \frac{2r}{n(n-1)} + \frac{4(n+2)r(r-1)}{(n+1)n(n-1)} + \frac{8(n^2-2n-1)r(r-1)(r-2)}{(n+1)n(n-1)(n^2-n-4)}, \\
A_{22} &= \frac{2r}{n(n-1)} + \frac{4(4+n)r(r-1)}{(n+1)n(n-1)} + \frac{16(n^2-5)r(r-1)(r-2)}{(n+1)n(n-1)(n^2-n-4)} \\
&\quad + \frac{16(n-2)r(r-1)(r-2)(r-3)}{(n+2)n(n-1)(n^2-n-4)}.
\end{aligned}$$

Furthermore, by using

$$\begin{aligned}
E\left(\sum_{i=1}^{n} X_i\right)^3 &= (2r)^3 \\
&= nA_3 + \frac{3!}{2!}2\binom{n}{2}A_{21} + 3!\binom{n}{3}A_{111}
\end{aligned}$$

we find that

$$A_{111} = \frac{12r(r-1)}{(n+1)n(n-2)} + \frac{8(n^2-2n-2)r(r-1)(r-2)}{(n+1)n(n-2)(n^2-n-4)}.$$

5 Degree moments in Bernoulli graphs of order n and edge probability p

Theorem 2 *For the Bernoulli* (n,p)*-graph the degrees* X_i *have moments* $A_m = EX_1^m$ *and* $A_{m_1m_2} = EX_1^{m_1}X_2^{m_2}$ *given by*

$$A_m = \sum_{k=1}^{m} S_{mk}k!\binom{n-1}{k}p^k, \tag{5.1}$$

$$A_{m_1m_2} = A_{m_1}A_{m_2} + \left(\frac{q}{p}\right)\left(A_{m_1} - A_{m_1}^*\right)\left(A_{m_2} - A_{m_2}^*\right), \tag{5.2}$$

where A^* *is obtained from* A *by replacing* n *by* $n-1$.

Proof. $A_m = \sum_{k=1}^{m} S_{mk}B_k$ and $B_k = (n-1)^{(k)}\,p^k$ according to Johnson, Kotz & Kemp (1992).

Write $X_1 = X_{12} + U_1$ and $X_2 = X_{12} + U_2$ where $U_i = \sum_{j=3}^{n} X_{ij}$ for $i = 1, 2$, and $n \geq 3$. Now X_{12}, U_1, U_2 are independent *Bernoulli* (p), *Bin* $(n-2,p)$, and *Bin* $(n-2,p)$. Thus we obtain by conditioning on X_{12}

$$\begin{aligned} A_{m_1m_2} &= E\left(X_{12}+U_1\right)^{m_1}\left(X_{12}+U_2\right)^{m_2} \\ &= qEU_1^{m_1}EU_2^{m_2} + pE\left(1+U_1\right)^{m_1}E\left(1+U_2\right)^{m_2}. \end{aligned}$$

Now

$$A_m = E\left(X_{12}+U_1\right)^m = qEU_1^m + pE\left(1+U_1\right)^m$$

and

$$EU_1^m = A_m^*$$

so that

$$E\left(1+U_1\right)^m = \frac{A_m - qA_m^*}{p}.$$

By substitution into the formula for $A_{m_1m_2}$

$$A_{m_1m_2} = qA_{m_1}^*A_{m_2}^* + \frac{\left(A_{m_1} - qA_{m_1}^*\right)\left(A_{m_2} - qA_{m_2}^*\right)}{p}$$

which simplifies to the expression given for $A_{m_1m_2}$. ∎

Note: it is also possible to use

$$A_{m_1m_2} = qA^*_{m_1}A^*_{m_2} + p\sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_1}\binom{m_1}{k_1}\binom{m_2}{k_2}A^*_{k_1}A^*_{k_2}.$$

Theorem 3 *For the Bernoulli* (n,p)*-graph the degrees* X_i *have moments* $A_{m_1m_2m_3} = EX_1^{m_1}X_2^{m_2}X_3^{m_3}$ *given by*

$$\begin{aligned}
A_{m_1m_2m_3} &= q^3A^{**}_{m_1}A^{**}_{m_2}A^{**}_{m_3} \\
&\quad +pq^2\frac{A^{**}_{m_1}\left(A^*_{m_2}-qA^{**}_{m_2}\right)\left(A^*_{m_3}-qA^{**}_{m_3}\right)}{p^2} \\
&\quad +pq^2\frac{\left(A^*_{m_1}-qA^{**}_{m_1}\right)A^{**}_{m_2}\left(A^*_{m_3}-qA^{**}_{m_3}\right)}{p^2} \\
&\quad +pq^2\frac{\left(A^*_{m_1}-qA^{**}_{m_1}\right)\left(A^*_{m_2}-qA^{**}_{m_2}\right)A^{**}_{m_3}}{p^2} \\
&\quad +p^2q\frac{\left(A_{m_1}-2qA^*_{m_1}+q^2A^{**}_{m_1}\right)\left(A^*_{m_2}-qA^{**}_{m_2}\right)\left(A^*_{m_3}-qA^{**}_{m_3}\right)}{p^4} \\
&\quad +p^2q\frac{\left(A^*_{m_1}-qA^{**}_{m_1}\right)\left(A_{m_2}-2qA^*_{m_2}+q^2A^{**}_{m_2}\right)\left(A^*_{m_3}-qA^{**}_{m_3}\right)}{p^4} \\
&\quad +p^2q\frac{\left(A^*_{m_1}-qA^{**}_{m_1}\right)\left(A^*_{m_2}-qA^{**}_{m_2}\right)\left(A_{m_3}-2qA^*_{m_3}+q^2A^{**}_{m_3}\right)}{p^4} \\
&\quad +p^3\frac{\left(A_{m_1}-2qA^*_{m_1}+q^2A^{**}_{m_1}\right)\left(A_{m_2}-2qA^*_{m_2}+q^2A^{**}_{m_2}\right)}{p^8} \\
&\quad \times\left(A_{m_3}-2qA^*_{m_3}+q^2A^{**}_{m_3}\right) \qquad (5.3)
\end{aligned}$$

where A^{**} *is obtained from* A *by replacing* n *by* $n-2$.

Proof.

$$\begin{aligned}
A_m &= E\left(X_{12}+X_{13}+U_1\right)^m \\
&= q^2EU_1^m + 2pqE\left(1+U_1\right)^m + p^2E\left(2+U_1\right)^m
\end{aligned}$$

where U_1 is $Bin\left(n-3,p\right)$. Let $EU_1^m = A_m^{**}$. Now

$$E\left(1+U_1\right)^m = \frac{A_m^* - qA_m^{**}}{p}$$

and

$$\begin{aligned} E\left(2+U_1\right)^m &= \frac{A_m - q^2A_m^{**} - 2pqE\left(1+U_1\right)^m}{p^2} \\ &= \frac{A_m - q^2A_m^{**} - 2q\left(A_m^* - qA_m^{**}\right)}{p^2} \\ &= \frac{\left(A_m - 2qA_m^* + q^2A_m^{**}\right)}{p^2}. \end{aligned}$$

Theorem 3 follows by substituting these expressions into

$$A_{m_1m_2m_3} = E\left(X_{12}+X_{13}+U_1\right)^{m_1}\left(X_{12}+X_{23}+U_2\right)^{m_2}\left(X_{13}+X_{23}+U_3\right)^{m_3}.$$

■

Theorem 4 *For the Bernoulli (n,p)-graph with degree sequence $(X_1,...,X_n)$ the mixed moment $A_{1,...,1} = EX_1\cdots X_t = A_{[t]}$ is given by*

$$\begin{aligned} A_{[t]} &= \sum_{m=0}^{\left\lfloor \frac{t}{2}\right\rfloor} c_{tm}p^{t-m} \\ &= \sum_{m=0}^{\left\lfloor \frac{t}{2}\right\rfloor} \frac{t!\left(n-1\right)^{t-2m}}{m!\left(t-2m\right)!2^m}p^{t-m}q^m \end{aligned} \qquad (5.4)$$

where

$$c_{tm} = \sum_{k=0}^{\left\lfloor \frac{t}{2}\right\rfloor - m} \left(-1\right)^k \frac{t!\left(n-1\right)^{t-2m-2k}}{m!k!\left(t-2m-2k\right)!2^{m+k}}.$$

Proof.

$$EX_1\cdots X_t = E\sum_{\substack{j_1=1\\ j_1\neq 1}}^{n}\cdots\sum_{\substack{j_t=1\\ j_t\neq t}}^{n} X_{1j_1}\cdots X_{tj_t}$$

Let c_{tm} be the number of sequences $(j_1, ..., j_t)$ with m "doubles" and $t-2m$ "singles". Here (j_r, j_s) is a "double" if $j_r = s$ and $j_s = r$ and $j_r < j_s$. Let e_{ij} be the event that (i,j) is a "double" for $1 \leq i < j \leq t$. We define

$$S_1 = \sum_{i<j} P(e_{ij}) \ , \ S_2 = \sum_{i<j} \sum_{k<l} P(e_{ij} \cap e_{kl}) \ , \ ...$$

which equals

$$S_1 = \binom{t}{2} \left(\frac{1}{n-1}\right)^2 \ , \ S_2 = \binom{t}{2} \binom{t-2}{2} \frac{1}{2} \left(\frac{1}{n-1}\right)^4 \ , \ ...$$

In general we have

$$\begin{aligned} S_m &= \binom{t}{2}\binom{t-2}{2} \cdots \binom{t-2(m-1)}{2} \frac{1}{m!} \left(\frac{1}{n-1}\right)^{2m} \\ &= \frac{t!}{(t-2m)!m!2^m (n-1)^{2m}} \ , \text{ for } m \leq \left\lfloor \frac{t}{2} \right\rfloor . \end{aligned}$$

Now $c_{tm}/(n-1)^t$ is the probability of exactly m "doubles" and it can be obtained as (See Feller (1970).)

$$\frac{c_{tm}}{(n-1)^t} = \sum_{k=0}^{\lfloor \frac{t}{2} \rfloor - m} (-1)^k \binom{m+k}{k} S_{m+k}$$

which simplifies to

$$c_{tm} = \sum_{k=0}^{\lfloor \frac{t}{2} \rfloor - m} (-1)^k \frac{t! (n-1)^{t-2m-2k}}{m!k! (t-2m-2k)! 2^{m+k}} \ .$$

Each term with exactly m "doubles" contributes p^{t-m} to the expected value $A_{[t]}$ and therefore

$$A_{[t]} = \sum_{m=0}^{\lfloor \frac{t}{2} \rfloor} c_{tm} p^{t-m}.$$

■

Generally for the *Bernoulli* (n,p)-model we have,

$$\begin{aligned} A_{m_1 \ldots m_t} &= E \prod_{i=1}^{t} E\left(X_i^{m_i} | X_i - U_i\right), \ t < n \\ &= E \prod_{i=1}^{t} E\left(U_i + x_i\right)^{m_i} = E \prod_{i=1}^{t} a_{m_i}\left(x_i\right) \end{aligned}$$

where

$$U_i = \sum_{j=t+1}^{n} X_{ij} \sim Bin\left(n-t, p\right)$$

are independent for $i = 1, ..., t$ and also independent of $\left(X_1 - U_1, ..., X_t - U_t\right)$, and $\left(x_1, ..., x_t\right) = \left(X_1 - U_1, ..., X_t - U_t\right)$ is the degree sequence of a *Bernoulli* (t,p) model. Hence

$$\begin{aligned} E\left(U_i + x\right)^{m_i} &= a_{m_i}\left(x\right) \\ &= \sum_{k=0}^{m_i} \binom{m_i}{k} x^{m_i - k} E U_i^k \qquad (5.5) \\ &= x^{m_i} + \sum_{k=1}^{m_i} \binom{m_i}{k} x^{m_i - k} \sum_{j=1}^{k} S_{kj} \left(n-t\right)^{(j)} p^j. \end{aligned}$$

Although the calculations based on (5.5) are straight forward, they are somewhat cumbersome and they have to be followed by taking the expectation over $\left(X_1 - U_1, ..., X_t - U_t\right)$. For large t this is prohibitive. An alternative approach is given by Frank (1979). Below follows the formulas for the moments of total orders $2, 3, 4$, and 6.

$$A_2 = \left(n-1\right) p + \left(n-1\right)\left(n-2\right) p^2$$

$$A_{11} = p + n\left(n-2\right) p^2$$

$$A_3 = (n-1)\,p + 3\,(n-1)\,(n-2)\,p^2 + (n-1)\,(n-2)\,(n-3)\,p^3$$

$$A_{21} = p + (n+2)\,(n-2)\,p^2 + (n-2)\left(n^2 - 2n - 1\right)p^3$$

$$A_{111} = (n-1)^3\,p^3 + 3\,(n-1)\,p^2 q$$

$$\begin{aligned} A_4 &= (n-1)\,p + 7\,(n-1)\,(n-2)\,p^2 + 6\,(n-1)\,(n-2)\,(n-3)\,p^3 \\ &\quad + (n-1)\,(n-2)\,(n-3)\,(n-4)\,p^4 \end{aligned}$$

$$\begin{aligned} A_{31} &= p + (n+6)\,(n-2)\,p^2 + 3\,(n-2)\,(n\,(n-1)-4)\,p^3 \\ &\quad + (n-2)\,(n-3)\,(n\,(n-2)-2)\,p^4 \end{aligned}$$

$$\begin{aligned} A_{22} &= p + (n-2)\,(n+4)\,p^2 + [(2n-4)\,((n-1)\,(n-3) + 4\,(n-2))]\,p^3 \\ &\quad + (n+1)\,(n-2)^2\,(n-3)\,p^4 \end{aligned}$$

$$\begin{aligned} A_{211} &= (3n-1)p^2 + [(n+4)(n-1)(n-3) + 2(n+1)(n-2)]p^3 \\ &\quad +[(n+1)(n-1)\,(n-2)\,(n-3) - n(n-3)]p^4 \end{aligned}$$

$$A_{1111} = (n-1)^4\,p^4 + 6\,(n-1)^2\,p^3 q + 3p^2 q^2$$

$$\begin{aligned} A_6 &= (n-1)\,p + 31\,(n-1)\,(n-2)\,p^2 + 90\,(n-1)\,(n-2)\,(n-3)\,p^3 \\ &\quad +65\,(n-1)\,(n-2)\,(n-3)\,(n-4)\,p^4 \\ &\quad +15\,(n-1)\,(n-2)\,(n-3)\,(n-4)\,(n-5)\,p^5 \\ &\quad + (n-1)\,(n-2)\,(n-3)\,(n-4)\,(n-5)\,(n-6)\,p^6 \end{aligned}$$

$$\begin{aligned} A_{51} &= p + (n-2)\,(n+30)\,p^2 + 15\,(n-2)\left(n^2 + 3n - 16\right)p^3 \\ &\quad +5\,(n-2)\,(n-3)\left(5n^2 - 2n - 42\right)p^4 \\ &\quad +5\,(n-2)\,(n-3)\,(n-4)\left(2n^2 - 3n - 11\right)p^5 \\ &\quad + (n-2)\,(n-3)\,(n-4)\,(n-5)\left(n^2 - 2n - 4\right)p^6 \end{aligned}$$

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

$$
\begin{aligned}
A_{411} &= (3n+11)\,p^2 + \left(n^3+32n^2-69n-64\right)p^3 \\
&\quad +(n-3)\left(7n^3+28n^2-128n-22\right)p^4 \\
&\quad +3(n-3)\left(2n^4-7n^3-17n^2+62n-4\right)p^5 \\
&\quad +(n-3)(n-4)\left(n^4-5n^3+20n-4\right)p^6
\end{aligned}
$$

$$
\begin{aligned}
A_{33} &= p + (n-2)(n+12)\,p^2 + 6(n-2)\left(n^2+5n-16\right)p^3 \\
&\quad +(n-2)(n-3)\left(11n^2+23n-94\right)p^4 \\
&\quad +3(n-2)^2(n-3)\left(2n^2-n-19\right)p^5 \\
&\quad +(n-2)^2(n-3)^2(n-4)(n+2)\,p^6
\end{aligned}
$$

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

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

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

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

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

$$
A_{111111} = \left(n-1\right)^6 p^6 + 15\left(n-1\right)^4 p^5 q + 45\left(n-1\right)^2 p^4 q^2 + 15p^3 q^3
$$

6 Applications to degree statistics

6.1 Degree mean

In the $Uniform\left(n,r\right)$ model, r is fixed and the degree mean $\overline{x} = 2r/n$ is not random. Under the $Bernoulli\left(n,p\right)$ model with $N = \binom{n}{2}$ we have

$$
\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{2R}{n} \sim \frac{2}{n} Bin\left(N,p\right)
$$

and

$$
E\overline{X}^m = \left(\frac{2}{n}\right)^m \sum_{k=1}^{m} S_{mk} N^{(k)} p^k.
$$

In particular

$$
E\overline{X} = \left(n-1\right)p, \tag{6.1}
$$

and

$$
Var\overline{X} = \frac{2\left(n-1\right)}{n} pq. \tag{6.2}
$$

6.2 Degree variance

The moments of the degree variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ are essential for the approximate distributions of S^2. Hagberg (2000 and 2003) shows that S^2 is approximately gamma distributed with parameters obtained from the first two moments of S^2. Another application of the degree variance under the $Uniform\,(n,r)$-model, can be found in Snijders (1981b).

Theorem 5 *For the* $Uniform\,(n,r)$*-graph the degree variance* S^2 *has expected value*

$$ES^2 = \frac{2r\left(n^2-n-2r\right)}{n^2\left(n+1\right)} \tag{6.3}$$

and variance

$$VarS^2 = \frac{8r\left(r-1\right)\left(n^2-n-2r\right)\left(n^2-n-2r-2\right)}{n^2\left(n+1\right)^2\left(n+2\right)\left(n^2-n-4\right)}. \tag{6.4}$$

Proof. First note that $EX_i = A_1 = \overline{x} = 2r/n$. By writing $S^2 = \frac{1}{n}\sum X_i^2 - A_1^2$ we obtain

$$\begin{aligned} ES^2 &= A_2 - A_1^2, \\ ES^4 &= \frac{1}{n^2}\sum\sum EX_i^2X_j^2 - 2A_1^2\frac{1}{n}\sum EX_i^2 + A_1^4 \\ &= \frac{1}{n}A_4 + \frac{n-1}{n}A_{22} - 2A_1^2A_2 + A_1^4. \end{aligned}$$

Substitution leads to

$$ES^2 = \frac{2r\left(n^2-n-2r\right)}{n^2\left(n+1\right)}$$

and

$$\begin{aligned} ES^4 &= \frac{4\left(n-2\right)^2 r}{n^4} + \frac{4\left(n^3-n^2-16n+28\right)r\left(r-1\right)}{n^4\left(n+1\right)} \\ &\quad - \frac{16\left(n^3-n^2-14n+24\right)r\left(r-1\right)\left(r-2\right)}{n^4\left(n+1\right)\left(n^2-n-4\right)} \\ &\quad + \frac{16\left(2n-8+n^2\right)r\left(r-1\right)\left(r-2\right)\left(r-3\right)}{n^4\left(n+1\right)\left(n+2\right)\left(n^2-n-4\right)}. \end{aligned}$$

The variance of S^2 is given by

$$\begin{aligned} VarS^2 &= ES^4 - \left(ES^2\right)^2 \\ &= \frac{1}{n}A_4 + \frac{n-1}{n}A_{22} - A_2^2 \end{aligned}$$

which by substitution becomes

$$VarS^2 = \frac{8r\,(r-1)\,(n^2-n-2r)\,(n^2-n-2r-2)}{n^2\,(n+1)^2\,(n+2)\,(n^2-n-4)}.$$

■

The mean and variance of S^2 can also be found in Snijders (1981a and 1981b).

Theorem 6 *For the Bernoulli (n,p)-graph the degree variance S^2 has expected value*

$$ES^2 = \frac{(n-1)\,(n-2)}{n}pq \tag{6.5}$$

and variance

$$VarS^2 = \frac{2(n-1)(n-2)^2}{n^3}pq\,(1+(n-6)pq)\,. \tag{6.6}$$

The correlation coefficient between the degree mean $\overline{X}$ and the degree variance S^2 is equal to

$$Corr\left(S^2,\overline{X}\right) = \frac{q-p}{\sqrt{1+(n-6)\,pq}},\; n>2. \tag{6.7}$$

In particular, for $p=\frac{1}{2}$, the degree mean and the degree variance are uncorrelated for $n>2$.

Proof. For the *Bernoulli* (n,p) -model it is convenient to use that

$$S^2 = \frac{1}{n}\sum X_i^2 - \frac{1}{n^2}\sum X_iX_j$$

We obtain

$$\begin{aligned} ES^2 &= A_2 - \frac{1}{n^2}\left(nA_2 + n\left(n-1\right)A_{11}\right) \\ &= \frac{n-1}{n}\left(A_2 - A_{11}\right). \end{aligned}$$

and substitution yields

$$ES^2 = \frac{\left(n-1\right)\left(n-2\right)}{n}pq.$$

Now

$$S^4 = \frac{1}{n^2}\sum\sum X_i^2X_j^2 - \frac{2}{n^3}\sum\sum\sum X_i^2X_jX_k + \frac{1}{n^4}\sum\sum\sum X_iX_jX_kX_l$$

and

$$\begin{aligned} ES^4 &= \frac{1}{n^2}\left(nA_4 + n\left(n-1\right)A_{22}\right) \\ &\quad -\frac{2}{n^3}\left(nA_4 + 2n\left(n-1\right)A_{31} + n\left(n-1\right)A_{22} + n\left(n-1\right)\left(n-2\right)A_{211}\right) \\ &\quad +\frac{1}{n^4}[(nA_4 + 4n\left(n-1\right)A_{31} + 3n\left(n-1\right)A_{22} \\ &\quad +6n\left(n-1\right)\left(n-2\right)A_{211} + n\left(n-1\right)\left(n-2\right)\left(n-3\right)A_{1111})] \end{aligned}$$

Substitution from the formulas given in Section 5 yields

$$\begin{aligned} ES^4 &= \frac{\left(n-1\right)\left(n-2\right)^2}{n^3}pq \\ &\quad +\frac{\left(n-1\right)\left(n-2\right)^2\left(n-3\right)\left(n+4\right)}{n^3}p^2q^2 \end{aligned}$$

and we obtain

$$VarS^2 = \frac{2(n-1)(n-2)^2}{n^3}pq\left(1 + (n-6)pq\right).$$

Furthermore

$$Corr\left(S^2, \overline{X}\right) = \frac{ES^2\overline{X} - ES^2E\overline{X}}{\sqrt{VarS^2}\sqrt{Var\overline{\overline{X}}}}$$

where

$$
\begin{aligned}
ES^2\overline{X} &= E\left(\frac{\overline{X}}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^3\right) \\
&= \frac{1}{n^2}\left(nA_3+n\left(n-1\right)A_{21}\right)-\frac{1}{n^3}\left(nA_3+6\binom{n}{2}A_{21}+6\binom{n}{3}A_{111}\right) \\
&= \frac{1}{n^2}\left(\left(n-1\right)A_3+\left(n-1\right)\left(n-3\right)A_{21}-\left(n-1\right)\left(n-2\right)A_{111}\right)
\end{aligned}
$$

and $E\overline{X}$, $Var\overline{X}$, ES^2, and $VarS^2$ are given by (6.1), (6.2), (6.3), and (6.4). Substitution yields

$$
Corr\left(S^2,\overline{X}\right)=\frac{q-p}{\sqrt{1+\left(n-6\right)pq}}, n>2.
$$

■

Furthermore, since R is binomially distributed and $\sum X_i = 2R$, we do not need $A_{[t]}$ for ES^t. We illustrate this for $t=6$.

$$
\begin{aligned}
ES^6 &= \frac{1}{n^6}E\left(\left(\sum_{i=1}^{n}X_i-\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2\right)\right)^3 \\
&= \frac{1}{n^6}\left(n^6A_6-3n^4EX_1^4\left(\sum_{i=1}^{n}X_i\right)^2+3n^2EX_1^2\left(\sum_{i=1}^{n}X_i\right)^4-E\left(2R\right)^6\right) \\
&= \frac{n^2\left(3+n\left(n-3\right)\right)A_6-6n\left(n^{(3)}\right)A_{51}+3n^2\left(n-1\right)\left(n\left(n-3\right)+7\right)A_{42}}{n^6} \\
&\quad+\frac{-3n\left(n^{(3)}\right)\left(n-6\right)A_{411}-6n\left(n^{(3)}\right)A_{33}-12n\left(n^{(3)}\right)\left(n-4\right)A_{321}}{n^6} \\
&\quad+\frac{12n\left(n^{(4)}\right)A_{3111}+n\left(n^{(3)}\right)\left(n\left(n-3\right)+9\right)A_{222}}{n^6} \\
&\quad+\frac{-3n\left(n^{(4)}\right)\left(n-6\right)A_{2211}+3n\left(n^{(5)}\right)A_{21111}-E\left(2R\right)^6}{n^6}
\end{aligned}
$$

By substitution this becomes

$$ES^6 = \frac{4n^{(3)}(n-2)^2}{n^6}pq + \frac{2n^{(4)}(3n-4)\left[(n-2)(n+6)-8\right]}{n^6}p^2q^2 + \frac{n^{(4)}\left[n^4(n+3)-4(3n-4)\left[3(n-2)(n+6)-n-4\right]\right]}{n^6}p^3q^3.$$

6.3 Squared degree deviation from the mean

Let

$$nS^2 = \sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 = \sum_{i=1}^{n} Q_i,$$
$$Q_i = \left(X_i-\overline{X}\right)^2.$$

Theorem 7 *For the $Uniform\,(n,r)$-graph the squared degree deviations from the mean, $Q_i = \left(X_i-\overline{X}\right)^2$, have expected values*

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

variances

$$VarQ_i = \frac{2(n-2)(n-4)^2 r}{n^4} + \frac{4\left(2n^4-11n^3+43n^2-48n-112\right)r(r-1)}{n^4(n+1)^2} - \frac{32(n-3)\left(n^3-2n^2+12n+16\right)r(r-1)(r-2)}{n^4(n+1)^2\left(n^2-n-4\right)} + \frac{32\left(n^3-2n^2+12n+16\right)r(r-1)(r-2)(r-3)}{n^4(n+1)^2(n+2)\left(n^2-n-4\right)}$$

and correlation coefficients

$$Corr\left(Q_i,Q_j\right) = \frac{n^2(n+1)\left(n^3-7n^2+10n-8\right)-8\left(n^2-6n-8\right)\left(n^2-n-2r\right)r}{-n^2(n+1)(n-1)^2\left(n^2-10n+8\right)-4(n-1)\left(n^3-2n(n-6)+16\right)\left(n^2-n-2r\right)r}.$$

In particular, $n^2 Corr\,(Q_i, Q_j)$ *tends to* $-\frac{1-4pq}{2pq}$ *for increasing* n *and* $2r/n\,(n-1)$ *tending to* p*. For* N *even and* $r = N/2$

$$Corr\,(Q_i, Q_j) = -\frac{4}{n\,(n-1)^2}\ ,$$

and for N *odd and* $r = (N/2) \pm \frac{1}{2}$

$$\begin{aligned} Corr\,(Q_i, Q_j) &= -\frac{4}{n\,(n-1)^2} \\ &\quad -\frac{8\,(n-4)^2}{n\,(n-1)^2\left((n-2)^5 + 5\,(n-2)^4 + 10\,(n-2)^3 + 4\,(n-2)^2 - 16\right)}. \end{aligned}$$

Proof. Due to exchangeability

$$\begin{aligned} Corr\,(Q_i, Q_j) &= \frac{EQ_1Q_2 - (EQ_1)^2}{VarQ_1} \\ &= \frac{EQ_1Q_2 - (EQ_1)^2}{EQ_1^2 - (EQ_1)^2} \text{ for } i \neq j. \end{aligned}$$

Here $Q_i = X_i^2 - 2A_1X_i + A_1^2$ so that

$$EQ_1 = A_2 - 2A_1^2 + A_1^2,$$

$$\begin{aligned} EQ_1^2 &= E\left(X_1^2 - 2A_1X_1 + A_1^2\right)^2 \\ &= A_4 + 4A_1^2A_2 + A_1^4 - 4A_1A_3 + 2A_1^2A_2 - 4A_1^3A_1 \\ &= A_4 - 4A_1A_3 + 6A_1^2A_2 - 3A_1^4\ , \end{aligned}$$

$$
\begin{aligned}
VarQ_1 &= EQ_1^2 - (EQ_1)^2 \\
&= A_4 - 4A_1A_3 + 6A_1^2A_2 - 3A_1^4 - \left(A_2 - 2A_1^2 + A_1^2\right)^2 \\
&= A_4 - 4A_1A_3 + 8A_1^2A_2 - A_2^2 - 4A_1^4 \\
&= \frac{2(n-2)(n-4)^2 r}{n^4} + \frac{4(2n^4 - 11n^3 + 43n^2 - 48n - 112)r(r-1)}{n^4(n+1)^2} \\
&\quad - \frac{32(n-3)(n^3 - 2n^2 + 12n + 16)r(r-1)(r-2)}{n^4(n+1)^2(n^2-n-4)} \\
&\quad + \frac{32(n^3 - 2n^2 + 12n + 16)r(r-1)(r-2)(r-3)}{n^4(n+1)^2(n+2)(n^2-n-4)},
\end{aligned}
$$

and

$$
\begin{aligned}
EQ_1Q_2 &= E\left(X_1^2 - 2A_1X_1 + A_1^2\right)\left(X_2^2 - 2A_1X_2 + A_1^2\right) \\
&= A_{22} - 2A_1A_{21} + A_1^2A_2 - 2A_1A_{21} + 4A_1^2A_{11} \\
&\quad -2A_1^3A_1 + A_1^2A_2 - 2A_1^3A_1 + A_1^4 \\
&= A_{22} - 4A_1A_{21} + 4A_1^2A_{11} + 2A_1^2A_2 - 3A_1^4
\end{aligned}
$$

Hence

$$
Corr(Q_1, Q_2) = \frac{A_{22} - 4A_1A_{21} + 4A_1^2A_{11} - A_2^2 + 4A_1^2A_2 - 4A_1^4}{A_4 - 4A_1A_3 + 8A_1^2A_2 - A_2^2 - 4A_1^4}
$$

$$
= \frac{n^2(n+1)(n^3 - 7n^2 + 10n - 8) - 8(n^2 - 6n - 8)(n^2 - n - 2r)r}{-n^2(n+1)(n-1)^2(n^2 - 10n + 8) - 4(n-1)(n^3 - 2n(n-6) + 16)(n^2 - n - 2r)r}.
$$

For increasing n and $2r/n(n-1)$ tending to p, we have

$$
n^2 Corr(Q_1, Q_2) \to \frac{1 - 4p + 4p^2}{-2p(1-p)} = -\frac{1 - 4pq}{2pq}. \tag{6.8}
$$

By substituting $r = N/2$ for N even, and $r = (N/2) \pm \frac{1}{2}$ for N odd into the formula for $Corr\left(Q_1, Q_2\right)$ we obtain

$$Corr\left(Q_1, Q_2\right) = -\frac{4}{n\left(n-1\right)^2} \text{ , for } N \text{ even and } r = N/2,$$

and

$$Corr\left(Q_1, Q_2\right) = -\frac{4}{n\left(n-1\right)^2}$$
$$-\frac{8\left(n-4\right)^2}{n\left(n-1\right)^2\left(\left(n-2\right)^5 + 5\left(n-2\right)^4 + 10\left(n-2\right)^3 + 4\left(n-2\right)^2 - 16\right)} \text{ ,}$$
$$\text{for } N \text{ odd and } r = (N/2) \pm \frac{1}{2}.$$

■

Theorem 8 *For the Bernoulli (n, p)-graph the squared degree deviations from the mean, $Q_i = \left(X_i - \overline{X}\right)^2$, have expected values*

$$EQ_i = \frac{\left(n-1\right)\left(n-2\right)}{n} pq,$$

variances

$$VarQ_i = \frac{\left(n-1\right)\left(n-2\right)\left(n\left(n-6\right)+12\right)}{n^3} pq + \frac{2\left(n-1\right)\left(n-2\right)\left(n^2\left(n-6\right)+20n-36\right)}{n^3} p^2 q^2$$

and correlation coefficients

$$Corr\left(Q_i, Q_j\right) = \frac{n\left(n+2\right) - 12 - 4\left(n\left(n+4\right)-18\right)pq}{\left(n-1\right)\left(\left(n\left(n-6\right)+12\right) + 2\left(n^2\left(n-6\right)+20n-36\right)pq\right)}$$

In particular, $n^2 Corr\left(Q_i, Q_j\right)$ tends to $\frac{1-4pq}{2pq}$ for increasing n. For $p = \frac{1}{2}$,

$$Corr\left(Q_i, Q_j\right) = -\frac{4\left(n-3\right)}{\left(n-1\right)\left(n^3 - 4n^2 + 8n - 12\right)}.$$

Proof. Here

$$\begin{aligned} Q_i &= X_i^2 - 2X_i\overline{X} + \overline{X}^2 \\ &= X_i^2 - \frac{2}{n}X_i\sum_j X_j + \frac{1}{n^2}\sum_i\sum_j X_iX_j \end{aligned}$$

so that, using the exchangeability,

$$\begin{aligned} EQ_1 &= A_2 - \frac{2}{n}\left(A_2 + (n-1)A_{11}\right) + \frac{1}{n^2}\left(nA_2 + n(n-1)A_{11}\right) \\ &= \frac{n-1}{n}\left(A_2 - A_{11}\right) \\ &= \frac{(n-1)(n-2)}{n}pq\ , \end{aligned}$$

$$\begin{aligned} EQ_1^2 &= A_4 - \frac{4}{n}\left(A_4 + (n-1)A_{31}\right) \\ &\quad + \frac{6}{n^2}\left(A_4 + (n-1)A_{22} + 2(n-1)A_{31} + (n-1)(n-2)A_{211}\right) \\ &\quad - \frac{4}{n^3}[A_4 + 4(n-1)A_{31} + 3(n-1)A_{22} + 6(n-1)(n-2)A_{211} \\ &\quad + (n-1)(n-2)(n-3)A_{1111}] \\ &\quad + \frac{1}{n^4}\left(nA_4 + 8\binom{n}{2}A_{31} + 6\binom{n}{2}A_{22} + 36\binom{n}{3}A_{211} + 24\binom{n}{4}A_{1111}\right) \\ &= \frac{1}{n^3}[(n-1)\left(n^2-3n+3\right)\left(A_4 - 4A_{31}\right) + 3(n-1)(2n-3)A_{22} \\ &\quad +3(n-1)(n-2)(n-3)\left(2A_{211} - A_{1111}\right)] \end{aligned}$$

$$\begin{aligned} VarQ_1 &= EQ_1^2 - \left(EQ_1\right)^2 \\ &= \frac{1}{n^3}\left((n-1)\left(n^2-3n+3\right)\left(A_4 - 4A_{31}\right) + 3(n-1)(2n-3)A_{22}\right) \\ &\quad + \frac{1}{n^3}3(n-1)(n-2)(n-3)\left(2A_{211} - A_{1111}\right) \\ &\quad - \frac{1}{n^2}(n-1)^2\left(A_2 - A_{11}\right)^2 \\ &= \frac{(n-1)(n-2)(n(n-6)+12)}{n^3}pq \\ &\quad + \frac{2(n-1)(n-2)\left(n^2(n-6)+20n-36\right)}{n^3}p^2q^2, \end{aligned}$$

and

$$
\begin{aligned}
EQ_1Q_2 &= A_{22} - \frac{4}{n}\left(A_{31} + A_{22} + (n-2)\,A_{211}\right) \\
&\quad + \frac{2}{n^2}\left(A_4 + (n-1)\,A_{22} + 2\,(n-1)\,A_{31} + (n-1)\,(n-2)\,A_{211}\right) \\
&\quad + \frac{4}{n^2}\left(2A_{31} + (n-2)\,A_{211} + 2A_{22} + 4\,(n-2)\,A_{211} + (n-2)\,(n-3)\,A_{1111}\right) \\
&\quad - \frac{4}{n^3}[A_4 + 4\,(n-1)\,A_{31} + 3\,(n-1)\,A_{22} + 6\,(n-1)\,(n-2)\,A_{211} \\
&\quad + (n-1)\,(n-2)\,(n-3)\,A_{1111}] \\
&\quad + \frac{1}{n^4}\left(nA_4 + 8\binom{n}{2}A_{31} + 6\binom{n}{2}A_{22} + 36\binom{n}{3}A_{211} + 24\binom{n}{4}A_{1111}\right) \\
&= \frac{1}{n^3}[(2n-3)\,(A_4 - 4A_{31}) + \left(n^2\,(n-2) - 3\,(n-3)\right)A_{22} \\
&\quad - (n+3)\,(n-2)\,(n-3)\,(2A_{211} - A_{1111})].
\end{aligned}
$$

Hence

$$
Corr\,(Q_1, Q_2) = \frac{a}{b},
$$

where

$$
\begin{aligned}
a &= (2n-3)\,(A_4 - 4A_{31}) + \left(n^2\,(n-2) - 3\,(n-3)\right)A_{22} \\
&\quad - (n+3)\,(n-2)\,(n-3)\,(2A_{211} - A_{1111}) - n\,(n-1)^2\,(A_2 - A_{11})^2
\end{aligned}
$$

and

$$
\begin{aligned}
b &= (n-1)\left(n^2 - 3n + 3\right)(A_4 - 4A_{31}) + 3\,(n-1)\,(2n-3)\,A_{22} \\
&\quad + 3\,(n-1)\,(n-2)\,(n-3)\,(2A_{211} - A_{1111}) - n\,(n-1)^2\,(A_2 - A_{11})^2\,.
\end{aligned}
$$

Substitution yields

$$
Corr\,(Q_1, Q_2) = \frac{n\,(n+2) - 12 - 4\,(n^2 + 4n - 18)\,pq}{(n-1)\,((n\,(n-6) + 12) + 2\,(n^3 - 6n^2 + 20n - 36)\,pq)},
$$

and

$$n^2 Corr\,(Q_1, Q_2) \to \frac{1-4pq}{2pq}. \tag{6.9}$$

For $p = \frac{1}{2}$

$$\begin{aligned} Corr\,(Q_1, Q_2) &= \frac{n\,(n+2) - 12 - (n^2 + 4n - 18)}{(n-1)\left((n\,(n-6)+12) + \frac{1}{2}\,(n^3 - 6n^2 + 20n - 36)\right)} \\ &= -\frac{4\,(n-3)}{(n-1)\,(n^3 - 4n^2 + 8n - 12)}. \end{aligned}$$

■

For both the models $Q_1, ..., Q_n$ are almost uncorrelated for large n and S^2 can be regarded as a sum of almost uncorrelated random variables. Under the $Uniform\,(n, r)$-modell $Corr\,(Q_1, Q_2)$ is negative and close to zero when $(2r/n\,(n-1)) \approx \frac{1}{2}$. Under the $Bernoulli\,(n, p)$-modell $Corr\,(Q_1, Q_2)$ is zero when $p = \frac{1}{2} \pm \left(\sqrt{2}\,(n-3)\,/\sqrt{54 + n\,(n+6)\,(n-5)}\right)$. We also see that the limiting correlations given by (6.8) and (6.9) have the same absolute value.

References

[1] Feller, W. (1970). *An Introduction to Probability Theory and its Applications, Vol 1, 3rd ed.* New York: Wiley.

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

[3] Graham, R., Knuth, D., and Patashnik, O. (1994). *Concrete mathematics: a foundation for computer science, 2nd ed.* Reading: Addison-Wesley.

[4] Hagberg, J (2000). *Centrality Testing and the Distribution of the Degree Variance in Bernoulli Graphs.* Stockholm University, Department of Statistics.

[5] Hagberg, J (2003). *Random Graph Distribution of Degree Variance.* Stockholm University, Department of Statistics.

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

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

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

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

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