Research Report

Department of Statistics

No. 2003:9

Extreme Values and Other Attained Values of the Degree Variance in Graphs

Jan Hagberg

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

Extreme Values and Other Attained Values of the Degree Variance in Graphs

Jan Hagberg

Abstract

Extreme values and other attained values of the degree variance are treated for two classes of graphs. In Class 1 the order of the graph is a fixed number n and the size of the graph is a fixed number r, $0 \leq r \leq \binom{n}{2}$. In Class 2, n is fixed but r is not fixed. The structure of the optimal graphs is investigated and it is shown that the maximum value of the degree variance can be obtained from integer sequences associated to the triangular numbers. Explicit formulas for the number of possible values and recurrence relations for the attained values of the degree variance are developed.

Key words: Degree Variance, Maximum Sum of Squares, Minimum Sum of Squares, Integer Sequences

1 Introduction

This paper is concerned with attained values of the degree variance in graphs. The degree variance of a graph on n vertices with r edges is defined as

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

where x_i is the degree of vertex i and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 2r/n$. If $Q = \sum\limits_{i=1}^{n} x_i^2$ denotes the sum of the squares of the degrees we have that $s^2 = \frac{Q}{n} - \left(\frac{2r}{n}\right)^2$.

Hence, if we multiply s^2 by n^2 we get an integer valued quantity, denoted z,

$$\begin{aligned} z &= n^2 s^2 \\ &= nQ - 4r^2. \end{aligned}$$

For graphs with fixed values on n and r, there is a one-to-one correspondence between possible distinct values on Q and s^2.This paper will show that there are some integer sequences associated to z and Q, that are of theoretical and/or practical interest. An example is the connection between the integer sequence

$$1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, ...$$

and the number of possible values attained by Q for fixed n and r. A related sequence is

$$1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, ...$$

The distinct values attained by Q are also connected to these sequences. The importance of knowing the possible values of Q is illustrated by Hagberg (2003b) who shows that for certain random graphs it is possible to improve on common approximations to the probability distributions of the degree variance.

Two classes of graphs are treated here. In Class 1 the order of the graph is a fixed number n and the size of the graph is a fixed number r, $0 \leq r \leq \binom{n}{2}$. In Class 2, r is not fixed. Section 2 treats the maximum value of the degree variance and Section 3 the minimum value of the degree variance in the two classes. In Section 4, we develop a counting formula for the number of possible values of the degree variance. Finally, Section 5 treats other attained values of the degree variance.

2 The maximum value of the degree variance

2.1 Class 1

For graphs of order n and size r, the maximum value of $Q = \sum_{i=1}^{n} x_i^2$ and/ or $s^2 = \frac{Q}{n} - \left(\frac{2r}{n}\right)^2$has been treated in at least seven papers and one textbook known to the author: Snijders (1981a, b), Boesch et al.(1990) , Mahadev and Peled (1995), de Caen (1998) , Peled et al. (1999), Caro and Yuster (2000) and Hagberg (2000). The proof by Caro and Yuster is based on the connection between the number of triangles in the graph G and the maximum value of Q. Another method is to use the connection between threshold graphs and the maximum value of Q. Several equivalent properties may be used to characterize the threshold graphs and they are treated in detail by Mahadev and Peled (1995). Three of the equivalent properties are given below:

(1) There exists some hyperplane that strictly separates the characteristic vectors of the independent sets of vertices of G from those of the non-independent sets.

(2) Every three distinct vertices i, j, k of G satisfy the following condition in terms of the adjacency indicators x_{ij} for vertices i and j in G:

$$x_{ik} \leq x_{jk} \text{ for all } k \neq i, j \text{ whenever } x_i \leq x_j.$$

(3) G can be constructed from the one-vertex graph by repeatedly adding an isolated vertex or a universal one (a vertex adjacent to every other vertex).

The first property gives rise to the name "threshold graphs". We use here the second property and the fact that a graph G and its complement $\overline{G}$ has the same degree variance s^2.

Theorem 1 *Consider graphs of order n and size r. If G is such a graph of maximum degree variance, then G has the threshold property that $x_{ik} \leq x_{jk}$ for all $k \neq i, j$ whenever $x_i \leq x_j$.*

Proof. We prove the theorem by contradiction. If G does not have the threshold property, then a graph G^* can be constructed which has larger degree variance. To see this, assume that $x_i \leq x_j$ and $x_{ik} > x_{jk}$ for some distinct i, j, k. Hence $x_{ik} = 1$ and $x_{jk} = 0$. Let G^* be the graph obtained from G by moving the edge from (i, k) to (j, k) so that $x^*_{ik} = 0$, $x^*_{jk} = 1$, $x^*_i = x_i - 1$, and $x^*_j = x_j + 1$. It follows that $Q^* = Q + 2 + 2x_j - 2x_i \geq Q + 2$ since $x_j \geq x_i$, i.e. $Q^* > Q$ and $s^{*2} > s^2$. ■

Theorem 2 *Consider graphs of order n and size r satisfying the threshold property that $x_{ik} \leq x_{jk}$ for all $k \neq i, j$ whenever $x_i \leq x_j$. If G is such a graph of maximum degree variance, then G consists of either*

1) a complete subgraph, at most one other vertex adjacent to some of the vertices in the subgraph, and possibly some isolated vertices, or

2) a set of saturated vertices, possibly some independent vertices, and at most one other vertex adjacent to all the saturated and some of the independent vertices.

Proof. The optimal graph of type 1 has $r = \binom{m-1}{2} + t$ with $0 < t < m$ and $m \leq n$ if there are m connected vertices and $n-m$ isolated vertices. Note that $t = m - 1$ implies that $r = \binom{m}{2}$ and that there is a complete subgraph and isolated vertices only. The optimal graph of type 2 has $r = (n-1)\,m - \binom{m}{2} + t$ with $0 \leq t < n - (m+1)$ and $m \leq n$ if there are m saturated vertices and $n-m$ non-saturated vertices of which one is adjacent to t of the non-saturated vertices. Note that $m = n$ implies that $t = 0$ and the graph is complete. For $m < n$ and $t = 0$ all non-saturated vertices have degree m.

In order to prove the theorem by contradiction we assume that G is a graph satisfying the threshold property but not being of type 1 or type 2 and we show that there exist a graph G^* satisfying the threshold property and being of larger degree variance than G.

Let G satisfy the threshold property and consist of a complete subgraph of order $m-1$ and two (or more) other vertices connected to t_1 and t_2 vertices in the subgraph, respectively, where $t_1 + t_2 = t$, $0 < t_2 \leq t_1$. Figure 1 illustrates a case of the adjacency matrix of G. Assume $x_{t_2,m+1} = 1$ and $x_{t_1+1,m} = 0$. Then $x_{t_2} = m$, $x_{m+1} = t_2$, $x_{t_1+1} = m - 2$, $x_m = t_1$.

	1	2	.	.	.	$m-1$	m	$m+1$	...	n
1	**0**	1	1	1	1	1	1	1	0	0
2	1	**0**	1	1	1	1	1	1	0	0
.	1	1	**0**	1	1	1	1	0	0	0
.	1	1	1	**0**	1	1	0	0	0	0
.	1	1	1	1	**0**	1	0	0	0	0
$m-1$	1	1	1	1	1	**0**	0	0	0	0
m	1	1	1	0	0	0	**0**	0	0	0
$m+1$	1	1	0	0	0	0	0	**0**	0	0
$\vdots$	0	0	0	0	0	0	0	0	**0**	0
n	0	0	0	0	0	0	0	0	0	**0**

Figure 1. The adjacency matrix of the threshold graph G.

Let G^* be the graph obtained from G by moving the edge at $(t_2, m+1)$ to (t_1+1, m). Now

$$\begin{aligned}
x^*_{t_2} &= x_{t_2} - 1 = m - 1, \\
x^*_{m+1} &= x_{m+1} - 1 = t_2 - 1, \\
x^*_{t_1+1} &= x_{t_1+1} + 1 = m - 1, \\
x^*_m &= x_m + 1 = t_1 + 1,
\end{aligned}$$

and it follows that

$$\begin{aligned}
Q^* &= Q + 4 + 2\left(x_{t_1+1} + x_m - x_{t_2} - x_{m+1}\right) \\
&= Q + 4 + 2\left(m - 2 + t_1 - m - t_2\right) \\
&= Q + 2\left(t_1 - t_2\right).
\end{aligned}$$

Hence $Q^* \geq Q$ with strict inequality if $t_2 < t_1$. When t_2 initially is equal to t_1, repeated application of this technique leads to a graph with $Q^* > Q$. When initially there are more than two other vertices connected to the complete subgraph, the same technique can be repeatedly applied to move the "last" 1 to the "first" 0 in the columns of the adjacency matrix corresponding to the extra vertices.

A similar argument can be applied to a graph satisfying the threshold property that consists of a set of saturated vertices, a set of independent vertices, and two or more other vertices connected to some of the independent vertices. ■

The optimal graph of type 1 with $r = \binom{m-1}{2} + t$ and $0 < t \leq m-1$ has

t vertices of degree $m-1$
$m-1-t$ vertices of degree $m-2$
1 vertex of degree t
$n-m$ vertices of degree 0,

and

$$\begin{aligned}
Q &= t\left(n-1\right)^2 + \left(m-1-t\right)\left(m-2\right)^2 \\
&= t^2 + \left(2m-3\right)t + \left(m-1\right)\left(m-2\right)^2.
\end{aligned}$$

Note that m is given by $m = 1 + \left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor$ where $\left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor$ yields the sequence

$$1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, ...$$

for $r = 1, 2, ...$ For details about the sequence, see, Graham et al. (1994) or Sloane (2003).

The optimal graph of type 2 with $r = (n-1)\,m - \binom{m}{2} + t$ and $0 \leq t \leq n - m - 2$ has

m vertices of degree $n-1$
1 vertex of degree $m+t$
t vertices of degree $m+1$
$n-m-1-t$ vertices of degree m,

and

$$\begin{aligned} Q &= m\,(n-1)^2 + (m+t)^2 + (m+1)^2\,t + (n-m-1-t)\,m^2 \\ &= t^2 + (4m+1)\,t + m\,(n-1)^2 + (n-m)\,m^2. \end{aligned}$$

Here m is given by $m = n - 1 - \left\lfloor \frac{1}{2} + \sqrt{n\,(n-1) - 2r} \right\rfloor$.

The sequence of optimal graphs of type 1 for $r = 1, 2, ...$ can be considered as locally grown with edges added one by one in the order $(1,2)\,,(1,3)\,,(2,3)\,,$ $(1,4)\,,(2,4)\,,(3,4)\,,(1,5)\,,...,$ so that the rth edge is adjacent to vertices t and m for $r = \binom{m-1}{2} + t$, $t = 1, ..., m-1$, $m = 2, 3, ...$ Note that t and m are uniquely determined by r. We write $m = m_r$ and $t = t_r$ if we need to specify m and t for distinct r.

The optimal graph of size r has m connected vertices, and the last edge connects vertices of degrees $m-1$ and t. The other edges all connect vertices of degree $m-1$. Before the last edge was added, its incident vertices had degrees $m-2$ and $t-1$. Therefore the increment in Q, say $Q_r - Q_{r-1}$, when the rth edge is added is seen to be

$$\begin{aligned} Q_r - Q_{r-1} &= (m-1)^2 - (m-2)^2 + t^2 - (t-1)^2 \\ &= 2\,(m+t-2) \end{aligned}$$

for $r = 1, 2, ...,$ with $Q_0 = 0$. It follows that

$$\begin{aligned} Q_r &= \sum_{k=1}^{r} 2\left(m_k + t_k - 2\right) \\ &= r\left(r-1\right) - r\left(a-3\right)a + 6\binom{a+1}{4} \end{aligned}$$

for $r = 1, 2, ...,$ and $a = \left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor$. The rth increment in Q is given as element (t, m) in the following matrix.

	$m=2$	$m=3$	$m=4$	$m=5$	$\cdots$
$t=1$	2	4	6	8	$\cdots$
$t=2$		6	8	10	$\cdots$
$t=3$			10	12	$\cdots$
$t=4$				14	$\cdots$
$\vdots$					$\cdots$

We see that $\frac{Q_r - Q_{r-1}}{2} = r - 1 + a - \binom{a}{2}$ yield the sequence

$$1, 2, 3, 3, 4, 5, 4, 5, 6, 7, 5, 6, 7, 8, 9, 6, 7, 8, 9, 10, 11, ...$$

We collect these results as the following theorem

Theorem 3 *For type 1 graphs of order n and size r, the maximum degree variance is given by*

$$s_{\max}^2\left(r\right) = \frac{Q_{\max}}{n} - \left(\frac{2r}{n}\right)^2$$

where

$$Q_{\max} = t^2 + \left(2m-3\right)t + \left(m-1\right)\left(m-2\right)^2$$

and t and m are determined by $r = \binom{m-1}{2} + t$ and $0 < t < m$ so that $m = 1 + \left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor$. Alternatively

$$
\begin{aligned}
Q_{\max} &= 2\sum_{k=1}^{r}(m_k+t_k-2)\\
&= r(r-1)-r(a-3)a+6\binom{a+1}{4}
\end{aligned}
$$

where t_k *and* m_k *are the values of* t *and* m *corresponding to* $r=k$ *for* $k=1,2,...,r$ *and* $a=\left\lfloor \frac{1}{2}+\sqrt{2r}\right\rfloor$.

Corollary 1 *For all graphs of order* n *and size* r*, the maximum degree variance is given by the largest of the two values* $s^2_{\max}(r)$ *and* $s^2_{\max}\left(\binom{n}{2}-r\right)$.

The optimal graph of type 2, mentioned in Theorem 2, with $r=(n-1)m-\binom{m}{2}+t$ and $0\le t\le n-m-2$ has

m vertices of degree $n-1$
1 vertex of degree $m+t$
t vertices of degree $m+1$
$n-m-1-t$ vertices of degree m,

and

$$
\begin{aligned}
Q &= m(n-1)^2+(m+t)^2+(m+1)^2t+(n-m-1-t)m^2\\
&= t^2+(4m+1)t+m(n-1)^2+(n-m)m^2.
\end{aligned}
$$

Here m is given by $m=n-1-\left\lfloor \frac{1}{2}+\sqrt{n(n-1)-2r}\right\rfloor$.

A similar incremental procedure can be advised for graphs of the second type. The sequences of optimal graphs of type 2 for $r=1,2,...,\binom{n}{2}$ are considered as globally grown with edges in the order $(1,2),(1,3),...,(1,n),(2,3)$, $(2,4),...,(2,n),(3,4),....$ We omit the further details and close this section with the following conjecture:

Conjecture 1 *For graphs of order* n *and size* r *we have*

$$Q_{2\max}-Q_{1\max}\ge 0 \text{ if } r\le \left\lfloor \frac{\binom{n}{2}}{2}-\frac{n}{3}\right\rfloor=\left\lfloor\frac{n(3n-7)}{12}\right\rfloor$$

$$Q_{2\max}-Q_{1\max}\le 0 \text{ if } r\ge \left\lfloor \frac{\binom{n}{2}}{2}+\frac{n}{3}\right\rfloor=\left\lfloor\frac{n(3n+1)}{12}\right\rfloor$$

where $Q_{1\max}$ *is the maximum sum of squares for graphs of type 1 and* $Q_{2\max}$ *is the maximum sum of squares for graphs of type 2.*

The conjecture is true at least for $n \leq 150$.

2.2 Class 2

Theorem 4 *A disconnected graph of order n has maximal degree variance if and only if it consists of k isolated vertices and a complete subgraph on the other $n-k=m$ vertices. A connected graph of order n has maximal degree variance if and only if it consists of k vertices of degree $n-1$ and no other edges.*

Proof. We use the fact that a graph G and its complement $\overline{G}$ have the same degree variance s^2. Let G_1 denote graphs of type 1. From Section 2.1 we have

$$\begin{aligned} n^2 s^2 (G_1) &= n\left((m-1)(m-2)^2 + 2(m+t-2)\right) - 4r^2 \\ &= n\left((m-1)(m-2)^2 + 2(m+t-2)\right) - 4\left(\binom{m-1}{2} + t\right) \\ &= (n-4)t^2 + [n(2m-3) - 4(m-1)(m-2)]t \\ &\quad + (n-m+1)(m-1)(m-2)^2 \\ &= f_m(t) \end{aligned}$$

For $n \leq 4$ it is easy to show that a complete graph on n vertices and no more edges yield the maximum value of $f_m(t)$. For $n > 4$ we see that $f_m(t)$ is a convex function of t, and for a fixed m, $f_m(t)$ has a minimum in t for a unique $t = t_m$. That is

$$\max f_m(t) \text{ occurs for } \begin{cases} t = m-1 \text{ if } t_m \leq 0 \\ t = 0 \text{ or } m-1 \text{ if } 0 < t_m < m-1 \\ t = 0 \text{ if } t_m \geq m-1. \end{cases}$$

In all the cases $f_m(0)$ or $f_{m+1}(0) = f_m(m-1)$ is the optimal (maximal) value of $f_m(t)$. Therefore

$$\max_r f(r) = \max_m f\left(\binom{m}{2}\right) = \max_m f\left(\binom{m_{opt}}{2}\right).$$

■

Theorem 4 is identical to the conjecture given by Caro and Yuster (2000), hence the conjecture is proved.

To determine the optimal m, we let $n - m_{opt} = k$.

Theorem 5 *For graphs of order n, the maximal degree variance is given by*

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

Proof.

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

that is

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

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

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

it follows that

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

and

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

Further, denote by k_n any value of k for which

$$f_n\,(k \pm 1) \leqslant f_n\,(k) \ .$$

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

$$\begin{aligned} \mathbf{A} \ :\ 0 &\leqslant k(n-k)(n-k-1)^2 - (k+1)(n-k-1)(n-k-2)^2 \\ &4k^2 - (5n-8)\,k + (n-2)^2 \leqslant 0 \\ &a_n \leqslant k \leqslant A_n \end{aligned}$$

where

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

$$\begin{aligned} \mathbf{B} \; : \; 0 \leqslant k(n-k)(n-k-1)^2 - (k-1)(n-k+1)(n-k)^2 \\ 4k^2 - 5nk + n\,(n+1) \geqslant 0 \\ k \leqslant b_n \;\text{ or } k \geqslant B_n \end{aligned}$$

where.

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

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

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

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

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

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

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

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. ■

3 The minimum value of the degree variance

3.1 Class 1

The degree variance is zero for any m-regular graph and we need the following result from graph theory.

Theorem 6 *(Chartrand and Lesniak (1996)) 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 $m = 0, 2, 4, ..., n-1$.*

Theorem 7 *Consider graphs of order n and size r. If G is such a graph of minimum degree variance, then G has k vertices of degree $m+1$ and $n-k$ vertices of degree m, where $m = \left\lfloor \frac{2r}{n} \right\rfloor$ and $k = 2r - nm$. The minimum degree variance is given by*

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

where

$$\begin{aligned}\theta &= \frac{2r}{n} - m \\ &= \frac{2r}{n} - \left\lfloor \frac{2r}{n} \right\rfloor .\end{aligned}$$

Proof. *The mean degree $\overline{x} = m + \theta$, $0 \leq \theta < 1$, and $2r = nm + k$, $0 \leq \theta < n$. If $\frac{2r}{n} = m$ where m is an integer, then $\theta = 0$, n is even if m is odd, 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 $m+1$. To obtain an optimal graph, we either construct an m-regular or an $(m+1)$-regular graph. If the m-regular graph is constructed, we add $k/2$ edges with no common vertices. If the $(m+1)$-regular graph is constructed, we remove $(n-k)/2$ edges with no common vertices. Hence,*

$$s^2_{\min} = \frac{n(1-\theta)(m-\overline{x})^2 + n\theta (m+1-\overline{x})^2}{n}$$

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

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

■

The optimal graph is not unique, for instance, consider the three graphs in Figure 2. They are all of order $n = 7$ size $r = 9$ and have $m = 2, k = 4$.

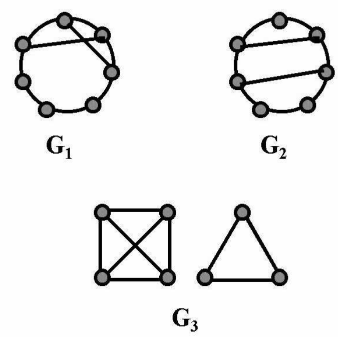

Figure 2

Both G_1 and G_2 are obtained by adding $(k/2) = 2$ disjoint edges to an m-regular graph for $m = 2$, $n = 7$. However, there may also be other ways to construct an optimal graph. The graph G_3 in Figure 2 is the disjoint union of two regular subgraphs and for $n - k > m$ and $k > m + 1$, it is possible to construct such optimal subgraphs . This follows since

$$n \text{ even } \Rightarrow k \text{ even so} \left\{ \begin{array}{c} \exists\ m\text{-reg of order } n-k>m \\ \exists\ (m+1)\text{-reg of order } k>m+1 \end{array} \right\rfloor$$

$$n \text{ odd, } m \text{ odd } \Rightarrow \left\{ \begin{array}{c} k \text{ odd so } \exists\ m\text{-reg of order } n-k>m \\ n-k \text{ even so } \exists\ (m+1)\text{-reg of order } k>m+1 \end{array} \right\rfloor$$

$$n \text{ odd, } m \text{ even } \Rightarrow \left\{ \begin{array}{c} k \text{ even so } \exists\ m\text{-reg of order } n-k>m \\ n-k \text{ odd so } \exists\ (m+1)\text{-reg of order } k>m+1 \end{array} \right\rfloor .$$

The minimum value of Q is given by

$$\begin{aligned} Q_{\min} &= n\left(s_{\min}^2 + \overline{x}^2\right) \\ &= n\left[(\overline{x} - m)(1 - \overline{x} + m) + \overline{x}^2\right] \\ &= 2r - \left\lfloor \frac{2r}{n} \right\rfloor \left(n - 4r + n\left\lfloor \frac{2r}{n} \right\rfloor\right). \end{aligned}$$

In particular, $r \leq \frac{n}{2}$ implies that $Q_{\min} = 2r$.

3.2 Class 2

In Class 2 there is no restriction on the size of the graph and any m-regular graph is obtainable for n even and any $2m$-regular graph is obtainable for n odd. The degree variance for any m-regular graph is zero. Hence, $\min s^2 = 0$ in Class 2.

4 Counting the number of possible values of the degree variance

Introduce the notation $M_{n,r}$ for the number of possible distinct values of Q among the graphs of order n and size r. For small and large values of r, we have the following explicit result. Note that $M_{n,r}$ does not depend on n for the range of r considered.

Theorem 8 *Consider graphs of order n and size $r \leq \frac{n}{2}$ or $r \geq \binom{n}{2} - \frac{n}{2}$. For such graphs we have*

$$M_{n,r} = \binom{r+3}{2} - a\,(r+2) + \binom{a+1}{3},$$

where $a = \left\lfloor \frac{1}{2} + \sqrt{2\,(r+3)} \right\rfloor$. An upper bound is given by $M_{n,r} \leq \binom{r}{2} + 1$.

Proof. From Section 2.1 we have that the maximum value of Q for $r \leq \frac{n}{2}$ is obtained if one vertex has degree r and r vertices have degree 1. All other vertices are isolated. This graph is a star and the maximum value of Q for such a graph is $r\,(r+1)$. The graph that yields the second highest value, has one vertex of degree $r-1$, two vertices of degree 2 and $r-3$ vertices of degree 1 and so on (see Figure 3). Let $A\,(t)$ denote the maximum value and $B\,(t)$ the minimum value of Q when we move t edges from the center of the star. $A\,(t)$ is obtained when we let t moved edges join one of the peripheral vertices with t other peripheral vertices. Then we have 1 vertex of degree $r-t$, 1 vertex of degree $1+t$, t vertices of degree 2 and $r-2t-1$ vertices of degree 1. We have that

$$A\,(t) = (r-t)^2 + t\,(t+4) + r$$

for $t \leq \frac{r-1}{2}$. The minimum value $B(t)$ is obtained when we let the t moved edges become isolated, so that we have 1 vertex of degree $r-t$ and $r-t+2t$ $= r+t$ vertices of degree 1. We have that

$$B(t) = (r-t)(r-t+1) + 2t.$$

To see that the difference between any two consecutive values of Q is 2 from $A(t)$ to $B(t)$, we first note that the $A(t)$ graph has t vertices of degree 2. Then, consider an edge connecting the vertex of degree $t+1$ with a vertex of degree 2. If we let this edge connect the vertex of degree $t+1$ with a previously isolated vertex, we reduce Q by 2. By repeating this t times, we reduce Q by $2t$. This yield a graph with 1 vertex of degree $(r-t)$, 1 vertex of degree $t+1$ and $r-1$ vertices of degree 1 (See the third graph for $t=2$ in Figure 3.). Move one of the edges connecting the vertex of degree $t+1$ with a vertex of degree 1 to two previously isolated vertices. Now we have reduced Q by $t^2 - (t+1)^2 + 1 = -2t$. By connecting the $t-1$ edges of the vertex of degree t with $t-1$ star vertices of degree 1, we increase Q by $2(t-1)$. The overall change of Q is $2(t-1) - 2t = -2$. We now have an $A(t-1)$ graph plus one edge connecting two previously isolated vertices. We repeat the process until we obtain a star with $r-t$ edges plus $2t$ vertices of degree 1 (See the last graph for $t=2$ in Figure 3.) i.e. a $B(t)$ graph.

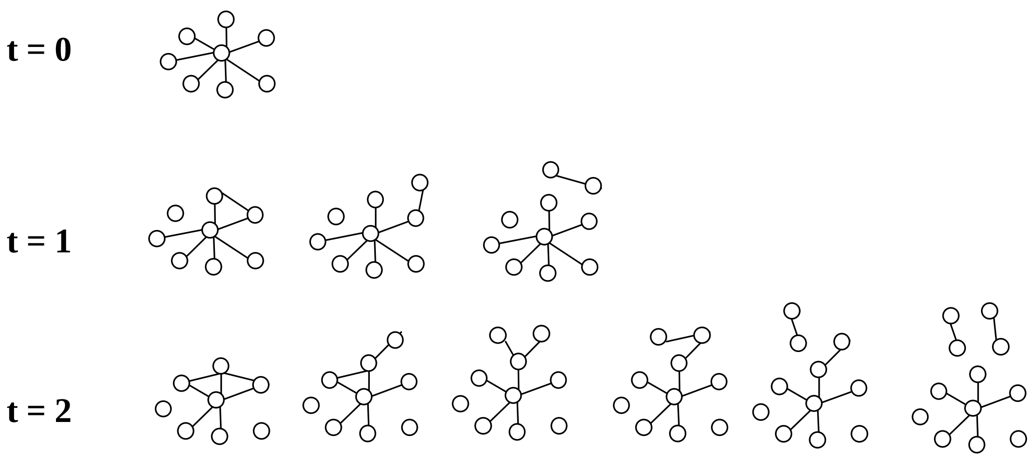

Figure 3. Graphs with r=7 edges and t=0,1,2.

						$A(t)-k$						
							k					
t	$A(t)$	2	4	6	8	10	12	14	16	18	$\cdots$	$2\left(\binom{t+2}{2}-1\right)$
0	r^2+r	-	-	-	-	-	-	-	-	-	-	-
1	$(r-1)^2+(1+4)+r$	x	x	-	-	-	-	-	-	-	-	-
2	$(r-2)^2+2(2+4)+r$	x	x	x	x	x	-	-	-	-	-	-
3	$(r-3)^2+3(3+4)+r$	x	x	x	x	x	x	x	x	x	-	-
.	.	x	x	x	x	x	x	x	x	x	$\cdots$	-
m	$(r-m)^2+m(m+4)+r$	x	x	x	x	x	x	x	x	x	x	x

Table 1. The distinct possible values of $Q(t)$ for each t. Each x means a possible value.

From Table 1 we see that there are $\binom{t+2}{2}$ distinct values for each t. When $A(t+1) \geq B(t)$ the difference between any two consecutive values is 2 from $A(t)$ down to the minimum value $2r$, which also can be seen from Table 1.

The difference $B(t) - A(t+1)$ is

$$B(t) - A(t+1) = 2(r-3) - t(t+7).$$

and from the inequality

$$2(r-3) \geq t(t+7)$$

solved for t, we have

$$t \leq \left\lfloor -\frac{7}{2} + \sqrt{2(r+3)} \right\rfloor = a - 4$$

where $a = \left\lfloor \frac{1}{2} + \sqrt{2(r+3)} \right\rfloor$. Hence

$$\begin{aligned} M_{n,r} &= \binom{r}{2} + 1 - \sum_{t=0}^{a-4} \left[\frac{B(t) - A(t+1)}{2} - 1 \right] \\ &= \binom{r}{2} + 1 - \left(\sum_{t=0}^{a-4} \left[\frac{2(r-3) - t(t+7)}{2} - 1 \right] \right) \\ &= \frac{(r+2)(r+3)}{2} - \frac{1}{6} a \left(6r + 13 - a^2\right) \\ &= \binom{r+3}{2} - a(r+2) + \binom{a+1}{3}. \end{aligned}$$

■

For $r = 1, ...,$ $M_{n,r}$ yields the sequence

$$1, 2, 4, 7, 10, 14, 19, 25, 31, 38, 46, 55, 65, 75, ...$$

According to the The On-Line Encyclopedia of Integer Sequences (Sloane (2003)) the convolution of the natural numbers $1, 2, ..., r$ with the sequence

$$1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, ...$$

denoted by A023532 in Sloane (2003), yields the sequence $1, 2, 4, 7, 10, 14, 19, 25, 31, 38, 46, 55, ...,$ here denoted by c_r or A023536 in Sloane (2003). One alternative, but not closed formula for c_r is given by the author in The On-Line Encyclopedia of Integer Sequences (Hagberg. (2002)).

We state the following theorem:

Theorem 9 $M_{n,r} = c_r$ *for graphs of order* n *and size* $r \le \frac{n}{2}$ *or* $r \ge \binom{n}{2} - \frac{n}{2}$.

Proof. The sequence A023532 takes the value zero for $m(m+3)/2$, and the inequality

$$\frac{m\,(m+3)}{2} \le r \ ,$$

solved for m, gives

$$m \le \left\lfloor -\frac{3}{2} + \sqrt{2r + \frac{9}{4}} \right\rfloor .$$

Consider the case $r = 9$ in Table 2.

									Sum
9	8	7	6	5	4	3	2	1	45
1	0	1	1	0	1	1	1	0	
9	0	7	6	0	4	3	2	0	31

Table 2. The convolution of the natural numbers $1, 2, ..., 9$ with A023532.

We see from Table 2 that c_9 is given by

$$c_9 = \frac{9\,(9+1)}{2} - (3\,(9+1) - (2+5+9)) = 31,$$

and in general, c_r is given by

$$\begin{aligned} c_r &= \frac{r\,(r+1)}{2} - \sum_{m=1}^{\left\lfloor -\frac{3}{2}+\sqrt{2r+\frac{9}{4}} \right\rfloor} \left((r+1) - \frac{m\,(m+3)}{2} \right) \\ &= \frac{r\,(r+1)}{2} - \left(\left\lfloor -\frac{3}{2} + \sqrt{2r+\frac{9}{4}} \right\rfloor (r+1) - \sum_{m=1}^{\left\lfloor -\frac{3}{2}+\sqrt{2r+\frac{9}{4}} \right\rfloor} \frac{m\,(m+3)}{2} \right) \\ &= \frac{(r+2)\,(r+3)}{2} - \frac{1}{6} b \left(6r + 13 - b^2 \right) \\ &= \binom{r+3}{2} - b\,(r+2) + \binom{b+1}{3} \end{aligned}$$

where $b = \left\lfloor \frac{1}{2} + \sqrt{2r + \frac{9}{4}} \right\rfloor$

Alternatively, since A023532 or 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, ..., is given by

$$\left\lfloor \frac{1}{2} + \sqrt{2\,(r+1)} \right\rfloor + 1 - \left\lfloor \frac{1}{2} + \sqrt{2\,(r+1)+1} \right\rfloor,$$

the convolution of the natural numbers with A023532 is given by

$$\begin{aligned} c_r &= \sum_{t=1}^{r} \left(\left\lfloor \frac{1}{2} + \sqrt{2\,(t+1)} \right\rfloor + 1 - \left\lfloor \frac{1}{2} + \sqrt{2\,(t+1)+1} \right\rfloor \right) (r+1-t) \\ &= \frac{(r+2)\,(r+3)}{2} - \frac{1}{6} \left\lfloor \frac{1}{2} + \sqrt{2r+3} \right\rfloor \left(13 + 6r - \left\lfloor \frac{1}{2} + \sqrt{2r+3} \right\rfloor^2 \right). \end{aligned}$$

The two formulas obtained for c_r and the formula for $M_{n,r}$ look different, but they yield the same sequence. For $r = 1, 2, 3, ...$, the rth element of the sequence

$$1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, ...$$

is given by

$$\left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor .$$

It follows that

$$\left\lfloor \frac{1}{2} + \sqrt{2(r+3)} \right\rfloor$$

yields the sequence

$$3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, ...$$

and

$$\left\lfloor \frac{1}{2} + \sqrt{2(r+2)} \right\rfloor$$

yields the sequence

$$2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, ...$$

For $-\frac{7}{4} \leq \delta < \frac{1}{4}$

$$\left\lfloor \frac{1}{2} + \sqrt{2r+4} \right\rfloor = \left\lfloor \frac{1}{2} + \sqrt{2r+4+\delta} \right\rfloor .$$

The lower bound of δ is given by the solution to the inequality

$$\frac{1}{2} + \sqrt{\frac{2\left(a(a+1)-2\right)}{2} + 4 + \delta} = \frac{1}{2} + \sqrt{a(a+1)+2+\delta} \geq a+1$$

and the upper bound is given by the solution to

$$\frac{1}{2} + \sqrt{\frac{2\left(a(a+1)-4\right)}{2} + 4 + \delta} = \frac{1}{2} + \sqrt{a(a+1)+\delta} < a+1.$$

Hence

$$\left\lfloor \frac{1}{2} + \sqrt{2r + \frac{9}{4}} \right\rfloor$$

yields the sequence

$$2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, ...$$

for $r = 1, 2, 3, ...$

Furthermore, let

$$a_r = \left\lfloor \frac{1}{2} + \sqrt{2\left(r+3\right)} \right\rfloor \text{ and } b_r = \left\lfloor \frac{1}{2} + \sqrt{2\left(r+2\right)} \right\rfloor = a_{r-1}\ ,\ r = 1, 2, 3, ...$$

That is, if $a_r = a_{r-1}$ we see that

$$\left(6r+13\right)\left(a_r - a_{r-1}\right) - \left(a_r^3 - a_{r-1}^3\right) = 0,$$

and $a_r - a_{r-1} = 1 \Leftrightarrow r = \binom{a_r}{2} - 2$, i.e.

$$6\left(\frac{a\left(a-1\right)-4}{2}\right) + 13 - \left(a^3 - \left(a-1\right)^3\right) = 0,\ \ a = 3, 4, 5, ...$$

Thus

$$M_{n,r} = c_r.$$

■

For other values of $r \leq n-1$, we can determine $M_{n,r}$ by calculating the minimum value of Q and adjusting $B\left(t\right)$. When $r = n-1$, the minimum value of Q is $2\left(2r-1\right)$, $B\left(t\right) = \left(r-t\right)^2 + 4t + \left(r-t\right)$ and $B\left(t\right) - A\left(t+1\right) = 2\left(r-3\right) - t\left(t+5\right)$. This yields

$$\begin{aligned} M_{n,r} &= \frac{r\left(r+1\right)}{2} - \frac{2\left(2r-2\right)}{2} - 2 + a - \frac{1}{2}\left(\sum_{t=0}^{a-3}\left(2\left(r-3\right) - t\left(t+5\right)\right)\right) \\ &= \frac{\left(r-1\right)\left(r+2\right)}{2} + \frac{1}{6}a\left(5 - 6r + a^2\right) \\ &= \binom{r+1}{2} - 1 - a\left(r-1\right) + \binom{a+1}{3}, \end{aligned}$$

where $a = \left\lfloor \frac{1}{2} + \sqrt{2r} \right\rfloor$ and $r = n - 1 > 2$.

For $(n-1) < r \leq \frac{n(n-1)}{4}$, $M_{n,r}$ is more complicated to determine due to the structures of the graphs that yield the maximum value of Q. Anyhow, results from computer calculations show that the difference between two consecutive values attained by Q is 2 when Q is not too close to its maximum value.

5 Other attained values of the degree variance

5.1 Class 1

By ordering the distinct values of $z = nQ - 4r^2$, we get the integer sequence $z_1 < z_2 < \cdots < z_m$ where z_1 is the minimal value given in Section 3.1 and z_m is the maximal value given in Section 2.1. In Section 4 we have seen that the difference between two consecutive values is 2 when Q is not too close to its maximum value. Since $z = nQ - 4r^2$, the following recurrence relation holds for graphs of order n and size r :

$$z_j = 2n + z_{j-1} \text{ for } j = 2, ..., t \leq m - 1 \tag{5.1}$$

where z_1 is given in Section 3.1. If the edges are generated according to the Uniform(n, r) random graph model (See Hagberg (2003 a, b).) the probability to obtain a z-value close to its maximum value decreases when $r \rightarrow \binom{n}{2}$. For large n this probability is very small and the recurrence relation given by (5.1) is useful even if t is not known. All possible values of the degree variance times n^2 for graphs of order 7 are given in Table 3.

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

The possible values of z in graphs of order $3, ..., 6$ and size $0, ..., \binom{n}{2}$ are given in Hagberg (2003b).

5.2 Class 2

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

$$z_j = 2n + z_{j-\lambda} \text{ for } j \geq 2, ..., t \leq m-1 \tag{5.2}$$

where λ is a integer valued lag length of the recurrence relation. If the number of edges are generated according to the Bernoulli(n, p) random graph model (See Hagberg (2003a, b).) the probability to obtain a z-value close to its maximum value decreases for fixed p when n increases. For $n \geq 9$ this probability is very small, even for p close to 0.5, and the recurrence relation given by (5.2) is useful even if t is not known. All values in the left tail can be derived by the formula for the minimum value of s^2 given in section 3.1. 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, tell us in how many subsequences $z_1, ..., z_m$ can be separated so that the difference between any two consecutive terms, except for the right tail, is $2n$ in every subsequence. From Table 3 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 from $a_1 < a_2 < \cdots < a_m$. Denote the set $\{a_1, ..., a_m\}$ by A. Remove from A such elements a_j if $a_i + 2n = a_j$ for $i = 1, ..., m$ to obtain a new set $\{b_1, ..., b_m\}$ denoted Λ. The number of elements in Λ equals the lag length λ of the recurrence relation (5.2). Further, $\Lambda \subseteq I$, the set of initial values, and I consists of $\{b_1, ..., b_m\}$ and $(b_i + 2n) < b_m$ for $i = 1, ..., m$. It should be noted that I might contain elements $(b_i + 2n) \notin A$. The reason is that the minimum value z_1 for fixed r might be larger than z_2 for another r. More formally we have

$$\begin{aligned}
Z(n) &= \{z_j(n,r) : j = 1, 2, ... \ \ , r = 0, 1, ...\} \\
A(n) &= \left\{ z : z = \min_r Z(n) \right\} \\
D(n) &= \left\{ z : z + 2n = \min_r Z(n) \right\} \\
\Lambda(n) &= A(n) - D(n) \\
|\Lambda(n)| &= \lambda \\
Z'(n) &= \left\{ z : z \leqslant \max_r \min_j z_j(n,r) \in \Lambda(n) \right\} \\
I(n) &= \Lambda(n) \cup Z'(n)
\end{aligned}$$

For example, for $n = 15$ we have

$$\begin{aligned}
A(15) &= \{0, 14, 26, 36, 44, 50, 54, 56\}, \\
\Lambda(15) &= \{0, 14, 26, 36, 50, 54\}, \\
I(15) &= \{0, 14, 26, 30, 36, 44, 50, 54\}, \\
z_j &= 30 + z_{j-6} \text{ and} \\
z_9 &= 56, \ z_{10} = 60, \ z_{11} = 66, ...
\end{aligned}$$

The lag length of the recurrence relation (5.2) and the initial values for $n = 7, ..., 20$ are listed in Table 4. Note that for $n = 4, 12, 20, 28, ...$, (5.2) can be written as $z_j = n + z_{j-\lambda/2}$. For n even, the lag length $\lambda_{n-1} > \lambda_n$, and under the *Bernoulli*(n, p)- model this give rise to less smoother distribution of the degree variance when n is odd. The implications of the smoothness of the distribution are discussed in Hagberg (2003b).

n	λ	The initial values of $z = n^2s^2$
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	0,32,36,56,68,72,80
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 4. The lag length and initial values of
$$z_j = 2n + z_{j-\lambda}.$$

All possible values of z for graphs of order $n = 1, ..., 12$ are given in Hagberg (2003b).

References

[1] Boesch, F., Brigham, R., Burr, S., Dutton, R. and, Tindell., R. (1990). *Maximizing the Sum of the Square of the Degrees of a Graph. Tech report, Stevens Inst Tech, Hoboken, NJ.*

[2] Caro,Y., and Yuster, R. (2000). *Graphs with Large Variance.* Ars Combinatoria, 57, 151 - 162.

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

[4] de Caen, D. (1998) *An Upper Bound of the Sum of squares in a Graph.* Discrete Math. 185, 245 - 248.

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

[6] Hagberg, J. (2000). *Centrality Testing and the Distribution of the Degree Variance in Bernoulli Graphs.* Research Report 2000:8, Department of Statistics, Stockholm University.

[7] Hagberg, J. (2002). "*Sequence A023536 Corrected*". In N, Sloane, (Ed). (2003). The On-Line Encyclopedia of Integer Sequences, published electronically at http://www.research.att.com/~njas/sequences/.

[8] Hagberg, J. (2003a). *General Moments of Degrees in Random Graphs.* Research Report 2003:7, Stockholm University, Department of Statistics.

[9] Hagberg, J. (2003b). *Random Graph Distribution of Degree Variance.* Research Report 2003:8, Department of Statistics, Stockholm University.

[10] Mahadev, N., Peled, U. (1995). *Threshold Graphs and Related Topics.* Annals of Discrete Mathematics. Amsterdam: North-Holland.

[11] Peled, U., Petreschi, R., and Sterbini, A. (1999). *(n,e)-Graphs with Maximum Sum of Squares of Degrees.* Journal of Graph Theory, 31, 283 - 295.

[12] Sloane, N. J. A. (Ed). (2003), The On-Line Encyclopedia of Integer Sequences, published electronically at http://www.research.att.com/~njas/sequences/.

[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.