# Maximum value and null moments

# of the degree variance

Tom Snijders

Department of Mathematics
University of Groningen

TW-229

april 1981

# Maximum value and null moments of the degree variance.

## 0. *Introduction*

This is companion paper to Snijders (1981), which will henceforth be indicated by DV. It contains proofs of some statements and formulas which were left unproven in DV.

Section 1 introduces the concept of a monotone graph, which is used in Section 2. In Section 2, the maximum value of the degree variance is computed for undirected graphs with a given number of points and a given density. In Section 3, this maximum is computed without restrictions on the density.

Section 4 is independent of the other sections. In this section, the mean and variance of the degree variance are computed under some simple stochastic null models.

All notation and terminology of DV is used without introducing it again.

## 1. *Monotone graphs*

The concept of a monotone graph plays an essential role in Section 2.

### 1.1. *Monotone graphs: introduction*

Consider an undirected graph G with incidence matrix $(x_{ij})$. Suppose that there exists points i,j,k with $i \neq k$ and

(1.1.1) $$x_j \leq x_k, \quad x_{ij} = 1, \quad x_{ik} = 0.$$

Construct the graph G' from G by switching the line i-j to i-k (i.e., $x'_{ij} = 0$ and $x'_{ik} = 1$, while $x'_{st} = x_{st}$ for all other (s,t)). Then

$$x'_j = x_j - 1 \qquad x'_k = x_k + 1$$

$$x'_t = x_t \qquad \text{for } t \neq j,k$$

$$x'_{.} = x_{.} \quad , \quad d' = d \quad ,$$

and hence for every convex function $\varphi : \mathbb{N}_0 \to \mathbb{R}$ we have

$$\Sigma_t \varphi(x'_t) - \Sigma_t \varphi(x_t) =$$

$$= \varphi(x_j - 1) + \varphi(x_k + 1) - \varphi(x_j) - \varphi(x_k) \geq 0,$$

with strict inequality if $\varphi$ is strictly convex. It follows that

$V_\varphi(G) \leq V_\varphi(G')$. The set of graphs which do not have three points i,j,k satisfying (1.1.1) will be defined as the set of monotone graphs.

*Definition.* A monotone graph is an undirected graph which satisfies

$$(x_{ij} = 1, \quad x_j \leq x_k, \quad i \neq k) \Rightarrow x_{ik} = 1.$$

The discussion above implies that in order to find the maximum of $V_\varphi(G)$ for undirected graphs G with fixed g and d, one may restrict attention to monotone graphs.

Let G be a monotone graph; in the sequel it will always be assumed that $x_1 \geq x_2 \geq \cdots \geq x_g$. Then

$$x_{ij} = 1 \quad \text{for} \begin{cases} 1 \leq j \leq x_i & \text{if } x_i \leq i-1 \\ 1 \leq j \leq x_{i+1} & \text{if } (x_i \geq i \text{ and } j \neq i) \end{cases} \tag{1.1.2}$$

$$x_{ij} = 0 \quad \text{for all other } (i,j).$$

It follows that every monotone graph is determined by its vector of degrees $(x_1, x_2, \ldots, x_g)$. Define

$$I = \min\{i \mid x_i \leq i-1\}.$$

Then a monotone graph is already determined by the degrees $x_i$ for $I \leq i \leq g$. To see this, note that

$$\begin{aligned} x_{ij} &= 1 \quad \text{for } 1 \leq j < i \leq I-1 \\ x_{ij} &= 1 \quad \text{for } 1 \leq j \leq x_i,\ i \geq I \\ x_{ij} &= 0 \quad \text{for } x_i < j \leq g,\ i \geq I \end{aligned} \tag{1.1.3}$$

and that the whole incidence matrix $(x_{ij})$ is determined by (1.1.3) together with the restrictions

$$x_{ij} = x_{ji}, \quad x_{ii} = 0.$$

A graphical representation of monotone graphs can be given in the following way. A grid of $g \times g$ squares, both dimensions having coordinate values 1,2,...,g, will be called the (i,j) grid. In square (i,j) put the number $x_{ij}$. Then the highest 1 in column i is situated at $j = x_i^*$ where

$$x_i^* = \begin{cases} x_i + 1 & 1 \leq i \leq I-1 \\ x_i & I \leq i \leq g \end{cases} .$$

The borderline between the regions $\{x_{ij} = 1\}$ and $\{x_{ij} = 0\}$, not taking into account the main diagonal, is called the "incidence borderline". Figure 1 gives the incidence borderline for the monotone graph with $g = 11$ and

$$x_1 = x_2 = 8, \qquad x_3 = 7,\ x_4 = x_5 = 4$$

$$x_6 = x_7 = x_8 = 3,\ x_9 = 2,\ x_{10} = x_{11} = 0.$$

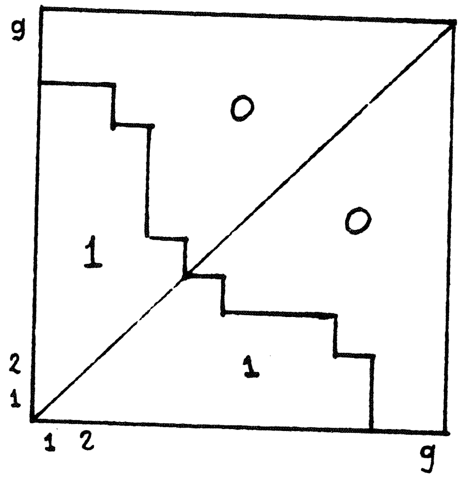


*Figure 1.*

The incidence borderline completely characterizes the monotone graph. A less redundant characterization is given by the "half incidence borderline", which is the part of the incidence borderline below the main diagonal, i.e. for $i \geq I$ only; see Figure 2.

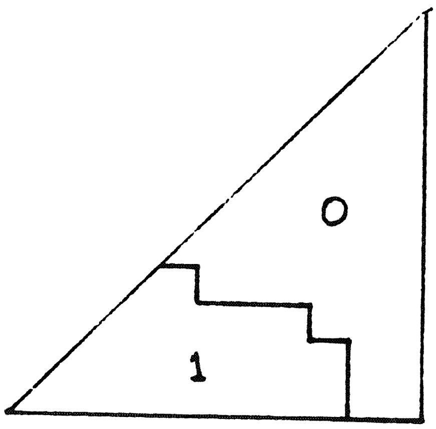


*Figure 2.*

### *1.2. Monotone graphs and majorization*

(This section will not be used in the sequel and may be skipped).

For two vectors $x = (x_1,\ldots,x_g)$ and $y = (y_1,\ldots,y_g)$, x is said to majorize y if

$$\Sigma_{i=1}^{g} x_i = \Sigma_{i=1}^{g} y_i$$

$$\Sigma_{i=1}^{k} x_{[i]} \geq \Sigma_{i=1}^{k} y_{[i]} \quad \text{for } k = 1,\ldots,g-1$$

where $x_{[1]},\ldots,x_{[g]}$ are the $x_i$'s arranged in descending order. See Marshall and Olkin (1979). This is equivalent to

$$\Sigma_{i=1}^{g} \varphi(x_i) \geq \Sigma_{i=1}^{g} \varphi(y_i) \text{ for all convex } \varphi,$$

and it is denoted by $x \succ y$. In the discussions at the start of Section 1.1 we have

$$(x_1', x_2', \ldots, x_g') \succ (x_1, \ldots, x_g).$$

The ordering of majorization is a partial ordering, which reflects the dispersion of the components of the vectors.

Denote

$$\mathbb{D} = \{(x_1,\ldots,x_g) \mid \text{there is an undirected graph with degrees } x_1, x_2, \ldots, x_g\}.$$

Then the result of the previous section can be formulated by saying that the class of maximal elements of $\mathbb{D}$ with respect to majorization is a subset of the class of vectors of degrees of monotone graphs. Corollary 6.18 of Chen (1971) states that if $x = (x_1,\ldots,x_g)$ is a vector of integers with $g-1 \geq x_1 \geq x_2 \geq x_3 \geq \cdots \geq x_g \geq 0$, then $x \in \mathbb{D}$ if and only if

(1.2.1)
$$\Sigma_{i=1}^{g} x_i \text{ is even}$$
$$\tilde{x} \succ x$$

where $\tilde{x} = (\tilde{x}_1,\ldots,\tilde{x}_g)$ is defined by

$$\tilde{x}_i = |\{j \neq i \mid x_j \geq i\}| \quad \text{if } 1 \leq i \leq i_o$$

$$\tilde{x}_i = |\{j \leq i_o \mid x_j + 1 \geq i\}| \quad \text{if } i_o + 1 \leq i \leq g,$$

where

$$i_o = \max\{i \mid x_i \geq i\} = I-1.$$

For monotone graphs (with degrees arranged in non-decreasing order) we precisely have $\tilde{x} = x$. On the other hand it is not hard to see that if $\tilde{x} = x$, then a (symmetric) incidence matrix of a monotone graph can be constructed with vector of degrees $x$ (compare the construction mentioned by Chen (1971) on page 411). So $x$ is the vector of degrees of a monotone graph if and only if $\tilde{x} = x$.

It can be proved that

(1.2.2) $$x \prec y \Rightarrow \tilde{x} \succ \tilde{y}.$$

(Compare 7.B.5 in Marshall and Olkin (1979), where the analogous result is proved for directed graphs.) With (1.2.1) this implies that if x is the vector of degrees of a monotone graph, then x is a maximal element of $\mathbb{D}$ with respect to majorization. Summarizing, we have obtained the following result.

*Theorem.* The maximal elements of $\mathbb{D}$ with respect to majorization are exactly the vectors of degrees of monotone graphs.

## 2. *The maximum value of the degree variance for a given density*

### 2.1. *Statement of the result*

In this section, the maximum value of V for undirected graphs with g points and degree sum

$$s = \Sigma_{i=1}^{g} x_i$$

will be obtained. The condition of a given degree sum is of course equivalent to the condition of a given density.

The monotone graph with

(2.1.1) $$x_i = \begin{cases} I & 1 \le i \le k \\ I-1 & k+1 \le i \le I \\ k & i = I+1 \\ 0 & I+2 \le i \le g \end{cases}$$

or, equivalently,

$$\begin{array}{lll} x_{ij} = 1 & 1 \le i,j \le I & i \ne j \\ x_{I+1,i} = x_{i,I+1} = 1 & 1 \le i \le k & \\ x_{ij} = 0 & \text{all other } (i,j) & \end{array}$$

where $0 \le k \le I-1$ and $1 \le I \le g-1$ or $(I,k) = (g,0)$, is denoted by $G(I,k)$. The incidence borderline of this graph is represented in Figure 3. The degree sum of $G(I,k)$ is

$$\begin{aligned} s(I,k) &= kI + (I-k)(I-1) + k \\ &= I(I-1) + 2k. \end{aligned}$$

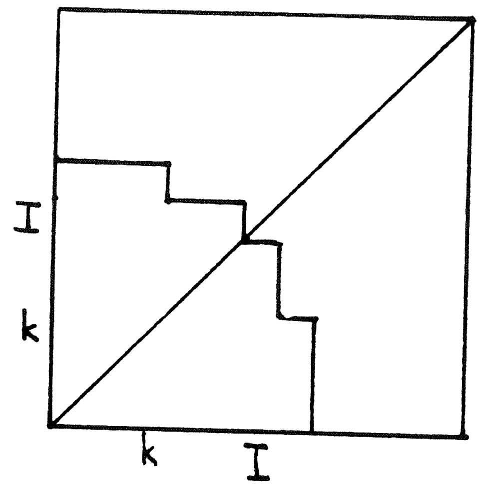


*Figure 3.*

For the mentioned values of (I,k), s(I,k) assumes all even integers from 0 to $g(g-1)$, which are all possible values for s for undirected graphs with $g$ points. The definition of G(I,k) can also be used for k = I, in which case one gets G(I,I) = G(I+1,0). The degree sum of squares of G(I,k) is

$$kI^2 + (I-k)(I-1)^2 + k^2 = I(I-1)^2 + k(2I+k-1).$$

whence the degree variance of G(I,k) is

(2.1.2) $$V(I,k) = \frac{1}{g}\{I(I-1)^2 + k(2I+k-1)\} - \{\frac{I(I-1) + 2k}{g}\}^2.$$

It will be proved that

(2.1.3) $$V_{max}(g,d) = \max V(I_d,k_d),\ V(I_d^c,\ k_d^c)$$

where $(I_d,k_d)$ and $(I_d^c,k_d^c)$ are the unique solutions of

(2.1.4)
$$I_d(I_d-1) + 2k_d = g(g-1)d$$
$$I_d^c(I_d^c-1) + 2k_d^c = g(g-1)(1-d)$$

when $(I_d,k_d)$ and $(I_d^c,k_d^c)$ are restricted to

(2.1.5) $$J = \{(I,k)\mid 0 \le k \le I-1,\ 1 \le I \le g-1\} \cup \{(g,0)\}.$$

It may be noted that the complement of $G(I_d^c,\ k_d^c)$ is a graph with degree sum s and variance equal to $V(I_d^c,k_d^c)$.

This result will be proved in the following way. Denote by $G(d)$ the set of all monotone graphs with g points and density d, and define

$$G_1(d) = \{G\in G(d)\mid G \text{ has no isolated points}\}$$

$$G_o(d) = \{G\in G(d)\mid G \text{ has at least one isolated point}\}.$$

For every $G \in \mathcal{G}_0(d)$, its complement is an element of $\mathcal{G}_1(1-d)$ and has the same degree variance as G, and vice versa. It well be demonstrated that for every $G \in \mathcal{G}_0(d)$, either the variance of G is equal to $V(I_d, k_d)$, or there exists a graph in $\mathcal{G}(d)$ with a greater degree variance than G. This implies (2.1.3).

## 2.2. *Changing a monotone graph.*

Let G be a monotone graph.

Define $z_1, z_2, \ldots, z_f$ as the different values assumed by the degrees $x_i$ for $i \geq I$, ordered so that $z_1 > z_2 > \ldots > z_f \geq 0$.

Define $i_h$ as the largest i with $x_i = z_h$. Then $I \leq i_1 < \ldots < i_f = g$, and $(i_h, z_h)$ are the coordinates of the right-upper angles in the half incidence borderline. Now change G into G' by disconnecting the points $i_h$ and $z_h$, while connecting the points $i_e + 1$ and $z_{e+1}+1$. Then G' is again a monotone graph with the same density as G; the degree sum of squares of G' minus that of G is (see Figure 4)

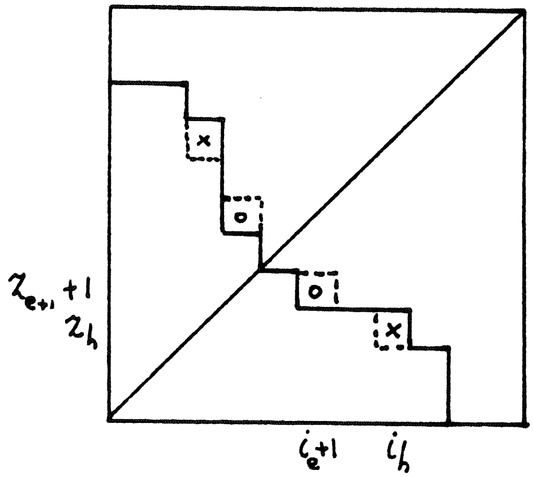


*Figure 4.*

$$(z_{e+1}+1)^2 + i_e^2 - \{z_{e+1}^2 + (i_e-1)^2\}$$

$$+ (z_h-1)^2 + (i_h-2)^2 - \{z_h^2 + (i_h-1)^2\}$$

(2.2.1) $$= 2(z_{e+1}+1 + i_e+1) - 2(i_h+z_h).$$

In the graphical representation, (2.2.1) is the sum of the coordinates of the squares in the (i,j) grid where 0 is replaced by 1, minus the coordinate sum of the squares where 1 is replaced by 0. Because of the symmetry with respect to the main diagonal we only need to consider the part of the (i,j) grid below the main diagonal.

In the following subsections, we consider changes of monotone graphs consisting of a succession of changes of the type above. In the (i,j) grid, disconnected pairs are marked by × and connected pairs by ∘. It can be easily verified graphically, that the proposed change of G leaves the degree variance unaffected or increases it. If the ∘ and × signs can be so paired that in each pair (consisting of an ∘ and an ×), both squares have the same coordinate sum i+j, then the degree variance is unaffected. If in each pair the coordinate sum of ∘ is not less than the coordinate sum of ×, while it is greater in some pairs, then the degree variance is increased. For example, in Figure 4 the degree variance is decreased.

The cases $f \geq 4$, $f = 3$, $f = 2$ and $f = 1$ will be studied successively. We denote the subclass of $G_o(d)$ with a given value of f by $G_o(d,f)$ and

$$y_h = i_{h+1} - i_h \qquad 1 \leq h \leq f-1$$

$$y_o = i_1 - I$$

$$d_h = z_h - z_{h+1} \qquad 1 \leq h \leq f-1.$$

Then $y_h \geq 1$, $d_h \geq 1$ for $1 \leq h \leq f-1$ and $y_o \geq 0$. For G with at least one isolated point, we have $x_g = z_f = 0$.

## 2.3. *The case $f \geq 4$*

It will be proved in this subsection that for every graph in $G_o(d,f)$ with $f \geq 4$, there exists a graph in $G(d)$ with a larger degree variance. Consider a graph in $G_o(d,f)$ with $f \geq 4$. For $h = 1,2$, consider the following set of changes.

**A.** Suppose $d_{h+1} \leq d_h \leq y_h$. Disconnect $j$— $i_h$ for $z_h - d_{h+1} + 1 \leq j \leq z_h$ and connect $j$— $(i_{h+1}+1)$ for $z_{h+2} + 1 \leq j \leq z_{h+1}$. (Figure 5). This increases the degree variance.

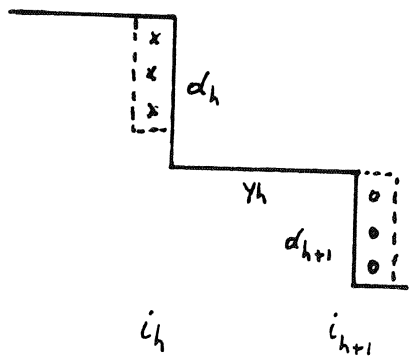


*Figure 5.*

B. Suppose $d_h \le d_{h+1} \le y_h$. Disconnect $j$ —— $i_h$ for $z_{h+1} + 1 \le j \le z_h$ and connect $j$ —— $(i_{h+1}+1)$ for $z_{h+2} + 1 \le j \le z_{h+2} + d_h$. (Figure 6). This increases the degree variance.

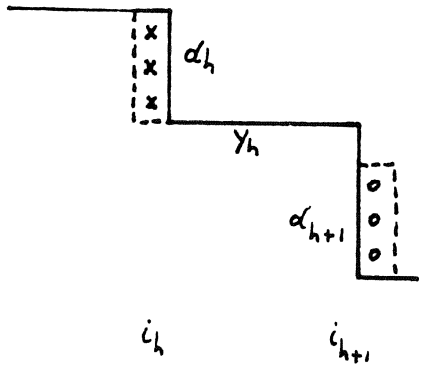


*Figure 6.*

C. Suppose $d_{h+1} \ge 2$ and $2 \le y_h \le d_h$. Disconnect $z_{h+1}$ —— $i$ for $i_h + 2 \le i \le i_{h+1}$, disconnect $(z_{h+1}-1)$ —— $i_{h+1}$ and connect $j$ —— $(i_h+1)$ for $z_{h+1}+1 \le j \le z_{h+1}+y_h$. (Figure 7). This increases the degree variance.

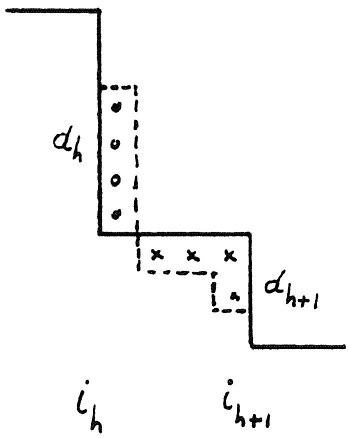


*Figure 7.*

D. Suppose $d_h \ge 2$ and $2 \le y_h \le d_{h+1}$. Disconnect $j$ —— $i_{h+1}$ for $z_{h+1} - y_h + 1 \le j \le z_{h+1}$, connect $(z_{h+1}+1)$—— $i$ for $i_h + 1 \le i \le i_{h+1}-1$ and connect $(z_{h+1}+2)$ —— $(i_h+1)$. (Figure 8). This increases the degree variance.

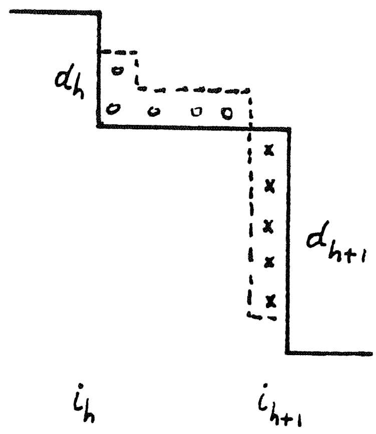


*Figure 8.*

If none of the conditions for changes A, B, C or D is satisfied, then necessarily

(2.3.1) $y_h = 1$ or ($d_h = 1$ and $y_h < d_{h+1}$) or ($d_{h+1} = 1$ and $y_h < d_h$).

By applying this for h = 1, it is seen that a necessary condition for a graph in $G_o(d,f)$ to have the maximum degree variance is

$$y_1 = 1 \text{ or } (d_1 = 1 \text{ and } y_1 < d_2) \text{ or } (d_2 = 1 \text{ and } y_1 < d_1).$$

If $d_1 = 1$ and $y_1 < d_2$, then disconnecting j —— $i_2$ for $z_2 - y_1 + 2 \leq j \leq z_2$ and connecting $z_1$ —— i for $i_1 + 1 \leq i \leq i_2 - 1$ yields a graph in $G_o(d,f)$ with the same degree variance and $y_1 = 1$. (Figure 9).

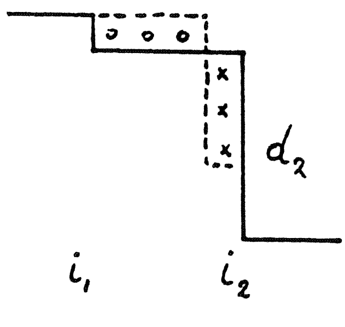


*Figure 9.*

If $d_2 = 1$ and $y_1 < d_1$, then disconnecting $z_2$ —— i for $i_1 + 2 \leq i \leq i_2$ and connecting j —— $(i_1+1)$ for $z_2 + 1 \leq j \leq z_2 + y_1 - 1$ yields a graph in $G_o(d,f)$ with the same degree variance and $y_1 = 1$. (Figure 10).

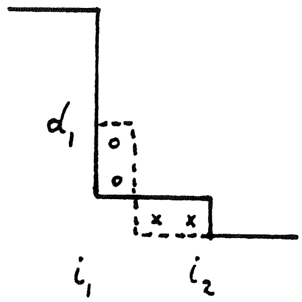


*Figure 10.*

This implies that in order to achieve the aim of this subsection it is sufficient to show that for every graph in $\mathcal{G}_o(d,f)$ with $f \geq 4$ and $y_1 = 1$, there exists a graph in $\mathcal{G}(d)$ with a greater degree variance. By considering the changes A, B, C and D for $h = 2$, it is seen that all these graphs which cannot be changed (with increasing degree variance) in one of these ways satisfy (cf. 2.3.1) $y_1 = 1$ and

$$y_2 = 1 \text{ or } (d_2 = 1 \text{ and } y_2 < d_3) \text{ or } (d_3 = 1 \text{ and } y_2 < d_2),$$

which implies

(2.3.2) $$y_1 = 1 \text{ and } (y_2 \leq d_2 \text{ or } (d_2 = 1 \text{ and } y_2 < d_3)).$$

If $y_1 = 1$ and $y_2 \leq d_2$, then disconnecting $z_3$ — $i_3$ and connecting $(z_2+1)$ — $i_2$ increases the degree variance. (Figure 11).

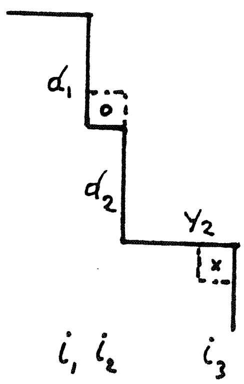


*Figure 11.*

If $y_1 = d_2 = 1$ and $y_2 \leq d_3$, then disconnecting $j$ — $i_3$ for $z_3 - y_2 + 1 \leq j \leq z_3$, connecting $z_2$ — $i$ for $i_2 + 1 \leq i \leq i_3-1$ and connecting $(z_2+1)$ — $i_2$ increases the degree variance. (Figure 12).

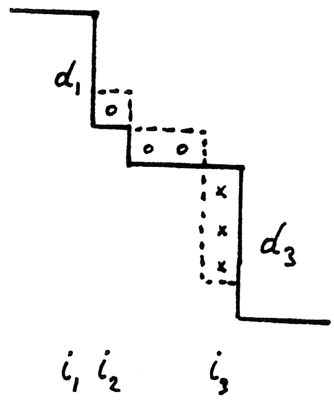


*Figure 12.*

This completes the demonstration that if $f \geq 4$, then the degree variance cannot be maximal.

## 2.4. *The case* $f = 3$.

It will be demonstrated in this subsection that if $1 \leq k_d \leq I-2$, then $G(I_d, k_d)$ has in $G_o(d,3)$ the maximal degree variance; while if $k_d = 0$ or $I-1$, then for every graph in $G_o(d,3)$ there is another graph in $G_o(d)$ with a greater degree variance.

All changes mentioned in Subsection 2.3, until and including the change represented by Figure 10, can also be applied if $f = 3$. This implies that the maximum degree variance for graphs in $G_o(d)$ with $f = 3$ is assumed by a graph satisfying $y_1 = 1$. (Figure 13).

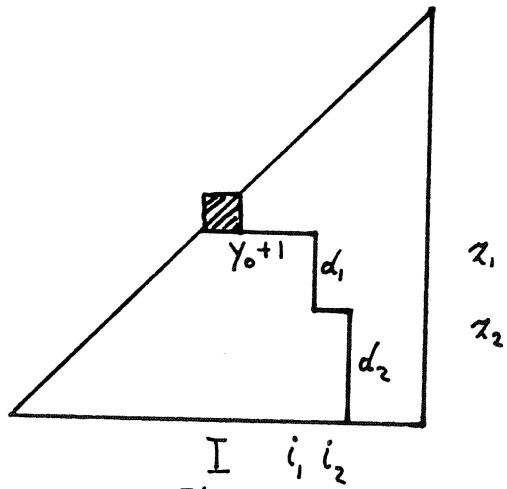


*Figure 13*

Consider a graph in $G_o(d)$ with $f = 3$ and $y_1 = 1$. If $1 \le y_o \le d_2$, then disconnecting $j$ — $i_2$ for $z_2 - y_o + 1 \le j \le z_2$ and connecting $(z_1+1)$ — $i$ for $I + 1 \le i \le i_1$ increases the degree variance. (Figure 14).

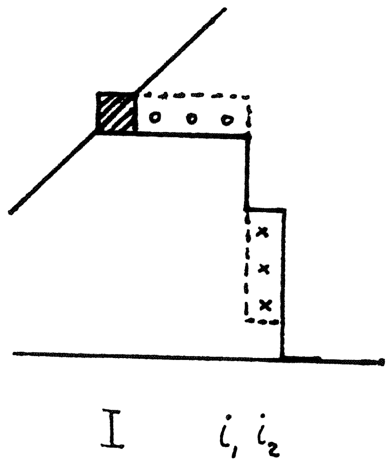


*Figure 14.*

If $d_2 \le y_o \le d_1 + d_2 - 1$, then disconnecting $j$ — $i_2$ for $1 \le j \le z_2$ and connecting $(z_1+1)$ — $i$ for $I + 1 \le i \le I + z_2$ increases the degree variance (Figure 15).

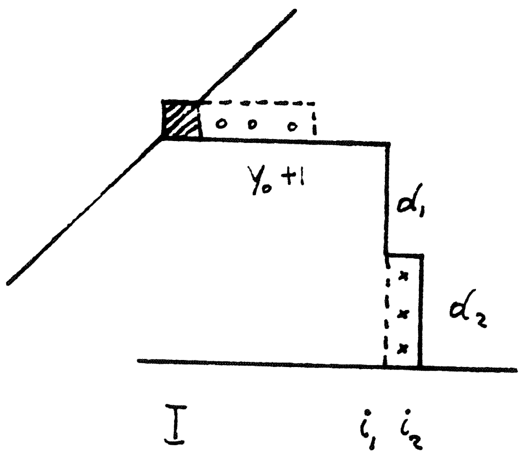


*Figure 15.*

If $y_o \ge d_1 + d_2$, then disconnecting $z_1$ — $i$ for $i_1 - d_1 - d_2 + 2 \le i \le i_1$, connecting $j$ — $i_2$ for $z_2 + 1 \le j \le z_1 - 1$ and connecting $j$ — $(i_2+1)$ for $1 \le j \le z_2$ increases the degree variance. (Figure 16).

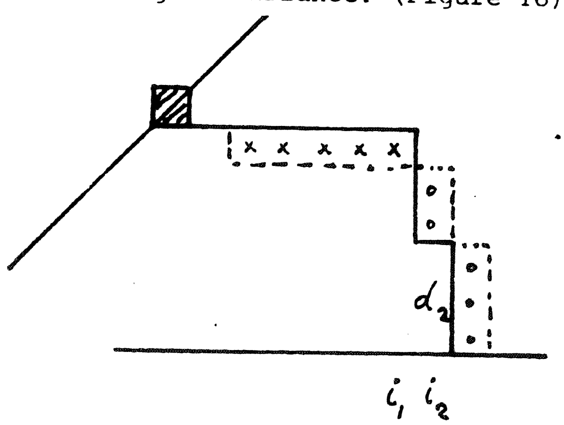


*Figure 16.*

The only case remaining of $f = 3$ is $y_o = 0$, $y_1 = 1$. This is just the graph $G(I_d, k_d)$, if $1 \le k_d \le I-2$; if $k_d = 0$ or $I-1$, then there is no graph in $G_o(d)$ with $y_o = 0$, $y_1 = 1$.

## 2.5. *The case $f = 2$.*

It will be demonstrated in this subsection that for every graph in $G_o(d,2)$, either there is a graph in $G(d)$ with a greater degree variance, or the degree variance is equal to $V(I_d, k_d)$ or $V(I_d^c, k_d^c)$.

For $z_1 = g - 2$, we encounter $G(g-1,0)$; for $z_1 = g-3$ and $f = 2$, we encounter $G(g-2,0)$ or $G(g-2,g-3)$. Now consider a graph in $G_o(d,2)$ with $z_1 \le g-4$. If $3 \le y_o \le d_1$, then disconnecting $j$ — $i_1$ for $z_1 - y_o + 1 \le j \le z_1$, connecting $(z_1+1)$ — $i$ for $I + 1 \le i \le i_1 - 1$ and connecting $(z_1+2)$ — $(I+2)$ increases the degree variance. (Figure 17).

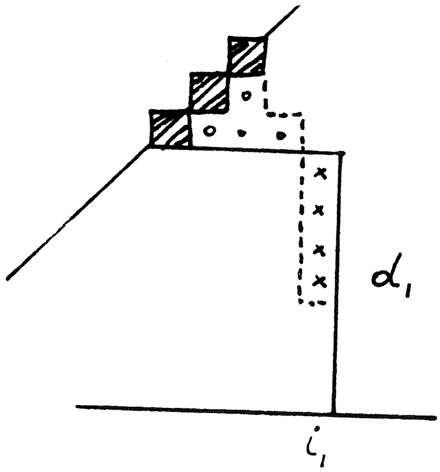


*Figure 17.*

If $y_1 \ge 2$, $d_1 \le y_o$, then disconnecting $z_1$ — $i$ for $i_1 - d_1+1 \le i \le i_1$, connecting $j$ — $(i_1+1)$ for $1 \le j \le z_1 - 1$ and connecting $1$ — $(i_1+2)$ increases the degree variance (Figure 18).

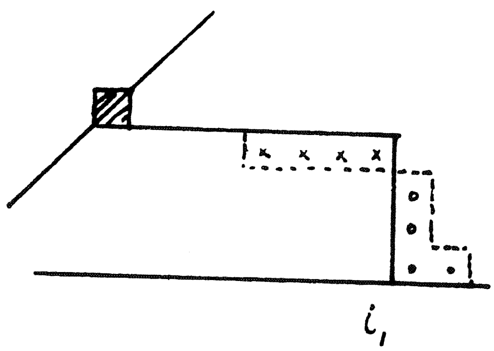


*Figure 18.*

If $y_1 = 1$, $d_1 \leq y_o$, then disconnecting $z_1$ — $i$ for $i_1 - d_1 + 2 \leq i \leq i_1$ and connecting $(i_1+1)$ — $j$ for $1 \leq j \leq z_1 - 1$ (note that $i_1 = g-1$) yields a graph with the same degree variance. This graph is just the complement of $G(I_d^c, k_d^c)$, and hence its degree variance is $V(I_d^c, k_d^c)$. (Figure 19).

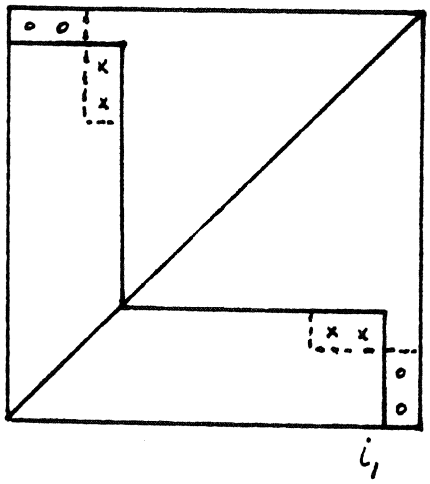


*Figure 19.*

If $y_o = 2$, then disconnecting $z_1$ — $i_1$ and connecting $(z_1+1)$ — $(i_1+1)$ does not affect the degree variance and yields the graph $V(I_d, k_d)$. (Figure 20).

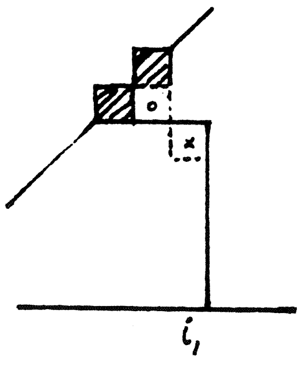


*Figure 20.*

For $y_o = 1$, we encounter $G(I_d, I_d-1)$. For $y_o = 0$, we encounter $G(I_d, 0)$.

*2.6. The case $f = 1$.*

If $f = 1$ for a graph with at least one isolated point, we have $x_I = x_g = 0$, which implies that the graph is completely disconnected and is equal to $G(0,0)$.

## 3. *The maximum value of the degree variance.*

It follows from Section 2 (cf. (2.1.3)) that

$$V_{max}(g) = \max_{0 \leq s \leq g(g-1)} V(I_d, k_d)$$

with $(I_d, k_d)$ determined by (2.1.4), (2.1.5). Formula (2.1.2) shows that, for fixed I, V(I,k) is a quadratic function of k with a nonnegative coefficient of $k^2$ whenever $g \geq 4$. As G(I,I) = G(I+1,0), this implies that

$$\max_{0 \leq k \leq I-1} V(I,k) \leq \max\{V(I,0),\ V(I+1,0)\}$$

and hence

$$V_{max}(g) = \max_{0 \leq I \leq g-1} V(I,0).$$

Define

$$v(I) = g^2 V(I,0) = I(I-1)^2(g-I).$$

Then

$$v(I-1) - v(I) = (I-1)(4I^2-(3g+8)I+4g).$$

This expression is 0 for I = 1 and

$$I = \frac{3g + 8 \pm \sqrt{9g^2 - 16g + 64}}{8}.$$

For $g \geq 5$ we have $3g - 8 < \sqrt{9g^2 - 16g + 64} < 3g$, so there is a $y \in (0,1)$ with

$$2 \leq I \leq \frac{3}{4}g + y \Rightarrow v(I-1) < v(I)$$

$$\frac{3}{4}g + y \leq I \leq g \Rightarrow v(I-1) > v(I).$$

This shows that for integer I, v(I) is maximal for I = I* with $\frac{3}{4}g - 1 < I^* < \frac{3}{4}g + 1$. Considering each of the cases g = 4h + t for integers h and t with $0 \leq t \leq 3$ shows that in all these cases, I* can be represented by

(3.1) $$I^* = [\tfrac{3}{4}g + \tfrac{1}{2}].$$

For g = 3,4 the number of possibilities is very small and it is easily seen that here, too, $V_{max}(g) = V(I^*,0)$ with I* as in (3.1). So we have obtained the result

(3.2) $$V_{max}(g) = V(I^*,0) = \frac{I^*(g-I^*)(I^*-1)^2}{g^2} .$$

It is straightforward to see that $g^{-2}V_{max}(g)$ is an increasing function of g which converges to

$$\lim_{g\to\infty} V_{max}(g) = \frac{27}{256} .$$

## 4. *Expectation and variance of the degree variance under some simple stochastic models.*

This section contains the computations for EV and var V under the null models of DV (Section 3).

### 4.1. *A simple general formula.*

In all models considered, the restriction

$$\Sigma_{i=1}^{g} X_i = s$$

is made, while $X_1, X_2 \ldots, X_g$ are permutation invariant. This yields

$$V = g^{-1}\Sigma_{i=1}^{g}(X_i - \frac{s}{g})^2 = g^{-1}\Sigma_{i=1}^{g} X_i^2 - (\frac{s}{g})^2$$

(4.1.1)
$$EV = \text{var } X_1$$
$$\text{var } V = g^{-1} \text{ var } X_1^2 + g^{-1}(g-1) \text{ cov } (X_1^2, X_2^2).$$

### 4.2. *A general formula for directed and undirected graphs.*

For directed and undirected graphs, we have

$$X_i = \Sigma_{\substack{j=1 \\ j\neq i}}^{g} X_{ij}.$$

Define

(4.2.1)
$$\pi_1 = P\{X_{12} = 1\}$$
$$\pi_2 = P\{X_{12} = X_{13} = 1\}$$
$$\pi_3 = P\{X_{12} = X_{13} = X_{14} = 1\}$$
$$\pi_4 = P\{X_{12} = X_{13} = X_{14} = X_{15} = 1\}.$$

The permutation invariance of $X_{12}, X_{13}, \ldots, X_{1g}$ yields

(4.2.2)
$$\text{var } X_1 = (g-1) \text{ var } X_{12} + (g-1)(g-2) \text{ cov } (X_{12}, X_{13})$$
$$= (g-1)\pi_1(1-\pi_1) + (g-1)(g-2)(\pi_2 - \pi_1^2)$$
$$EX_1^2 = \text{var } X_1 + (EX_1)^2 = (g-1)\pi_1 + (g-1)(g-2)\pi_2$$
$$EX_1^4 = E(\Sigma_{j=2}^{g} X_{1j})^4 = (g-1)EX_{12}^2 + 4(g-1)(g-2)EX_{12}^3 X_{13} +$$
$$+ 3(g-1)(g-2)EX_{12}^2 X_{13}^2 + 6(g-1)(g-2)(g-3)EX_{12}^2 X_{13} X_{14} +$$

$$+ (g-1)(g-2)(g-3)(g-4)EX_{12}X_{13}X_{14}X_{15}$$

$$= (g-1)\pi_1 + 7(g-1)(g-2)\pi_2 + 6(g-1)(g-2)(g-3)\pi_3 + (g-1)(g-2)(g-3)(g-4)\pi_4.$$

With

$$\rho_2 = P\{X_{12} = X_{21} = 1\}$$
$$\rho_3 = P\{X_{12} = X_{21} = X_{13} = 1\}$$
$$\rho_4 = P\{X_{12} = X_{21} = X_{13} = X_{23} = 1\},$$

the permutation invariance of $X_{13}, X_{14}, \ldots, X_{1g}, X_{23}, X_{24}, \ldots, X_{2g}$ yields

$$EX_1^2X_2^2 = E(X_{12}+\Sigma_{j=3}^g X_{1j})^2(X_{21}+\Sigma_{j=3}^g X_{2j})^2$$

$$= EX_{12}^2X_{21}^2 + 4(g-2)EX_{12}^2X_{21}X_{23} + 2EX_{12}^2(\Sigma_{j=3}^g X_{2j})^2 + 4(g-2)^2EX_{12}X_{13}X_{21}X_{23} + 4(g-2)EX_{12}X_{13}(\Sigma_{j=3}^g X_{2j})^2 + E(\Sigma_{j=3}^g X_{ij})^2)(\Sigma_{j=3}^g X_{2j})^2$$

$$= \rho_2 + 4(g-2)\rho_3 + 4(g-2)^2\rho_4 + g(g-2)\pi_2 + 2(g-2)(g^2-2g-1)\pi_3 + (g+1)(g-2)^2(g-3)\pi_4.$$

Hence, with (4.1.1),

$$\text{var } V = g^{-1}EX_1^4 + g^{-1}(g-1)EX_1^2X_2^2 - (EX_1^2)^2$$

$$= g^{-1}(g-1)\{\rho_2+4(g-2)\rho_3 + 4(g-2)^2\rho_4\}$$

(4.2.3)
$$+ g^{-1}(g-1)\{\pi_1+(g+7)(g-2)\pi_2 + 2(g-2)(g^2+g-10)\pi_3 + (g-2)(g-3)(g^2-6)\pi_4\}$$

$$- \{(g-1)\pi_1 + (g-1)(g-2)\pi_2\}^2.$$

## 4.3. *A general formula for bipartite graphs.*

For bipartite graphs, we have

$$X_i = \Sigma_{j=1}^{h} X_{ij}.$$

A comparison with Section 4.2 shows that

(4.3.1)
$$\text{var } X_1 = h\pi_1(1-\pi_1) + h(h-1)(\pi_2-\pi_1^2)$$
$$EX_1^2 = h\pi_1 + h(h-1)\pi_2$$
$$EX_1^4 = h\pi_1 + 7h(h-1)\pi_2 + 6h(h-1)(h-2)\pi_3 + h(h-1)(h-2)(h-3)\pi_4$$
$$EX_1^2X_2^2 = h^2\pi_2 + 2h^2(h-1)\pi_3 + h^2(h-1)^2\pi_4,$$

where $\pi_1$ to $\pi_4$ are as in (4.2.1). Hence, with (4.1.1) again,

(4.3.2)
$$\begin{aligned}\text{var } V &= g^{-1}EX_1^4 + g^{-1}(g-1)EX_1^2X_2^2 - (EX_1^2)^2 \\ &= g^{-1}\{h\pi_1 + h(gh+6h-7)\pi_2 + 2h(h-1)(gh+2h-6)\pi_3 \\ &\qquad + h(h-1)(gh(h-1)-4h+6)\pi_4\} \\ &\quad - \{h\pi_1 + h(h-1)\pi_2\}^2.\end{aligned}$$

## 4.4. *Undirected graph.*

For undirected graphs we have $X_{ij} = X_{ji}$, and hence $\rho_i = \pi_{i-1}$. Further,

$$\pi_1 = \frac{s}{g(g-1)}$$
$$\pi_2 = \frac{s(s-2)}{(g+1)g(g-1)(g-2)}$$
$$\pi_3 = \frac{s(s-2)(s-4)}{(g+1)g(g-1)(g-2)(g^2-g-4)}$$
$$\pi_4 = \frac{s(s-2)(s-4)(s-6)}{(g+1)g(g-1)(g-2)(g-3)(g+2)(g^2-g-4)}.$$

With (4.2.2) and (4.2.3), lengthy computations yield

$$E\{V|s\} = \frac{s(g^2-g-s)}{g^2(g+1)}$$

$$\text{var}\{V|s\} = \frac{2s(s-2)(g^2-g-s)(g^2-g-s-2)}{g^2(g+1)^2(g+2)(g^2-g-4)}.$$

## 4.5. *Directed graph: U|man distribution*

For directed graphs with the U|man distribution, we have

$$\pi_1 = \frac{2m + a}{g(g-1)}$$

$$\pi_2 = \frac{4m(m-1) + 4ma + a(a-1)}{(g+1)g(g-1)(g-2)}$$

$$\pi_3 = \frac{8m(m-1)(m-2) + 12m(m-1)a + 6ma(a-1) + a(a-1)(a-2)}{(g+1)g(g-2)(g^2-g-4)}$$

$$\pi_4 = \frac{16m(m-1)(m-2)(m-3) + 32m(m-1)(m-2)a + 24m(m-1)a(a-1) + 8ma(a-1)(a-2) + a(a-1)(a-2)(a-3)}{(g+2)(g+1)g(g-1)(g-2)(g-3)(g^2-g-4)}$$

$$\rho_2 = \frac{2m}{g(g-1)}$$

$$\rho_3 = \frac{4m(m-1) + 2ma}{(g+1)g(g-1)(g-2)}$$

$$\rho_4 = \frac{8m(m-1)(m-2) + 8m(m-1)a + 2ma(a-1)}{(g+1)g(g-1)(g-2)(g^2-g-4)} .$$

With (4.2.2) and (4.2.3), lengthy computations yield

$$E\{V \mid m\,a\,n\} = \frac{(2m+a)(2n+a) + ga}{2g^2(g+1)}$$

$$\mathrm{var}\{V \mid m\,a\,n\} = \frac{\text{numerator}}{g^2(g+1)^2(g+2)(g^2-g-4)}$$

$$\text{numerator} = 2\{4m(m-1) + 4am + a(a-1)\}\{4n(n-1) + 4an + a(a-1)\} + 8(g+1)\,a\{4mn-a(a-1)\} + 2a(a-1)(g+1)(3g^2-4g-2).$$

It may be noted that the symmetry considerations for complementary graphs (DV, beginning of Section 2) imply that the formulas for $E\{V \mid m\,a\,n\}$ and $\mathrm{var}\{V \mid m\,a\,n\}$ must be symmetric in m and n.

It can be easily verified that for $m = \frac{1}{2}s$, $a = 0$, $n = \frac{1}{2}g(g-1)-m$, the formulas above are identical to those for undirected graphs with the $U|s$ distribution.

The identity

$$m + a + n = \tfrac{1}{2}g(g-1)$$

implies that the formulas above can be given many different expressions. I doubt whether other expressions are possible which are considerably simpler than those given here.

## 4.6. *Bipartite graph and directed graph with $U|s$ distribution.*

For bipartite graphs and directed graphs with the $U|s$ distribution, we have $\rho_i = \pi_i$; taking $h = g-1$ in (4.3.1) and (4.3.2) yields, in this case, formulas (4.2.2) and (4.2.3). Hence it is sufficient to consider only the case of bipartite graphs. We have

$$\pi_1 = \frac{s}{gh}$$

$$\pi_2 = \frac{s(s-1)}{gh(gh-1)}$$

$$\pi_3 = \frac{s(s-1)(s-2)}{gh(gh-1)(gh-2)}$$

$$\pi_4 = \frac{s(s-1)(s-2)(s-3)}{gh(gh-1)(gh-2)(gh-3)} .$$

With (4.3.1) and (4.3.2), lengthy computations yield

$$E\{V|s\} = \frac{s(gh-s)(g-1)}{g^2(gh-1)}$$

$$\mathrm{var}\{V|s\} = \frac{2(g-1)h(h-1)s(s-1)(gh-s)(gh-s-1)}{g^2(gh-1)^2(gh-2)(gh-3)} .$$

## 4.7. *Some general comments.*

* In the formulas for the means and variances of V, all factors in the numerator can be easily interpreted: if one of them is equal to zero, it is (by direct arguments, not from the formula) easy to see that the distribution of V is degenerated (in 0 if EV = 0, in one point if var V = 0).
* In the $U|s$ distribution, the marginal distribution of $X_i$ is hypergeometric; for the bipartite and directed cases, the simultaneous distribution of $(X_1,\ldots,X_g)$ is then even multivariate hypergeometric. The formulas for $E\{V|s\}$ are equal to the well-known formula for the variance of the hypergeometric distribution (with the correct parameter values).
* Several approaches are possible to compute the formulas obtained here. Those of Section 4.5 can be obtained from the theorems in Holland and Leinhard (1975), but with at least as much computational effort as was needed in our approach. Those from Section 4.6 can be obtained from the moments of the (multivariate) hypergeometric distribution, but, again, wit no less computational effort than was needed here. The approach followed was chosen because it admits a unified treatment for all cases considered in DV. I wonder whether, for the $U|s$ case, the simple (after all the

intermediate results) formulas for var $\{V|s\}$ can be obtained by simpler methods than those used here.

*References.*

Chen, W.K. (1971), Applied graph theory. North Holland Publishing Company, Amsterdam.

Holland, P.W. and S. Leinhardt (1975), Local structure in social networks. In D.R. Heise (ed.), Sociological Methodology -1976: 1-45. Jossey-Bass, San Francisco.

Marshall, A.W. and I. Olkin (1979), Inequalities: Theory of majorization and its applications. Academic Press, New York.

Snijders, T.A.B. (1981), The degree variance: an index of graph heterogeneity. To be published in Social Networks.