Social Networks, 3 (1981) 163–174

The Degree Variance: An Index of Graph Heterogeneity*

Tom A. B. Snijders
*University of Groningen***

In the analysis of empirically found graphs, the variance of the degrees can be used as a measure for the heterogeneity of (the points in) the graph. For several types of graphs, the maximum value of the degree variance is given, and the mean and variance of the degree variance under a simple stochastic null model are computed. These are used to produce normalized versions of the degree variance, which can be used as heterogeneity indices of graphs.

Key words: graph heterogeneity, graph centrality, random graphs, degree variance.

1. Introduction and notation

This paper is concerned with undirected, directed and bipartite graphs, without multiple lines or loops. The points of the graph are labeled 1, 2, ..., g for undirected and directed graphs; for bipartite graphs, the points of the first set are labeled 1, 2, ..., g while those of the second set are labeled 1, 2, ..., h. The incidence matrix is (x_{ij}); its element x_{ij} is 1 if there is a line from point i to point j, and 0 otherwise. Thus $x_{ij} = x_{ji}$ for undirected graphs and $x_{ii} = 0$ for undirected and directed graphs. For bipartite graphs (x_{ij}) is taken to be a $g \times h$ matrix and x_{ij} refers to the existence of a line from point i in the first set to point j in the second set.

The degree of point i is denoted by x_i, and can be defined by

$$x_i = \sum_j x_{ij}$$

For directed graphs, this is the out-degree of point i; if one wishes to consider in-degrees instead, one needs only to reverse the directions of the lines (which amounts to transposition of (x_{ij})). For bipartite graphs, only the degrees of the points in the first set are studied. The degree sum is

$$s = \sum_i x_i$$

*I thank the other members of the Study Group on Stochastic Graphs (Social Science Department, University of Groningen), chaired by Frans N. Stokman, for their stimulating comments.
**Department of Econometrics, University of Groningen, G.P.O. 800, Groningen, The Netherlands.

The (line) density is $d = s/g(g - 1)$ for undirected and directed graphs, and $d = s/gh$ for bipartite graphs. In the sequel, when discussing directed graphs, h will be defined by $h = g - 1$ because there will be several formulas which turn out to be identical for bipartite and directed graphs when this identification is made.

The integer part of a real number y (that is, the largest integer not exceeding y) is denoted by $[y]$.

2. The degree variance

In the investigation of graphs produced from empirical observations, it is desirable to have one or more descriptive statistics (or structural parameters) giving a global impression of the magnitude of the differences between the points as regards their centrality in the graph, or, briefly said, of the heterogeneity of (the points in) the graph. Often, instead of this rather aspecific concept of heterogeneity, the more specific concept of centralization has been studied. Freeman (1978) describes centralization as "the tendency of a single point to be more central than all other points in the network". He stresses the difference between this concept of centralization and the concept of graph centrality in the sense of "all points being close together", and gives a review of centralization parameters. Høivik and Gleditsch (1975), in a review article about graph parameters, describe centralization as "the dispersion in the set of vertex centralities", which is quite in the spirit of the heterogeneity concept mentioned above; but they go on to mention centralization parameters which are all operationalizations of Freeman's description. This concept of centralization implies the idea of one clearly defined center, preferably consisting of one point. The pinnacle of centralization, for undirected graphs with a fixed number of points, is the star: the graph with lines between point 1 (say) and every other point, and no other lines. All centralization parameters mentioned by Freeman (1978) assume their maximum in the star.

In many empirically found graphs, however, there is a vaguely outlined center, consisting of more than one point; or there are several centers, or just a gradual transition from more central to more peripheral points. So it seems appropriate to have an index of heterogeneity, which keeps account of the differences in point centrality between all points, and which can be used in addition to, or instead of, an index of centralization (which keeps account of the differences in point centrality only between the single most central point and all other points). In this paper, attention will be restricted to graphs where the degrees of the points are important characteristics; graphs representing sociometric choices or interlocking directorates provide examples of this situation. A possible operationalization of the concept 'graph heterogeneity' is then the dispersion of the degrees, as measured by their variance:

$$V = g^{-1} \sum_i (x_i - x_.)^2 = g^{-1} \sum_i x_i^2 - x_.^2$$

where $x_. = s/g$ is the mean degree.

Use of the density and the degree variance as descriptive statistics for graphs is analogous to the use of the mean and the variance as descriptive statistics for sets of numbers. The degree variance is mentioned by Coleman (1964:Chap. 14.1) as an intermediate step in the construction of a measure of hierarchization (in a sociometric context, graph heterogeneity can be interpreted as hierarchization); Holland and Leinhardt (1979) consider probability distributions of directed graphs conditionally on the density and the variances of the in- and out-degrees; but the degree variance seems to have received little attention as a descriptive statistic for graphs in its own right.

Of course, many other measures for the dispersion of the degrees can be proposed. For example, if $\phi(x)$ is a convex non-decreasing function (for nonnegative integer x), then

$$V_\phi = g^{-1} \sum_i \phi(x_i) - \phi(x_.)$$

will be such a measure; taking $\phi(x) = x \log x$ in this expression yields the entropy-based measure of hierarchization which is proposed, after normalization, by Coleman (1964:Chap. 14.2). Practical reasons for choosing the variance (coinciding with V_ϕ for $\phi(x) = x^2$) are that it is an established and widely used measure of dispersion, and that its simplicity makes it possible to derive a number of properties of the degree variance, which would be harder or impossible to derive for many other measures of graph heterogeneity. This will be done in the next Sections; these properties can serve as conceptual arguments for or against use of the degree variance in certain empirical contexts. In § 6, several normalizations for V will be proposed. All along, it will be assumed that the degree variance is used complementarily to the density, and hence several properties of V will be derived conditionally on the value of the density d (or, equivalently, conditionally on s or $x_.$).

3. Maximum values

The maximum value for V, for the three kinds of graphs considered, and for fixed g, will be denoted by $V_{\max}(g)$. The maximum value for V when g and d are fixed will be denoted by $V_{\max}(g, d)$. For bipartite graphs it would be more precise also to indicate in the notation the dependence on h (which will be a fixed number), but this is avoided in order to get a uniform notation for directed and bipartite graphs.

It will be seen that for the graphs where the maximum values $V_{\max}(g)$ and $V_{\max}(g, d)$ are assumed, the set of points is clearly divided into a center consisting of points of high degree, and a periphery consisting of points of low degree; there is sometimes one intermediate point, needed to obtain the given density d. In contrast to the situation for the centralization indices discussed by Freeman, however, the center consists of more than one point, except for very low values of g or d; the size of the center of the graph where $V_{\max}(g)$ is assumed is roughly proportional to g. This is an important characteristic of the degree variance.

Two graphs are called complementary if one has lines exactly where the other does not; their incidence matrices (x_{ij}) and (x_{ij}^c) satisfy $x_{ij}^c = 1 - x_{ij}$ (for $i \neq j$, in the cases of directed and undirected graphs). For two complementary graphs, the two densities sum to 1 and the degree variances are equal. This implies that always

$$V_{\max}(g, d) = V_{\max}(g, 1 - d)$$

3.1. Directed and bipartite graphs

Recall that for directed graphs h is defined by $h = g - 1$. It will be proved that the maximum degree variance for given g and d is assumed by the graph with $[gd]$ points having degree h (these are connected with all other points), $[g(1 - d)]$ points having degree 0, and (unless gd is an integer) one point having degree $hgd - h[gd]$.

Consider any other graph with density d; for such a graph there are at least two points i and j with degrees satisfying $1 \leqslant x_i \leqslant x_j \leqslant h - 1$. In this graph one line can be removed from i and one added to j; this keeps d constant.

Using the expression

$$V = g^{-1} \sum_i x_i^2 - (hd)^2$$

it is seen that the increase in V is

$$g^{-1} \{(x_j + 1)^2 + (x_i - 1)^2 - x_j^2 - x_i^2\} = 2g^{-1}(x_j - x_i + 1)$$

This is a positive number, so the graph considered does not have the maximum possible V. It follows that the graph mentioned at the beginning has the maximum degree variance; the degree variance of this graph is

$$V_{\max}(g, d) = h^2 g^{-1} \{[gd] + (gd - [gd])^2 - gd^2\}$$

It can be proved that for bipartite graphs all graphs assuming this maximum are isomorphic; for directed graphs this is the case only for certain values of d (for instance, if gd is an integer).

The maximum degree variance for given g, if g is even, is assumed by the graph for which half of the points have degree h and the other half have degree 0; the maximum value is

$$V_{\max}(g) = h^2/4 \qquad (g \text{ even})$$

and all graphs assuming this maximum are isomorphic. If g is odd, then V is maximal if and only if either $(g - 1)/2$ or $(g + 1)/2$ of the points have degree h, while the other points have degree 0; the maximum value is

$$V_{max}(g) = (1 - g^{-2})h^2/4 \qquad (g \text{ odd})$$

and there are two classes of isomorphism for which this maximum is assumed. These results are easily verified by noting that for any other graph V can be increased by adding or removing a suitably chosen line.

It may be remarked that Freeman (1978) 'proves' that the centralization indices mentioned by him are (for the class of undirected graphs) maximal for the star by demonstrating that adding, removing or switching a line in the star yields a graph with a smaller value of the centralization index. This does not prove the desired results, because there are graphs which cannot be obtained by adding, removing or switching just one line in the star. The method used above, which is in a certain sense the reverse of Freeman's method, can be used to give the same results as those obtained by him.

3.2. Undirected graphs

For undirected graphs, the situation is much more complicated, and more interesting. In § 3.1, changing a line affected the (out-)degree of only one point, while for undirected graphs changing a line affects the degrees of two points simultaneously. The proofs of the results of this Section are rather lengthy, and the interested reader can find them in Snijders (1981).

We denote by $G(I, k)$ the graph with g points defined by

$$x_{ij} = 1 \qquad 1 \leqslant i, j \leqslant I, \qquad i \neq j$$

$$x_{I+1,\, i} = x_{i,\, I+1} = 1 \qquad 1 \leqslant i \leqslant k$$

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

where $1 \leqslant I \leqslant g$ and $0 \leqslant k \leqslant I - 1$ if $1 \leqslant I \leqslant g - 1$, while $k = 0$ if $I = g$. For this graph, the points $1, 2, \ldots, I$ can be conceived as the center, and the points $I + 2, I + 3, \ldots, g$ can be conceived as the periphery of the graph; the point $I + 1$ occupies an intermediate position, and can for certain values of k be considered as a member of the center or of the periphery. The complement of $G(I, k)$ is denoted by $G^c(I, k)$. For $G^c(I, k)$, the points $1, 2, \ldots, I$ are peripheral and the points $I + 2, I + 3, \ldots, g$ are central points. In the case $k = 0$, these graphs are particularly simple. Using the notation of Harary (1969), one has $G(I, 0) = K_I \cup \bar{K}_{g-I}$ and $G^c(I, 0) = K_{g-I} + \bar{K}_I$.

The density of $G(I, k)$ is

$$\{I(I - 1) + 2k\}/g(g - 1)$$

and the density of $G^c(I, k)$ is

$$1 - \{I(I - 1) + 2k\}/g(g - 1)$$

The degree variance of $G(I, k)$ and $G^c(I, k)$ is

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

Let I_d and I_d^c, respectively, be the largest integers with

$$I_d(I_d - 1) \leq g(g-1)d = s$$

$$I_d^c(I_d^c - 1) \leq g(g-1)(1-d) = g(g-1) - s$$

and let

$$k_d = \{g(g-1)d - I_d(I_d - 1)\}/2$$

$$k_d^c = \{g(g-1)(1-d) - I_d^c(I_d^c - 1)\}/2$$

Then $G(I_d, k_d)$ and $G^c(I_d^c, k_d^c)$ both have density d, and one of these has the maximum degree variance for density d:

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

The maximum degree variance for all undirected graphs with g points is assumed by $G(I^*, 0)$ and $G^c(I^*, 0)$ where

$$I^* = [(3g + 2)/4]$$

and this maximum is

$$V_{max}(g) = I^*(I^* - 1)^2 \ (g - I^*)/g^2$$

It may be noted that as g tends to infinity we have $I^*/g \to 3/4$, and hence

$$g^{-2} \ V_{max}(g) \to 27/256$$

4. Null models

When interpreting the value of V found for a certain graph, one may ask the question "couldn't this value of V have been obtained by mere chance?". Or, put in more respectable terms, "would this value be likely under a simple stochastic null model?". In order to answer this question, the mean and variance of V will be computed under some null models. As the degree variance will be used as a descriptive statistic complementary to the density of the graph, all these null models will be conditional on the density (or, equivalently, on the degree sum s).

Firstly, the null model is considered where the total number of lines ($s/2$ for undirected graphs, s for directed and bipartite graphs) is distributed at random among all the possible pairs of points ($g(g-1)/2$ for undirected graphs, gh for directed and bipartite graphs): every admissible incidence matrix with exactly s entries equal to 1 is equally probable. The computations of the mean and variance of V under this null model, denoted by $E\{V|s\}$ and $\text{var}\{V|s\}$ are elementary but time-consuming. Only the results will be given; the computations can be found in Snijders (1981). For undirected graphs, we get

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

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

and for directed and bipartite graphs

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

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

In the language of Holland and Leinhardt (1975, 1979), this could be called the $U|s$ distribution: the uniform distribution, conditioned on s. (Holland and Leinhardt use the notations C and X_{++} for s in their 1975 and 1979 papers, respectively, and therefore speak about the $U|C$ or $U|X_{++}$ distribution.) For directed graphs, it can also be relevant to condition on the numbers m of mutual dyads (pairs of points (i, j) with $i < j$ and $x_{ij} = x_{ji} = 1$), a of asymmetric dyads (pairs (i, j) with $i < j$ and $x_{ij} \neq x_{ji}$) and n of null dyads (pairs (i, j) with $i < j$ and $x_{ij} = x_{ji} = 0$). The distribution where the $g(g-1)/2 = m + a + n$ dyads are randomly divided into three subsets of sizes m (mutual dyads), a (asymmetric dyads) and n (null dyads), respectively, and where the directions in the asymmetric dyads are randomly chosen, is called the $U|man$ distribution. Note that $s = 2m + a$. For directed graphs with the $U|man$ distribution, the mean and variance of V can be found from the theorems in Holland and Leinhardt (1975). In Snijders (1981) another approach is followed. The result is

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

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

with

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

If $m = s/2$ and $a = 0$, the $U|man$ distribution for directed graphs coincides with the $U|s$ distribution for undirected graphs, and the formulas for mean and variance of V are identical. For directed graphs, the formulas for mean and variance of the in-degrees are identical to those given here for V (the variance of the out-degrees).

5. The degree variance as an estimator

In the preceding Sections the degree variance was treated as a purely descriptive statistic. The stochastic null models of § 4 were not proposed as acceptable models for 'real life situations', but rather as a standard of emptiness with which the value found for V can be rated. In some situations, however, it will be sensible to regard the observed graph as the outcome of a random graph generated by some stochastic mechanism: for example, if the graph is produced by a sampling mechanism of some sort. In that case V (or rather $g^{-2}V$, as will be seen below) can be regarded as an estimator for a parameter of this stochastic mechanism. In order to investigate which parameter is being estimated, the expected value of V will be derived.

It will be assumed that the probability distribution of the random graph is invariant under permutations of the points. Random variables are designated by capital letters; for example, X_{ij} is a random variable with outcome x_{ij}, the indicator of a line from i to j. We define

$$p_1 = P\{X_{ij} = 1\}$$

$$p_{11} = P\{X_{ij} = X_{is} = 1\}$$

$$p_2 = P\{X_{ij} = X_{st} = 1\}$$

for distinct indices i, j, s and t. Because of the permutation invariance, these probabilities do not depend on i, j, s and t. If the lines (i.e., the X_{ij}) are independent, then

$$p_{11} = p_2 = p_1^2$$

The parameter $p_{11} - p_2$ can be interpreted as a measure of the stochastic dependence of two lines at a common point, as compared to lines at different points. If $p_{11} - p_2 > 0$ then the lines will tend to cluster at a few points with high degrees, whereas for $p_{11} - p_2 < 0$ the lines will tend to be evenly divided among the points.

Let us first consider the case of undirected graphs. Using the properties that $x_{ij} = x_{ij}^2 = x_{ji}$ and $x_{ii} = 0$, the following formulas are obtained. In all summations in this Section, the indices mentioned are understood not to assume identical values.

$$EX_{i\cdot}^2 = E\{\sum_j X_{ij}\}^2 = E\{\sum_j X_{ij} + \sum_{j,s} X_{ij}X_{is}\} = (g-1)p_1 + (g-1)(g-2)p_{11}$$

$$EX_{\cdot\cdot}^2 = g^{-2}E\{\sum_{i,j} X_{ij}\}^2 = g^{-2}E\{2\sum_{i,j} X_{ij} + 4\sum_{i,j,s} X_{ij}X_{is} + \sum_{i,j,s,t} X_{ij}X_{st}\}$$

$$= (1-g^{-1})\{2p_1 + 4(g-2)p_{11} + (g-2)(g-3)p_2\}$$

It follows that

$$Eg^{-2}V = E\{g^{-3}\sum_i X_{i\cdot}^2 - g^{-2}X_{\cdot\cdot}^2\}$$

$$= (g-1)g^{-3}\{(g-2)(g-3)(p_{11}-p_2) + (g-2)(p_1-p_{11})\}$$

For large g, this is approximately equal to the dependence measure $p_{11} - p_2$. It can be concluded that $p_{11} - p_2$ is approximately the parameter being estimated by $g^{-2} V$.

Instead of this approach where an interpretation is sought for the degree variance, one can also start from the other side and wish to have an unbiased estimator for the dependence measure $p_{11} - p_2$. Such an estimator is

$$\{g(g-1)(g-2)\}^{-1} \sum_{i,j,s} X_{ij}X_{is} - \{g(g-1)(g-2)(g-3)\}^{-1} \sum_{i,j,s,t} X_{ij}X_{st}$$

which is equal to

$$\{(g-1)(g-2)(g-3)\}^{-1}\{(g+1)V + X_.^2\} - \{(g-2)(g-3)\}^{-1}X_.$$

Under the null model of § 4, we have $p_{11} - p_2 = 0$. This is reflected in the property that $E\{V|s\}$ is of the order of magnitude of g, which implies that under this null model $g^{-2}V$ converges to 0 in probability as g tends to infinity.

It may be noted that $Eg^{-2}V$ is necessarily nonnegative (as V is nonnegative), whereas $p_{11} - p_2$ can be negative. From the expression for $Eg^{-2}V$ one obtains the following lower bound for $p_{11} - p_2$:

$$p_{11} - p_2 \geqslant -(p_1 - p_{11})/(g-3) \geqslant -(g-3)^{-1}$$

If g is large, negative values of $p_{11} - p_2$ must be very close to 0; and for an infinite random graph with a probability distribution which is invariant under finite permutations, negative dependence is altogether impossible. (This is similar to the property of exchangeable (i.e. permutation invariant) infinite sequences of random variables, that negative dependence is impossible; see, e.g., the remark of Kingman (1978:187).)

For directed and bipartite graphs the formula for $EX_.^2$, and hence that for EV, contains other probabilities besides p_1, p_{11} and p_2. The precise expressions for EV will not be given here; it suffices to mention that in these cases, too, $Eh^{-2}V$ is, for large g and h, approximately equal to $p_{11} - p_2$.

6. Normalization

Several approaches to the normalization of the degree variance will be mentioned, all depending on g and s or (m, a, n). One approach is a normalization with respect to the null models of § 4. This leads to

$$\tilde{V} = \frac{V - E\{V|s\}}{\sqrt{\operatorname{var}\{V|s\}}}$$

and for directed graphs also to

$$\tilde{V} = \frac{V - E\{V|m, a, n\}}{\sqrt{\operatorname{var}\{V|m, a, n\}}}$$

as an alternative possibility. The measure $\tilde{V}$ will be used especially to examine whether the value found for V points to a considerable departure from the null model. This may be done formally by testing the null model as a null hypothesis, using the property that under this model $\tilde{V}$ has approximately (for large g) a standard normal distribution (see Holland and Leinhardt 1970, 1979); or it may be done in a loose way, to the effect that if $\tilde{V}$ does not differ much from 0, say $|\tilde{V}| \leqslant 2$, then one refrains from giving a substantial interpretation of the heterogeneity in the graph.

Another approach is a normalization with respect to the maximum values of § 3. This leads to the index of graph heterogeneity

$$J^2 = V/V_{\max}(g, d)$$

In order to obtain an index which is proportional to the absolute instead of the squared differences between the degrees, it is preferable to use

$$J = \sqrt{J^2}$$

The indices J^2 and J assume values from 0 to 1. For fixed d, the lower value 0 can be attained only if $(g - 1)d$ is an integer; otherwise, the minimum possible value is very small but positive.

The three elements mentioned (null model, maximum value, square root) can be combined into the index of graph heterogeneity

$$H = \frac{\sqrt{V} - \sqrt{E\{V|s\}}}{\sqrt{V_{\max}(g, d)} - \sqrt{E\{V|s\}}}$$

with $s = g(g - 1)d$. This index has the advantage over J that it is 0 if the value found for V is exactly compatible with the null model in the sense that $V = E\{V|s\}$. H assumes negative values if V is smaller than would be expected under the null model. Since

$$\lim_{g \to \infty} E\{V|s = g(g - 1)d\}/V_{\max}(g, d) = 0$$

H and J will be close to each other if g is large.

Of these three standardized versions of the degree variance, the values of $\tilde{V}$ for graphs with different numbers of points g are not comparable owing to the effect of g on the variance of V under the null model. The values of J and of H for graphs with different values of g and d will be comparable, however.

7. An example from *Social Networks*

As an example, the heterogeneity indices of § 6 will be applied to the graph presented in Fig. 1 and Table 1. The points represent the articles in the first volume of *Social Networks*; two points are connected by a line if there is at least one publication which is contained in the lists of references of both articles. These lines can be interpreted as pointing towards a relation between the articles connected, concerning the substantial contents or the methods used. For this undirected graph we have

Figure 1.

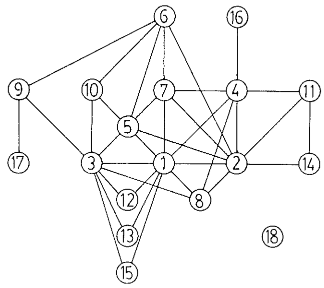

Table 1.

i	x_i	Author(s) and title
1	9	Ithiel de Sola Pool and Manfred Kochen Contacts and influence
2	8	Meindert Fennema and Huibert Schijf Analysing interlocking directorates: Theory and methods
3	8	Brian L. Foster Formal network studies and the anthropological perspective
4	6	Ronald S. Burt A structural theory of interlocking corporate directorates
5	6	Alvin W. Wolfe The rise of network thinking in anthropology
6	5	Linton C. Freeman Centrality in social networks: Conceptual clarification
7	5	Stephen D. Berkowitz *et al.* The determination of enterprise groupings through combined ownership and directorship ties
8	4	Stephen B. Seidman and Brian L. Foster A note on the potential for genuine cross-fertilization between anthropology and mathematics
9	3	Gary Coombs Opportunities, information networks and the migration-distance relationship
10	3	Forrest R. Pitts The medieval river trade network of Russia revisited
11	3	Gerrit Jan Zijlstra Networks in public policy: Nuclear energy in the Netherlands
12	2	Ove Frank Sampling and estimation in large social networks
13	2	Peter D. Killworth and H. Russell Bernard The reverse small-world experiment
14	2	Robert J. Mokken and Frans N. Stokman Corporate–governmental networks in the Netherlands
15	2	Lee Douglas Sailer Structural equivalence: Meaning and definition, computation and application
16	1	Ronald S. Burt Stratification and prestige among elite experts in methodological and mathematical sociology circa 1975
17	1	Davor Jedlicka Opportunities, information networks and international migration streams
18	0	Maureen T. Hallinan The process of friendship formation

$g = 18$	$s = 70$	$d = 0.23$
$V = 6.65$	$V_{\max}(g, d) = 21.77$	
$\widetilde{V} = 4.60$	$J = 0.55$	$H = 0.31$

The value of $\widetilde{V}$ clearly indicates that the null model $U|s$ is not adequate in the light of these data. For a more refined analysis, much more would be needed; e.g., one could take account of the differences in length between the lists of references of the articles considered.

The formulas which are presented in this paper and used in Section 6 for the various normalized heterogeneity indices have been programmed by Roel Popping and are part of the GRADAP computer package for analysis of graphs.

References

Coleman, J. S.
1964 *Introduction to Mathematical Sociology.* New York: Free Press.
Freeman, L. C.
1978 "Centrality in social networks: conceptual clarification". *Social Networks 1*:215–239.
Harary, F.
1969 *Graph Theory.* Reading, MA: Addison-Wesley.
Høivik, T. and N. P. Gleditsch
1975 "Structural parameters of graphs: a theoretical investigation". Pp. 203–223 in H. M. Blalock *et al.* (eds.), *Quantitative Sociology 1975.* New York: Academic Press.
Holland, P. W. and S. Leinhardt
1970 "A method for detecting structure in sociometric data". *American Journal of Sociology 70*:492–513.
1975 "Local structure in social networks". Pp. 1–45 in D. R. Heise (ed.), *Sociological Methodology – 1976.* San Francisco: Jossey - Bass.
1979 "Structural sociometry". Pp. 63–83 in P. W. Holland and S. Leinhardt (eds.), *Perspectives on Social Networks Research.* New York: Academic Press.
Kingman, J. F. C.
1978 "Uses of exchangeability". *Annals of Probability 6*:183–197.
Snijders, T. A. B.
1981 *Maximum Value and Null Moments of the Degree Variance.* TW-report 229 . Department of Mathematics, University of Groningen.