Social Networks, 3 (1981) 41–52

The Development of Structure in Random Networks: an Analysis of the Effects of Increasing Network Density on Five Measures of Structure*

Noah E. Friedkin

*University of California, Santa Barbara***

The density of ingroup relations continues to be proposed as an indicator of structural cohesion. Network density is obviously a misleading indicator of structural cohesion when a group has subgroups; in such circumstances, the cohesion may be entirely internal to the subgroups. However, it is plausible that network density is a useful indicator of structural cohesion when it can be assumed that a group lacks subgroups. In order to analyze this possibility, I construct a set of random networks, increase the density of relations in these networks, and observe how the networks' structure develops in terms of five measures. The results show that low densities in large networks may be associated with more structural cohesion than higher densities in smaller networks; it is suggested that in field studies, attempts to control for network size will encounter problems of nonlinearity and heteroscedasticity. I conclude that network density is not a useful indicator of structure and that direct measurement of structure is to be preferred.

Introduction

This paper investigates the relationship between network density and the structure of networks. Network density is a measure of the incidence of direct relations among the possible pairs of a network; structure refers to how the direct relations are combined or arranged in a network. Though network density is among the most commonly reported properties of networks, its implications for network structure have not been made explicit (cf. Harary *et al.* 1965:75).

The approach taken here is to construct a set of random networks that differ in size, to increase the density of relations in these networks, and to observe how the structure of each network develops. Stages of structural development in the random networks may be revealed with this approach. The consideration of networks of different sizes is included in the approach

*This paper has benefited from the comments of James Coleman, Lawrence Hubert, Eugene Johnsen, Edward Laumann, and several referees of this journal.
**Department of Education, University of California, Santa Barbara, CA 93106, U.S.A.

so that the possible significance of network size as a conditional factor can be addressed.

The intended scope of this inquiry does not include an attempt to arrive at precise expected values of network structure under different conditions of network size and density. Instead, the analysis seeks to obtain *qualitative* statements about the development and state of structure under different conditions of network density and size. To illustrate the kind of description being sought, I shall be concerned more with whether distinct stages of development may be discerned than with estimating precisely the values of network density which correspond to the emergence of each stage; I shall be concerned more with whether similar values of network density are associated with similar states of structure in networks of different size than with estimating precisely the expected states of structure under the different conditions.

Methods and constructs

The inquiry is based on an analysis of random networks constructed with a Monte Carlo procedure.[1] I follow Erdös and Renyi (1960) who work with an evolving random network: (1) the first edge of the network is randomly selected from the $N(N-1)/2$ possible edges between N members; (2) the second edge of the network is randomly selected from $[N(N-1)/2]-1$ remaining possible edges; (3) edges continue to be selected in this manner up to an arbitrarily set limit; (4) at various points, the process is interrupted to calculate statistics on the structure of the evolving network.[2]

Large numbers of Monte Carlo repetitions are necessary to achieve a highly precise estimate of the expected value of a structural parameter in a random network. Large numbers of repetitions were not practical, due to the computer expense involved in calculating the measures of structure in the networks. Some error, therefore, will be present in the absolute values reported in this analysis which arises from the relatively small number of Monte Carlo repetitions upon which the results are based.[3]

---

[1]See Coleman (1964) and Rapoport (1963) who discuss the uses of mathematical models and simulation techniques; also see Mayhew and Levinger (1976), Rapoport (1979), and Rapoport and Horvath (1961).

[2]A FORTRAN program, written by the author, is used to accomplish this task. The program 'stops' to calculate attributes of the evolving network each time the network's density has been increased approximately one percent. The author would be pleased to provide further details on the nature of this program upon request.

[3]The data in Figs. 1–5 are the mean results of five repetitions of the Monte Carlo procedure described in the body of the text. Exceptions are the data involving networks of twenty members which, owing to greater instability of the results, are based on ten repetitions. Despite the small number of repetitions used to produce each curve, the curves are quite smooth in appearance. This smoothness stems, in part, from the fact that a curve is based on the mean of several Monte Carlo repetitions and, also, from the fact that the structure of the network 'evolves', instead of being generated anew at each level of density; hence, discontinuities of structure are unlikely from one level of network density to the next. As is pointed out in the text which follows, the number of repetitions is sufficient for the purposes of this analysis. A qualitative comparison of the curves and a qualitative description of the general form of the curves are supported by the overall consistency of the findings, rather than by the achievement of a high level of precision in the estimation of each individual curve.

However, the analysis does not require a greater degree of precision given its limited purpose, that is, to obtain *qualitative* statements about network structure in relation to network density. The results will show that there is considerable internal consistency in the pattern of results across networks that differ in size. The observed degree of internal consistency provides powerful and sufficient support for the qualitative statements forwarded in this analysis.

Seven constructs are involved in the analysis. The simplest construct is network size which is defined as the number of members in a network; the remaining six constructs are considered in detail below.

Network density indicates how nearly a network is *complete*—a state in which each member is connected directly with every other member (Harary *et al.* 1965:7). The analysis deals with graphs (symmetric diagraphs); *i.e.* a direct connection of member i with member j implies a direct connection of j with i. In these circumstances, network density equals $2E/N(N-1)$, where E is the number of direct relations (edges) among N members. Thus, a network density of 0.5 indicates that a relation occurs in one-half of the possible pairs in a network.

Network density is involved in this analysis because over at least the past thirty years, it has been proposed as an indicator of group cohesion. Kephart (1950:549) may have been the first to suggest network density as such an indicator. Barnes (1969:61–64) has treated network density as an indicator of the extent to which a network is 'close-knit'; Bott (1957:250), following Barnes' usage, has associated network density with the idea of "connectedness". More recently, Blau (1977:136) has suggested that network density reflects the "strength of ingroup bonds or group cohesion". It is to be expected that network density, currently the most commonly reported property of networks, will become still more prevalent now that its estimation in large groups is practical (Granovetter 1976).

It is doubtful that a single property of structure can be found that is sufficient to characterize the overall extent of structural cohesion in a network (*cf.* Holland and Leinhardt 1978). It is likely that a battery of structural properties will provide a better indicator of structural cohesion in a network than any single property. It remains to be determined which among the many possibilities ought to be included in such a battery of indicators. The present analysis considers five measures of structure. They have been selected, not only because they are relatively easily calculated, but also because they are heavily implicated in noteworthy empirical and theoretical sociological studies:

(1) among the total $N(N-1)/2$ pairs, the proportion in which the members are mutually reachable;

(2) the average length of the geodesics (*i.e.* the shortest paths);

(3) the length of the longest geodesic (in networks in which all the members are mutually reachable, this property is equivalent to network diameter);

(4) the proportion of individual members that are not involved in a triad (*i.e.* in a completely connected subgraph of three members); and

(5) the number of such triads that individual members are involved in, on the average.

Measures (1)–(3) of the extent of reachability and the length of geodesics in a network bear on the analysis of interpersonal flows in networks. A path between two network members indicates that an opportunity exists for the occurrence of a flow (*e.g.* of information or influence) from one to the other. The distance separating the two members is a condition of the actual likelihood of a flow—the greater the length of the shortest path between two members, the less likely the flow.

With regard to measures (4) and (5) of triadic structure, Simmel (1950) first pointed to the possible significance of triads; Goode (1960) has discussed the role of third parties to relationships in a useful and interesting way. In general, triadic structures, these analyses suggest, imply the presence of third parties to social relationships who may act to mediate interpersonal conflict and to forestall the breaking-off of relationships during the process of social control. Triadic structure is at the basis of both Bott's (1957) distinction of close-knit from loose-knit networks and Laumann's (1973) distinction of interlocking from radial networks. For an additional perspective on triadic structure, see the work of Davis (1977), and Holland and Leinhardt (1976), and the literature cited therein.

Results

A network's structural cohesion generally increases with an increase in its network density. The relationship is markedly nonlinear in form for each of the five measures of network structure. Dramatic increases in structural cohesion occur at relatively low values of network density, *i.e.* in the range 0.0–0.5; in this range, all the pairs of a network rapidly become connected, the geodesics at first increase then rapidly decrease in length, and the involvement of individual network members in triads becomes both prevalent and intensive. In terms of the structural measures utilized here, it appears that most of the possible structural variation in the networks is exhausted at relatively low values of network density. Though one might expect that a comparable or higher level of network density would be required in a large network, in comparison to a smaller one, to achieve an equivalent level of structural cohesion, such is not the case; comparable levels of structural cohesion are achieved at lower levels of network density in large networks than in small ones.

Proportion of joined pairs

Erdös and Renyi (1960) find that when the number of relations in a network goes beyond one-half the number of members in the network the proportion of joined pairs abruptly increases. Ling and Killough (1976) show that Erdös and Renyi's study, as well as others that rely on asymptotic results, do not provide a necessarily reliable basis of inference about what may obtain in small finite samples (also see Schultz and Hubert, 1973). The present results (with the exception of those based on networks of size 20) are surprisingly consistent with the findings of Erdös and Renyi.

Erdös and Renyi's threshold, after which point the proportion of joined pairs begins to rapidly increase, occurs at a density of 0.05 in a network of size 20, at a density of 0.03 in a network of size 40, and at a density of 0.02 in a network of size 60. The results shown in Fig. 1 are not inconsistent with the predictions: in these small networks there is a rapid increase in the proportion of joined pairs around the predicted threshold values of network density.

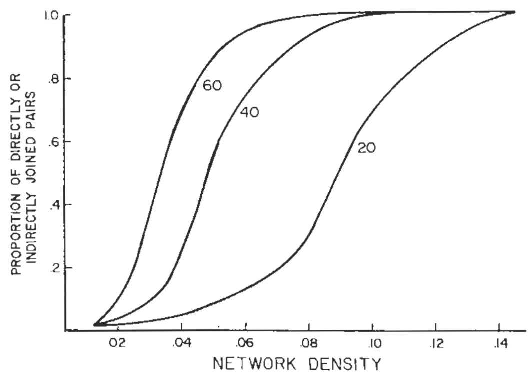
Note that the larger the size of a network, the *lower* the value of network density at which the proportion of joined pairs begins to increase most rapidly. Moreover, note that even in the smallest of the networks—the network with 20 members—all of the pairs are joined at a density which is well below 0.20. It is clear, then, that only the range of low values of network density is pertinent to an account of the structural cohesion of networks in terms of the extent of reachability.

A low network density in a large network *belies* the degree of connectivity which may exist in it. For example, the network of size 20 and densities of 0.04 have, on average, 0.05 of their pairs joined; at the same time, the network of size 60 and densities of 0.04 have, on the average, 0.77 of their pairs joined. It may not be assumed that a low density in a network which is larger than another necessarily implies that the larger network is less structurally cohesive in terms of the proportion of its joined pairs.

Length of geodesics

Beginning at low values of network density, increase of network density is at first associated with an increase both in the length of a network's longest

Figure 1. *Monte Carlo results on network density and the proportion of joined pairs in networks of 20, 40, and 60 members.*



geodesic and in the mean length of all the geodesics. This is seen in Figs. 2 and 3. At low densities there is little indirect connectivity in a network; most of the connected pairs are joined by direct relations. As the density of the network increases, more pairs are joined by way of paths through intermediaries and, hence, the length of the geodesics in the network increases.

The larger a network, the higher is the maximum expected length to which the geodesics will rise. After the point of maximum expected length has been reached, further increases of network density are associated with a decline in the length of the geodesics that is more rapid in the larger networks than in the smaller. Most of the possible reduction in geodesic length has occurred in the lower range of network density; beyond a density of 0.20, these data suggest that smaller reductions in geodesic lengths can be expected with further increases of network density.

Figure 2. *Monte Carlo results on network density and length of the longest geodesic (among the joined pairs) in networks of 20, 40, and 60 members.*

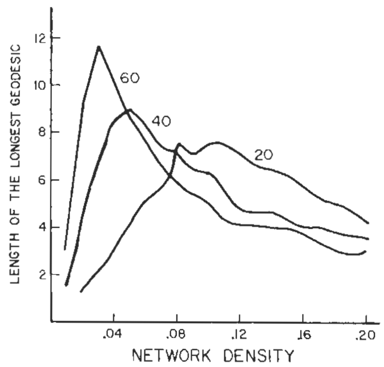
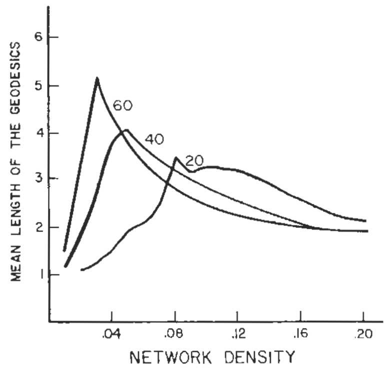


Figure 3. *Monte Carlo results on network density and the mean length of geodesics (among the joined pairs) in networks of 20, 40, and 60 members.*



Again, in these data it is apparent that a low network density in a large network does not necessarily correspond to a low degree of structural cohesion, relative to a smaller network with a higher density. A low density in a large network belies the actual degree of structural cohesion that is present in it.

Members participation in complete subgroups of size three (triads)

Network members' participation in triads increases with an increase of network density. Figure 4 shows that the proportion of members that are not involved in a triad declines with increasing network density; Fig. 5 shows that the number of triads, with which the average member is involved, increases dramatically with increasing network density.

Figure 4. *Monte Carlo results on network density and the proportion of network members who are not involved in a triad in networks of 20, 40, 60, and 80 members.*

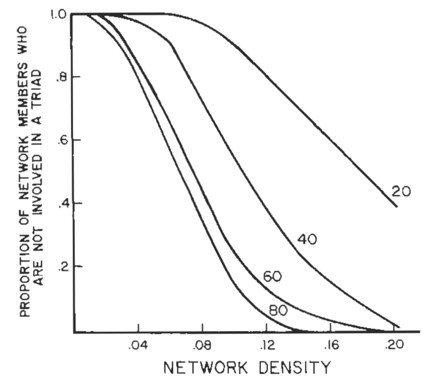
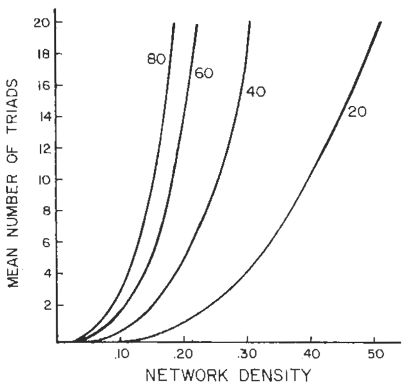


Figure 5. *Monte Carlo results on network density and the mean number of triads per member in networks with 20, 40, 60, and 80 members.*



With this structural property as with the others that have been examined, substantial gains in network cohesion occur at relatively low levels of network density. In all but the smallest networks, the proportion of network members not involved in a triad drops precipitously with increases in network density in the range 0.0–0.50; in the same range of values, the average number of triads individual members are involved in increases precipitously.

Furthermore, it is seen once again that the same network density in networks of different size may indicate substantially different degrees of structural cohesion; it requires a larger value of network density to achieve the same level of structural cohesion in a small network than in a larger one. Low densities in large networks may belie the degree of structural cohesion in them in terms of the prevalence and intensity of members participation in triads.

Conclusions

For some time, it has been recognized that there is a serious constraint upon the use of network density as a measure of structural cohesion. Festinger *et al.* (1950:94) have pointed out that an aggregate measure of group cohesion can be misleading if it does not take into account subgroup formations:

> "As an extreme illustration, there conceivably might be two subgroups of four people each, each member within each subgroup choosing every other member but without any choices at all between the subgroups. In this case each of the subgroups may have great cohesiveness but the cohesiveness of the group as a whole would be low."

It is plausible, however, that within groups where it is known or can be assumed that subgroups are absent, network density may be a useful measure of structural cohesion. In such circumstances, it is plausible that relatively high values of network density are associated with relatively high levels of structural cohesion and that low values of network density are associated with low levels of structural cohesion. Perhaps it is because of the extreme plausibility of these associations that we have been beguiled for so long into an uncritical acceptance of them.

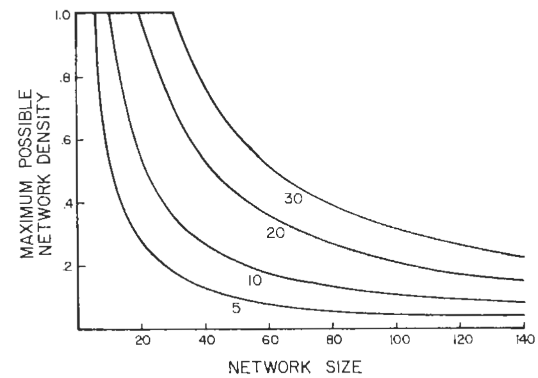
Holding network size constant, increases in network density generally do correspond with increases in networks' structural cohesion, as indicated by five measures of structure. Dramatic increases in structural cohesion occur within the lower (0–0.50) range of values of network density. These data suggest that, in field studies, observed variations of network density in the range 0.5–1.0 may not be as important in the account of phenomena as is variation in the lower range. But this conclusion is based on a set of structural measures that may be quite distinctive in their lack of ability to discriminate structural variation under conditions of high network density. Other measures of structure may be more useful in describing the development of structure with increase of network density in its upper range. With further work along these lines it may be possible eventually to describe the structural development of a network in terms of a battery of properties, some of which are salient in a network's initial stages of structural development and others which are salient in its later stages.

Network density ought not to be used as an indicator of structural cohesion when networks of different sizes are being compared. Similar values of network density in networks of different size correspond to substantially different levels of structural cohesion. It requires a larger value of network density to achieve the same degree of structural cohesion in a small than in a large network; accordingly, a low network density in a large network can belie the actual degree of its structural cohesion.

One is tempted to conclude that network density may be a useful general indicator of structural cohesion if only network size is controlled. However, I believe that there are serious problems involved in an attempt to control network size towards this end. Some evidence presented below suggests that the relationship between network size and density is nonlinear and heteroscedastic.

Figure 6 shows how the maximum possible network density declines in networks of different size: each curve involves an assumption about the maximum number of direct relations that individual network members can maintain. The curves suggest how network size in general may be associated with network density.[4]

Figure 6. *Declining maximum possible density in networks of different sizes. The curves assume, respectively, that each network member can maintain up to 5, 10, 20, and 30 direct relations with other members.*



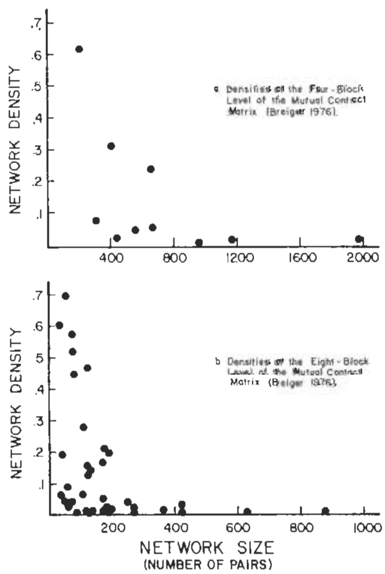
[4]Mitchell (1969:1920) writes that "in general it appears that there is probably a limit to the number of people with whom an individual might be in direct and regular contact" and it is his impression "that this limit in an urban environment may be about thirty persons". Network density on these curves equals $M/(N - 1)$, where M is a prescribed maximum number of direct contacts each person may be involved with among N persons (*i.e.* 5, 10, 20, or 30 contacts in the present illustration): E (the number of edges) = $M \times N/2$; therefore $2E/N(N - 1) = M \times N/N(N - 1) = M/(N - 1)$. The results of Fig. 6 are not based on the Monte Carlo procedures described earlier in the paper (no constraint on the maximum number of contacts per population member is involved in the Monte Carlo production of random graphs). Figure 6 is an analytical exercise meant to illustrate that variation in the number of contacts persons are able to maintain is associated with a progressively smaller degree of variation in network density in progressively larger populations and, therefore, that high network density in large populations may be expected to be rare.

Network density rapidly declines in a nonlinear manner with an increase of network size. Variation in the number of relations that network members can feasibly maintain is associated with a progressively smaller degree of variation in network density in progressively larger networks. In other words, we should in general be surprised to find a high value of network density in fairly large networks.

Figure 7 displays the same rapid decline of network density with increasing network size (here defined as the number of possible pairs in it). The data are from Breiger's (1976) four- and eight-block matrices of mutual contact among a sample of scientists. When the number of possible contacts in a block is relatively small, the network density of the block manifests its full range of variation, from 0.0 to 1.0; but as the size of the network increases, network density is more strictly confined to low values. It is worth noting that there is nothing intrinsic to the CONCOR algorithm that might account for these results on the relationship between block size and density.

Field researchers who wish to employ network density as a general indicator of structural cohesion may face a serious problem of control because of possible nonlinearity and heteroscedasticity in the relationship between network density and size. This problem, in combination with the previously mentioned difficulty of interpreting the structural significance of network density when subgroups are present, lead me to conclude that network density is not a generally useful indicator of network structure. Direct measurement of structure is to be preferred on the basis of the variety of measures that are currently available.

Figure 7. *The relationship of network size and density in a sample of scientists.*



References

Barnes, J. A.
1969 "Networks and political processes". In J. C. Mitchell (ed.), *Social Networks in Urban Situations*: 51–76. Manchester: Manchester University Press.

Blau, P. M.
1977 *Inequality and Heterogeneity*. New York: Free Press.

Bott, E.
1957 *Family and Social Network*. London: Tavistock.

Breiger, R. L.
1976 "Career attributes and network structure: a blockmodel study of a biomedical research speciality". *American Sociology Review 41*:117–135.

Coleman, J. S.
1964 "Mathematical models and computer simulation". In Robert Faris (ed.), *Handbook of Modern Sociology*: 1027–1062. Chicago: Rand-McNally.

Davis, J. A.
1977 "Sociometric triads as multi-variate systems". *Journal of Mathematical Sociology 5*:41–59.

Erdös, P. and A. Renyi
1960 "On the evolution of random graphs". *Publication of the Mathematical Institute of the Hungarian Academy of Sciences 5*:17–61. Reprinted in *Paul Erdös: The Art of Counting*: 569–573, J. Spencer (ed.). Cambridge, MA: MIT Press.

Festinger, L., S. Schachter and K. Back
1950 *Social Pressures in Internal Groups*. New York: Harper.

Goode, W. J.
1960 "Norm commitment and conformity to role-status obligations". *American Journal of Sociology 66*:246–258.

Granovetter, M.
1976 "Network sampling: some first steps". *American Journal of Sociology 81*:1287–1303.

Harary, F., R. Z. Norman and D. Cartwright
1965 *Structural Models: An Introduction to the Theory of Directed Graphs*. New York: Wiley.

Holland, P. W. and S. Leinhardt
1976 "Local structure in networks". In D. R. Heise (ed.), *Sociological Methodology 1976*. San Francisco: Jossey-Bass.
1978 "An omnibus test for social structure using triads". *Sociological Methods and Research 7*:227–256.

Kephart, W. M.
1950 "A quantitative analysis of intragroup relationships. *American Journal of Sociology 55*:544–549.

Laumann, E. O.
1973 *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: Wiley.

Ling, R. L. and G. G. Killough
1976 "Probability tables for cluster analysis based on a theory of random graphs". *Journal of the American Statistical Association 71*:293–300.

Mayhew, B. H. and R. L. Levinger
1976 "Size and density of interaction in human aggregates". *American Journal of Sociology 82*:86–110.

Mitchell, J. C.
1969 "The concept and use of social networks". In J. C. Mitchell (ed.), *Social Networks in Urban Situations*: 1–50. Manchester: Manchester University Press.

Rapoport, A.
1963 "Mathematical models of social interaction". In R. D. Luce, R. Bush, and E. Galanter (eds.), *Handbook of Mathematical Psychology II*: 493–579. New York: Wiley.
1979 "A probabilistic approach to networks". *Social Networks 2*:1–18.

Rapoport, A. and W. J. Horvath
1961 "A study of a large sociogram". *Behavioral Science 6*:279–291.

Schultz, J. V. and L. J. Hubert
1973 "Data analysis and the connectivity of random graphs". *Journal of Mathematical Psychology 10*:421–428.

Simmel, G.
1950 *The Sociology of Georg Simmel.* Translated, edited, and with an introduction by K. H. Woff. New York: Free Press.