to appear in
Social Networks

NETWORK STRUCTURE AND MINIMUM DEGREE

by

Stephen B. Seidman
Department of Mathematical Sciences
George Mason University
Fairfax, Virginia 22030

ABSTRACT

NETWORK STRUCTURE AND MINIMUM DEGREE

BY

Stephen B. Seidman
Department of Mathematical Sciences
George Mason University
Fairfax, Virginia 22030

Social network researchers have long sought measures of network cohesion. Density has often been used for this purpose, despite its generally admitted deficiencies. An approach to network cohesion is proposed that is based on minimum degree and which produces a sequence of subgraphs of gradually increasing cohesion. The approach also associates with any network measures of local density which promise to be useful both in characterizing network structures and in comparing networks.

I. Introduction

Most social network research has devoted serious attention to the construction of measures of network structure. Many such measures have been defined, addressing a wide variety of network structural features. One aspect of network structure that has been treated frequently is the question of network cohesiveness. Cohesiveness ideas have been used (at least in a metaphorical sense) for many years, but perhaps their first explicit appearance is in the work of Bott (1957). Bott distinguishes between "tightly-knitted" and "loosely-knitted" networks and uses the degree of "knittedness" of a couple's network to explain the degree to which the couple's conjugal roles are separate (Bott 1957:59). For Bott, the "knittedness" of a couple's network is measured by the degree to which the couple's friends know each other. This measure was soon formalized and generalized by Barnes, Mitchell and their students, who adopted a graph-theoretic perspective. From this perspective, the "knittedness" of a network was replaced by the concept of the density of a graph; that is, the ratio of the number of links in a graph to the maximum number of possible links (Barnes 1968:117).

Density, like most other network structural measures, has been used in several distinct ways. First, networks can be compared on the basis of their densities. Alternatively, network density can be used as a structural attribute of a single network. Both of these perspectives were used by Bott; they can be regarded as global applications of the density concept. Density has also been used to address local network structure. For example, in Kapferer's analysis of a conflict in a Zambian factory (Kapferer 1969), individuals are compared on the basis of the density of their first-order zones.

It was soon realized, however, that density is less than satisfactory as a cohesiveness measure. One serious problem with density arises from the fact that it can be obtained from an average of the degrees of the points in a network. Thus a particular density value may arise either from a rather "uniform" network or from a network consisting of a very cohesive region (with points of high degree) and a very sparse region (with points of low degree). Density is incapable of distinguishing between these situations.

The work described above had the goal of determining the cohesiveness of an entire network or of an ego-centered subnetwork. Alternatively, one can try to identify maximally cohesive subnetworks of a network. This perspective has been used for several decades in sociometry, and the term "clique" has long been used for such a maximally cohesive subnetwork. Although many heuristic and statistical definitions of cliques have been given, the first formal definition is due to Luce and Perry (1949). For Luce and Perry, a clique is a maximal complete subgraph of a network. Such a subgraph is certainly maximally cohesive. Unfortunately, two serious difficulties arose with the use of cliques in empirical analysis. Most empirical networks contain few nontrivial cliques, and in any case the cliques of a network overlap in a complex pattern. These shortcomings have been addressed by many scholars. On the one hand, generalized cliques were defined that retained strongly cohesive properties but which would tend to be larger in practice (see Luce 1950, Alba 1973, Seidman and Foster 1978a, Mokken 1979). On the other hand, algorithmic procedures have been devised to merge overlapping cliques into larger cohesive subsets (Alba n.d., Kappelhoff 1974). All of these methods have been reasonably successful in locating cohesive regions of networks.

But once the highly cohesive regions of a network have been located, we are still left with two problems. First, it is clear that all of the interesting network structure does not lie in the cohesive regions of the network. In cohesive network regions, individuals are likely to be linked by multiple, redundant paths, while individuals outside cohesive regions (or in different cohesive regions) are likely to be linked (if at all) by single paths. Granovetter (1973) has argued that it is just such weak links that may be crucial in the transmission of information. An analytical strategy that is limited to the identification of cohesive network regions will be unable to deal with the structure of the portion of the network lying outside (and between) the cohesive regions.

But if the cohesive subset strategy is unable to address the structure of the non-cohesive portion of a network, it is certainly unable to determine parameters that summarize the global structure of a network. Such summary parameters can be used as part of a network description or as the basis of a comparison of two or more networks. We have seen that density has been used as such a parameter, but that it has not been satisfactory. The purpose of this paper is to propose a way in which the advantages of the cohesive subset and density approaches can be combined. Using the approach outlined here, network regions will be identified that will contain all highly cohesive regions, even though the larger regions may not themselves be highly cohesive. Despite their possible lack of extreme cohesion, these regions will have interesting and clearly specifiable structural properties. In addition, this approach will produce global measures of network structure that are sensitive both to the cohesive regions of a network and to the portions of the network outside the cohesive regions.

The approach proposed in this paper is in some sense orthogonal to the standard cohesive subset approaches. While cohesive subsets are usually built up from small but highly cohesive cliques by gradually weakening some aspect of cohesion (see Seidman and Foster 1978b), the subsets obtained by the proposed approach will first be large, gradually diminishing in size. They can be seen as seedbeds, within which cohesive subsets can precipitate out.

II. Definition and Properties of k-cores

Let G be a graph. If H is a subgraph of G, $\delta(H)$ will denote the minimum degree of H; each point of H is thus adjacent to at least $\delta(H)$ other points of H. If H is a maximal connected (induced) subgraph of G with $\delta(H) \geq k$, we say that H is a k-core of G.

It is easy to see that the 1-cores of a graph are its nontrivial connected components. Figure 1 shows a graph and its k-cores for several values of k. It is clear that, in general, k-cores need not be highly cohesive, but that all cohesive subsets are contained in k-cores. Note that a connected graph can have at most one 2-core, and that trees have no 2-cores. Also, it is clear that a k-core must have at least k+1 points, and that points in different k-cores cannot be adjacent.

Although k-cores may not in practice be maximally cohesive, it is possible to say something explicit about their cohesiveness. Cohesiveness in social networks can be characterized in several distinct but related ways. Suppose that we are looking at some subset of the population making up the members of the network. If the subset is cohesive, it should be possible to remove some of its members without fragmenting the subset. To make this idea formal, we regard the social network as a graph, so that the cohesive subset corresponds to an induced subgraph. The connectedness of a graph is defined to be the minimum number of points whose removal will produce a disconnected graph (or the trivial graph on one point) (Harary 1969:43).

Cohesive subsets of social networks should correspond to subgraphs of high connectedness. From this perspective, cliques are seen to be maximally cohesive, since a clique on p points has connectedness equal to p-1. It is easy to see that the connectedness of a k-core must be between 1 and k. The following result shows that, for k-cores with relatively few points, we can say something more about the connectedness of k-cores. Proofs of all results will be found in an appendix.

Proposition 1: If $\ell \leq k$ and $p \leq 2k-\ell + 2$, then a k-core with p points has connectedness at least ℓ . Note that this bound is best possible. To see how the bound works in particular cases, let $\ell = 2$. We see that a k-core with at most 2k points must have connectedness at least 2 (i.e. it has no cut points). Thus while a 3-core with 6 points must have connectedness at least 2, a 3-core with 7 points may have a cut point.

An alternative way of interpreting the connectedness of a subgraph in a sociologically interesting way is provided by Menger's theorem. This result, originally proved by Menger (1927) has appeared in the literature in many variant forms (Harary 1969:47-52). For our purposes, the most interesting variant is due to Whitney (1932); it states that the connectedness of a graph is at least ℓ if and only if every pair of points is joined by at least ℓ point-disjoint paths. Thus a cohesive subset whose corresponding subgraph has high connectedness will also have high redundancy; messages can reach their destinations by many alternate (and independent) paths.

This approach to connectedness has ignored the length of the paths. It is often important that paths between individuals be short, for example to guarantee that messages be communicated with little degradation of content. A geodesic path between points s and t of a graph is a path from s to t with the fewest possible links, and the diameter of a connected graph is the length of the longest geodesic path between any pair of points of the graph. Thus if the diameter of a connected graph is d, any two points of the graph can be linked by a path with no more than d links. Just as we should expect cohesive subsets of a network to correspond to subgraphs of high connectedness, we should also expect them to correspond to subgraphs of small diameter. Once again, cliques are seen to be maximally cohesive; all cliques have diameter equal to 1. Just as we were able to show that small k-cores had high connectedness, we are able to show that small k-cores have small diameter. Note that [x] denotes the largest integer less than or equal to x, and that mod(p,q) is the remainder obtained when the integer p is divided by the integer q.

Proposition 2: (a) If K is a k-core with p points and if $k+1 < p < 2k+2$, then diam(K) = 2.

(b) If K is a k-core with p points and if $p \geq 2k+2$, than $\text{diam}(K) \leq 3\,[p/(k+1)] + a(p,k) - 3$,

$$\text{where}\quad a(p,k) = \begin{cases} 0 & \text{if } \mathrm{mod}(p,k+1) = 0 \\ 1 & \text{if } \mathrm{mod}(p,k+1) = 1 \\ 2 & \text{if } \mathrm{mod}(p,k+1) = 2 \end{cases}$$

This bound is best possible. A weaker version of this result (with a(p,k) replaced by 2) was obtained by Kramer(1972). We can conclude from part (a) that a 3-core with 5,6 or 7 points must have diameter 2. A 3-core with 8 points must have diameter at most 3; if it has 9 points, its diameter is at most 4, while if it has 10 or 11 points, its diameter is at most 5. In general, for a given value of k, smaller k-cores are more cohesive, whether cohesiveness is measured by connectedness or by diameter.

The examples that demonstrate that Proposition 2b gives a best possible bound all have cut points. It is natural to expect that better diameter bounds can be obtained for k-cores whose connectedness is known to be greater than 1. Such a bound is given in the next result, which can be seen as a generalization of Proposition 2b.

Proposition 3: If K is a k-core with p points and connectedness ℓ, and $p \geq 2k+2$, then

$$\mathrm{diam}(K) \leq 3\,[(p-2k-2)/\beta] + b(p,k,\ell) + 3, \text{ where}$$

$$\beta = \max\{k+1,\ 3\ell\} \text{ and}$$

$$b(p,k,\ell) = \begin{cases} 0 \text{ if } & \mathrm{mod}(p-2k-2,\beta) < \ell \\ 1 \text{ if } & \ell \leq \mathrm{mod}(p-2k-2,\beta) < 2\ell \\ 2 \text{ if } & 2\ell \leq \mathrm{mod}(p-2k-2,\beta) \end{cases}$$

Note that the hypotheses of Proposition 3 require that $1 \leq \ell \leq k < p$. If $\ell = 1$ is used in Proposition 3, we obtain Proposition 2b. The bounds given in Proposition 3 are also best possible. To see how much the bounds in Propostion 3 are improvements over those given in Proposition 2, we will look at 3-cores with connectedness 1, 2, or 3 and at most 12 points. Table 1 shows the relationship between the number of points, the connectedness, and the maximum diameter for 3-cores. Note that the diameter bounds given in Table 1 for connectedness equal to 1 are precisely those given by Proposition 2. It is clear that higher connectedness yields sharper diameter bounds. Once again, we see that smaller k-cores can be expected to be more cohesive.

Both of the measures of cohesiveness that we have discussed so far share the property that they assign a number to any graph; extreme values for that number indicate a high degree of cohesiveness. Another way of looking at cohesiveness is to ask whether a network contains a maximally cohesive nucleus. For example, a subset of a network may not itself be a clique, but if it contains a clique with 4 points, say, it is evidently more cohesive than a subset that contains only trivial cliques. Such a large clique could serve as a focus of activities that could tie the non-clique members more tightly into the subset. If we represent networks as graphs, we can then ask the question "When must a graph of a certain size contain a clique of a certain size?". Questions of this sort are treated in the branch of graph theory called extremal graph theory, and indeed the first results in extremal graph theory dealt with the question of when graphs must contain cliques of various sizes. In particular, a result of Turán(1941) states that any graph with p points and more than $t_{r-1}(p)$ edges must contain an induced subgraph with r points that is complete, where

$$t_{r-1}(p) = \sum_{0 \leq i < j < r-1} \left[\frac{p+i}{r-1}\right] \left[\frac{p+j}{r-1}\right] .$$

In this form, Turán's theorem is not useful for our purposes, but the following result(Bollobás 1978:295) is more directly applicable. A thorough survey of related results can be found in Bollobás(1978).

Proposition 4: A k-core with p points must contain a clique with r points if $p < \left(\frac{r-1}{r-2}\right) k$.

Thus if $p < 2k$, a k-core on p points must contain a clique with 3 points, and if $p < 3k/2$, a k-core on p points must contain a clique with 4 points. It is easy to find examples showing that these results are best possible. The dependence on p is perhaps easier to see if we rewrite the bound in Proposition 4 as $r < k/(p-k) + 2$. If $p = k+1$, this implies the existence of a clique with k+1 points inside the k-core, which is thus complete. If $p = k+2$, complete subgraphs of the k-core with r points can be found for $r < (k/2) + 2$, while if $p = k+3$, complete subgraphs of the k-core with r points can only be assured if $r < (k/3) + 2$. Thus all three measures of cohesiveness indicate that smaller k-cores tend to be more cohesive.

III. Remainders and Core Collapse Sequences

Although the cohesive regions of a social network are contained in the k-cores of the corresponding graph, we have argued above that interesting and important structure will be found outside the cohesive regions. In this section, we will investigate the structure of the complements of the k-cores, and we will use that structure to define global structural parameters for social networks.

Suppose that G is a graph. Let $G_0 = G$, and let G_k be the union of the k-cores of G for k=1,2,... . Define $R_k = G_k - G_{k+1}$. R_k is the set of points that are in k-cores but that are in no (k+1)-cores. R_k will be called the k-remainder of G. Note that R_0 is the set of isolates of G. If H is a set of points of G, the frontier F(H) of H is defined to be the set of points of G-H that are adjacent to points of H. Clearly, $F(G_{k+1}) \subseteq R_k$. Finally, a graph G is said to be k-degenerate (Bollobas 1978:222) if for every induced subgraph H of G, we have $\delta(H) \leq k$.

Proposition 5: The subgraph induced by $F(G_k)$ is (k-2)-degenerate (k = 2,3,...).

A particularly interesting example of the way this result can be applied is obtained when we examine the structure of $F(G_3)$. In this case, $F(G_3)$ is 1-degenerate, so that it can have no induced subgraph of minimum degree 2 or more. Thus $F(G_3)$ is acyclic, so that its connected components are trees. For example, in Figure 1b, the 3-cores are $\{a,b,c,d\}$ and $\{j,k,l,m\}$. It is easy to see that $F(\{a,b,c,d\}) = \{e,f,g\}$, and $F(\{j,k,l,m\}) = \{f,g,i\}$. Each of these frontiers is a forest (its connected components are trees). Suppose that the object of study is the network arising from the relation "talks with" on the set of participants in a scientific conference. A 3-core in this network is a maximal set of people, each of whom talks to at least three other members of the set. It is reasonable to believe that such sets correspond to strong interest subgroups. But members of such interest subgroups also talk with nonmembers; the frontier of a 3-core is the set of nonmembers who talk with members. We have seen that the frontier breaks up into components which are trees. These components could correspond to distinct facets of the interests that together make up the defining interests that correspond to the 3-core. Even more interestingly, the tree structure of the components of the frontier can be used to construct a hierarchy within each component that may correspond to relationships among the various interests.

Frontiers of k-cores for $k > 3$ have more complex structure that is not so easily interpreted. On the other hand, most naturally occurring networks will not contain significant k-cores for large values of k. Although the frontiers are very useful in analyzing the interface between the k-cores and their complements, it is also interesting to look at the entire remainder R_k. Arguments similar to those used in the proof of Proposition 5 show that the subgraph induced by R_k is k-degenerate. In order to obtain a structural argument similar to the one presented above for frontiers, we would need to know that a remainder was 1-degenerate. This only happens for R_1, but since it is clear that the complement of G_2 in G_1 can only consist of disjoint tree-like appendages, we obtain no additional insight.

It is intuitively reasonable that k-cores should be more dense than k-remainders. Suppose that the k-remainder R_k has r_k points. Then a result of Lick and White (1970) implies that R_k has at most $kr_k - k(k-1)/2$ edges, from which an upper bound on the density of R_k can easily be obtained. Similarly, a lower bound for the density of G_k can be obtained from the fact that (if G_k has n_k points) G_k has at least $kn_k/2$ edges. Unfortunately, these bounds are not very useful in practice, since the lower bound for G_k is derived on the assumption that all points of G_k have degree k, while the upper bound for R_k is derived on the assumption that R_k is a k-degenerate graph on r_k points with as many edges as possible. Neither assumption is likely to be realistic.

Despite the failure of a direct approach to the density of the cores and remainders, it still seems reasonable that interesting structural information is contained in the sequence of cores and remainders. One way to get at this information is to consider the sequence $\{r_k\}$ (k=0,1,2,...) of remainder sizes. r_0 is the number of isolates in the network, r_1 is the number of points in tree-like subnetworks, and $r_2, r_3, \ldots$ count the numbers of points in successively more cohesive network regions. This sequence certainly carries structural information, but how can it be used? Whether it is used as a global parameter to aid in the analysis of a particular network or to aid in network comparisons, it should be made independent of network size. This can be done by replacing r_k by $r_k/|V(G)|$ which represents the proportion of points of G that are in $G_k - G_{k+1}$. If we regard the sequence of core unions G_k as having been produced by a process starting with $G=G_0$ and gradually stripping off less cohesive regions, $r_k/|V(G)|$ is the proportion of points that are removed by the (k+1)st removal operation. We will call $\{r_k/|V(G)|\}$ the <u>core collapse sequence</u> (CCS) of a network. For the network that appears in Figure 1, the core collapse sequence is (0,5/21,8/21,8/21).

Before proceeding further with interpretation of the core collapse sequence, we should ask how long such a sequence can be. It is clear that if Δ is the <u>maximum</u> degree of a graph, then the core collapse sequence for that graph can have no more than $\Delta+1$ terms, and that that upper bound can be attained. The CCS is a way of replacing density by a sequence of measures of local density. Figure 2 shows how the CCS can be used to distinguish between two graphs that have the same density but quite different structure. It is of course possible for graphs to have the same CCS but different density, but the emphasis of the CCS on the existence of induced subgraphs of specified minimum degree tends to insure the structural similarity of graphs with similar CCS. The CCS should be very useful in the comparison of networks on the same or different populations.

If we look more closely at the way in which G_{k+1} is obtained from G_k, we see that an iterative process is involved. This process starts out with G_k, and each iteration removes all points of degree k from the previously obtained graph. Let i_k be the number of iterations that are needed to produce G_{k+1} from G_k. If G_k has n_k points, it is clear that $0 \leqslant i_k \leqslant n_k - k - 2$ (these values will be increased by 1 if the final iteration, producing no changes, is included). The value of i_k carries structural information about R_k. In particular, it can be used to investigate the existence of long paths in R_k.

<u>Proposition 6</u>: (a) There is a path in R_k of length $i_k - 1$.

(b) If there is a path in $F(G_3)$ of length t, then $i_2 \geqslant t/2$.

The converse of Proposition 6a is false. It would be interesting to generalize Proposition 6b to apply to G_k for k 3. In general, if i_k is small, R_k will be relatively "thin", while if i_k is large, R_k will be relatively "thick". Examples of graphs with large and small values of i_k (illustrating this distinction) are given in Figure 3. The structural significance of i_k seems relatively independent of the size of the original network, and thus provides another structural parameter that can be used along with the CCS in global descriptions of network structure.

IV. Conclusions

We have shown that the k-core concept can be used in two distinct but related ways. First, k-cores are subsets of a network whose cohesion increases as k increases. For small values of k, the k-cores tend to be large, diminishing in size as k increases. Furthermore, cohesive subsets defined in any reasonable way must occur as subsets of k-cores for some nontrivial value of k. Thus k-cores can be regarded as seedbeds, within which we can expect highly cohesive subsets to be found. On the other hand, k-cores can be used to define the core collapse sequence of a network. This sequence takes account both of the k-cores and their complements, and thus considers both regions of strong ties and regions of weaker ties. It can be used to characterize the "mesh" of a network in a far more subtle and interesting way than by using network density directly. Core collapse sequences should thus prove very useful in global network comparisons. In addition, the k-cores and core collapse sequences will be very easy and efficient to compute, since all that is really necessary is to access the row marginals of the adjacency matrix of the network.

Acknowledgements

The research on which this paper is based was partially supported by the National Science Foundation under Grant No. BNS 80-13507. The paper was written while the author was a visting scholar in the Department of Anthropology, Arizona State University. The author would like to thank Brian L. Foster for his helpful comments on an earlier version of the manuscript.

References

Alba, R.D.

1973 "A graph-theoretic definition of a sociometric clique." Journal of Mathematical Sociology 3:113-126.

n.d. "COMPLT:A program for the analysis of sociometric data and the clustering of similarity matrices." Ms. Columbia University.

Barnes, J.A.

1968 "Networks and political process." In M.J. Swartz (ed.), Local-Level Politics:107-130. Chicago:Aldine.

Bollobás, Béla

1978 Extremal Graph Theory. London: Academic Press.

Bott, Elizabeth

1957 Family and Social Network. London: Tavistock.

Granovetter, Mark

1973 "The strength of weak ties." American Journal of Sociology 68:1360-1380.

Harary, Frank

1969 Graph Theory. Reading, Mass.: Addison-Wesley.

Kapferer, Bruce

1969 "Norms and the manipulation of relations in a work context." In J.C. Mitchell (ed.), Social Networks in Urban Situations:181-244. Manchester: Manchester University Press.

Kappelhoff, Peter

1974 Die Bestimmung und Charakterisierung Primärer Umwelten mit Hilfe soziometrischer Daten im Rahmen eines Mehrebenenmodells. Dissertation (Köln).

Kramer, F.

1972 "Relations between the diameter of a graph and the degrees of the nodes of a graph." Rev. Anal. Numer. Teoria Approximatiei 1:125-131.

Lick, D.R. and A.T. White

1970 "K-degenerate subgraphs." Canadian Journal of Mathematics 22:1082-1096.

Luce, R.D.

1950 "Connectivity and generalized cliques in sociometric group structure." Psychometrika 15:169-190.

Luce, R.D. and A. Perry

1949 "A method of matrix analysis of group structure." Psychometrika 14:94-116.

Menger, K.

1927 "Zur allgemeinen Kurventheorie." Fundamenta Mathematicae 10:96-115.

Mokken, R.J.

1979 "Cliques, clubs and clans." Quality and Quantity 13:161-173.

Seidman, S.B. and B.L. Foster

1978a "A graph-theoretic generalization of the clique concept." Journal of Mathematical Sociology 6:139-154.

1978b "A note on the potential for genuine cross-fertilization between anthropology and mathematics." Social Networks 1:65-72.

Turán, Paul

1941 "On an extremal problem in graph theory." Mat.Fiz. Lapok 48:436-452.

Watkins, M.E.

1967 "A lower bound for the number of vertices of a graph." American Mathematical Monthly 74:297.

Whitney, H.

1932 "Congruent graphs and the connectivity of graphs." American Journal of Mathematics 54:150-168.

Proposition 1 is a restatement of Corollary I.1.4 in (Bollobás 1978). Similarly, Proposition 2a is a restatement of Theorem 2 in (Seidman and Foster 1978a). As noted in the test above, Proposition 2b is a special case of Proposition 3, obtained from the latter result by setting $\ell=1$. If S is a set, $|S|$ will denote its cardinality. The connectedness of a graph G will be denoted by $\kappa(G)$.

Proof of Proposition 3: Let K be a k-core with p points and connectedness ℓ. Set $\beta = \max\{k+1, 3\ell\}$, and suppose that diam(K) = $t \geq 3$. Note that if diam (K) < 3, the desired bound is trivially satisfied. Choose points u,v, in K at distance t, and define $N_i = \{x \in K \mid d(u,x) = i\}$, where d(u,x) is the distance from u to x, and $i = 0,1,2,\ldots,t$. $N_0 = \{u\}$, $N_i \neq \emptyset$ for $0 \leq i \leq t$, and $V(K) = \bigcup_{i=0}^{t} N_i$. We will now partition the points of K by grouping the sets $\{N_i\}$. Define

$$P_0 = N_0 \cup N_1,$$

$$P_i = N_{3i-1} \cup N_{3i} \cup N_{3i+1} \quad (i=1,2,\ldots,[(t-4)/3]),$$

$$Q = N_{t-1} \cup N_t,$$

and

$$R = V(K) - Q - \bigcup_{i=0}^{[(t-4)/3]} P_i.$$

Note that R may be empty. If R is nonempty, it is equal either to one of the $\{N_i\}$ or to the union of two of the $\{N_i\}$. Let $\mathcal{S} = \{P_0, P_1, \ldots, Q\}$, and put $|\mathcal{S}| = j$. Note that if $3 \leq t \leq 5$, then $\mathcal{S} = \{P, Q\}$ and j=2. Since the degree of each point of K is at least k, we see that for each $T \in \mathcal{S}$, $|T| \geq k+1$. But since $\kappa(G) = \ell$, $|N_i| \geq \ell$ for $1 \leq i \leq t-1$, so that $|P_i| \geq 3\ell$ for $i \geq 1$, and thus $|P_i| \geq \beta$ for $i \geq 1$. Since the members of $\mathcal{S}$ are disjoint subsets of V(K), we conclude that

$$(j-2)\beta + 2(k+1) \leq p, \text{ or that}$$

$$j \leq [(p-2k-2)/\beta] + 2.$$

Case 1: $\mathrm{mod}(p-2k-2,\beta) < \ell$.

If $j = [(p-2k-2)/\beta] + 2$, at most $\ell-1$ points remain unaccounted for by the minimum size calculations for the members of $\mathcal{S}$. Since $|N_i| \geq \ell$ for each i, these points cannot be included in R; they must therefore be included in various members of $\mathcal{S}$, so that $R = \emptyset$. We conclude that

$\mathrm{diam}(K) = 2(j-2) + 2(j-1) + 2 = 3j-3$, where the first term comes from the internal links in each P_i $(i \geq 1)$, the second term comes from the links between the members of $\mathcal{S}$, and the third term comes from the internal links in P_0 and Q. It thus follows that

$\mathrm{diam}(K) = 3\,[(p-2k-2)/\beta] + 3 = 3[(p-2k-2)/\beta] + b(p,k,\ell) + 3$, since $b(p,k,\ell)=0$ in this case.

If $j < [(p-2k-2)/\beta] + 2$, the points unaccounted for by the minimum size calculations may either be included in members of $\mathcal{S}$ or they may constitute one or two ungrouped N_i's. This may increase the diameter by 1 or 2, but since j is being decreased by at least 1, 3j is decreased by at least 3. We conclude that

$\mathrm{diam}(K) < 3\,[(p-2k-2)/\beta] + b(p,k,\ell) + 3$,

completing the discussion of Case 1.

Case 2: $\ell \leq \mathrm{mod}(p-2k-2,\beta) < 2\ell$

If $j = [(p-2k-2)/\beta] + 2$, then the number of points unaccounted for by the minimum size calculations is between ℓ and $2\ell-1$. These points may be included in members of $\mathcal{S}$ or they may make up one ungrouped N_i. If an ungrouped N_i appears, the diameter bound for K in case 1 is increased by 1. Since in this case we have $b(p,k,\ell)=1$, we conclude easily that

$\mathrm{diam}(K) \leq 3\;\;[(p-2k-2)/\beta] + b(p,k,\ell) + 3$, as desired.

If $j < [(p-2k-2)/\beta] + 2$, the arguments given above at the end of the discussion of Case 1 imply here too that

$\mathrm{diam}(K) < 3\;\;[(p-2k-2)/\beta] + b(p,k,\ell) + 3$.

Case 3: $mod(p-2k-2,\beta) \geq 2\ell$

If $j = [(p-2k-2)/\beta] + 2$, then the number of points unaccounted for by the minimum size calculations is between 2ℓ and $\beta - 1$. Recall that $\beta = \max\{3\ell, k+1\}$. If $\beta = 3\ell$, we see that no more than two ungrouped N_i's can appear. If $\beta = k+1$, we can draw the same conclusion. Hence the diameter bound for K in Case 1 can be increased by no more than 2. Since in this case we have $b(p,k,\ell) = 2$, we conclude that $diam(K) \leq 3[(p-2k-2)/\beta] + b(p,k,\ell) + 3$.

If $j < [(p-2k-2)/\beta] + 2$, the arguments at the end of Case 1 imply that $diam(K) < 3[(p-2k-2)/\beta] + b(p,k,\ell) + 3$. Thus the desired bound has been verified in all cases, completing the proof of Proposition 3.

We will now show that the bound given in Proposition 3 is best possible. Let p, k, and ℓ be given with $p \geq 2k+2$ and $k \geq \ell$; we will construct a graph $G(p,k,\ell)$ with p points, minimum degree k, connectedness ℓ, and whose diameter is given by the expression in Proposition 3. We will denote the join of graphs G_1 and G_2 by G_1+G_2 (Harary1969:21). The complete graph on n points will be denoted by K_n.

Case 1: $(3\ell \geq k+1)$

Let $j = [(p-2k-2)/\beta] = [(p-2k-2)/3\ell]$.

Define $N_0=K_1$, $N_1=K_k$, $N_2=N_3=\ldots=N_{j+1}=K_\ell + K_\ell + K_\ell$. Suppose first that $mod(p-2k-2,\beta)=0$. Then define $N_{j+2} = K_k$ and $N_{j+3}=K_1$. The graph $G(p,k,\ell)$ is formed by adding edges from each point of N_{i-1} to each point of N_i for $i=1,2,\ldots,j+2$. It is easy to see that $G(p,k,\ell)$ has the desired properties. If $0 < mod(p-2k-2,\beta) < \ell$, we can construct the desired graph by adding the additional $mod(p-2k-2,\beta)$ points to N_1. If $\ell \leq mod(p-2k-2,\beta) < 2\ell$, we define $N_{j+2}=K_{mod(p-2k-2,\beta)}$, $N_{j+3}=K_k$ and $N_{j+4}=K_1$. This increases the diameter of G by 1 without affecting the other properties. Finally, if

$a = \mathrm{mod}(p-2k-2, \beta) \geq 2\ell$, we define $N_{j+2}=K_\ell$, $N_{j+3}=K_{a-\ell}$, $N_{j+4}=K_k$ and $N_{j+5}=K_1$. In this case, the diameter of G is increased by 2 without affecting the other graph properties.

Case 2: $(3\ell < k+1)$

Once again, let $j = [(p-2k-2)/\beta] = [(p-2k-2)/(k+1)]$. In this case, the previous construction is modified slightly. The definition of N_i for i=0,1 and $i \geq j+2$ is unchanged, while $N_2=N_3=\ldots=N_{j+1} = K_\ell + K_\ell + K_{k-2\ell+1}$. It is easy to see that the desired conditions are satisfied by the graph $G(p,k,\ell)$.

The following useful bound is a simple restatement of Proposition 3.

Corollary 3.1: If G is a connected graph with p points, minimum degree δ, and connectedness κ, then

$$\mathrm{diam}(G) \leq 3\,[(p-2\delta-2)/\beta] + b(p,\delta,\kappa) + 3, \text{ where } \beta = \max\{\delta+1, 3\kappa\}, \text{ and}$$

$$b(p,\delta,\kappa) = \begin{cases} 0 & \text{if } \mathrm{mod}(p-2\delta-2,\beta) < \kappa \\ 1 & \text{if } \kappa \leq \mathrm{mod}(p-2\delta-2,\beta) < 2\kappa \\ 2 & \text{if } 2\kappa \leq \mathrm{mod}(p-2\delta-2,\beta). \end{cases}$$

By setting $\delta=\kappa$ in Corollary 3.1, we immediately obtain a bound due to Watkins (1967). Similarly, by setting $\kappa=1$ in Corollary 3.1, we obtain a best possible bound on the diameter of a connected graph as a function of its order and its minimum degree, which improves a bound due to Kramer (1972).

Proposition 4 is Corollary VI.1.3 in (Bollobas 1978).

Proof of Proposition 5: Suppose that the frontier of G_k were not (k-2)-degenerate. Let H be a subgraph of $F(G_k)$ with $\delta(H) \geq k-1$. But since $H \subseteq F(G_k)$, each point of H is adjacent to some point of G_k. Hence $\delta(H \cup G_k) \geq k$, so that each point of H would be contained in some k-core. But since $H \cap G_k = \emptyset$, this contradicts the maximality of the k-cores.

Proof of Proposition 6:

(a) Let $R_k(i)$ be the result of i iterations of the removal operation (with minimum degree k+1) applied to R_k. Note that $R_k(0) = R_k$, and $R_k(i_k) = \emptyset$. Suppose that $x \in R_k(i) - R_k(i+1)$, where $i \geq 1$. Thus x is removed by the (i+1)st iteration. Hence the degree of x in $R_k(i)$ is at most k, while the degree of x in $R_k(i-1)$ is at least k+1. x must therefore be adjacent to at least one point in $R_k(i-1) - R_k(i)$. Since $R_k(i)-R_k(i+1)$ is nonempty for $i=0,1,\ldots,i_k-1$, it is easy to see that we can construct a path in $R_k(0) = R_k$ with i_k points, and therefore with length i_k-1.

(b) Suppose P is a path of length t in $F(G_3)$. All interior points of P have degree at least 3, so that the only points that can be removed by the first removal iteration are the endpoints of the path. It follows by induction that $i_2 \geq t/2$.

Titles for Figures

Figure 1: A graph and its k-cores

Figure 2: Graphs having the same density but different core collapse sequences

Figure 3: Graphs having the same density and core collapse sequences, but different values of i_2.

Captions for Figures

Figure 1: Figure 1a represents a graph G with two connected components. Since G has no isolates, each connected component of G is a 1-core of G. Figure 1b shows the 2-cores of G, while Figure 1c shows the 3-cores of G. G has no k-cores for $k > 3$.

Figure 2: The graphs shown in Figures 2a and 2 b have density equal to 8/15, but the core callapse sequence for the graph of Figure 2a is (0,1/3,0,2/3), while that for the graph of Figure 2b is (0,0,1,0). It is clear from the core collapse sequences that the graph of Figure 2a consists of a dense core with some peripheral hangers-on, while that of Figure 2b is more uniform in mesh but lacks a highly cohesive core.

Figure 3: The graphs shown in Figures 3a and 3b have density equal to 1/4 and their core collapse sequences are (0,0,1/2,1/2). The value of i_2 for the graph of Figure 3a is 1, while that for the graph of Figure 3b is 4.

p \ ℓ	1	2	3
4	1	1	1
5	2	2	2
6	2	2	2
7	2	2	2
8	3	3	3
9	4	3	3
10	5	4	3
11	5	4	4
12	6	5	4

Table 1

Maximum diameter for 3-cores with p points and connectedness ℓ.

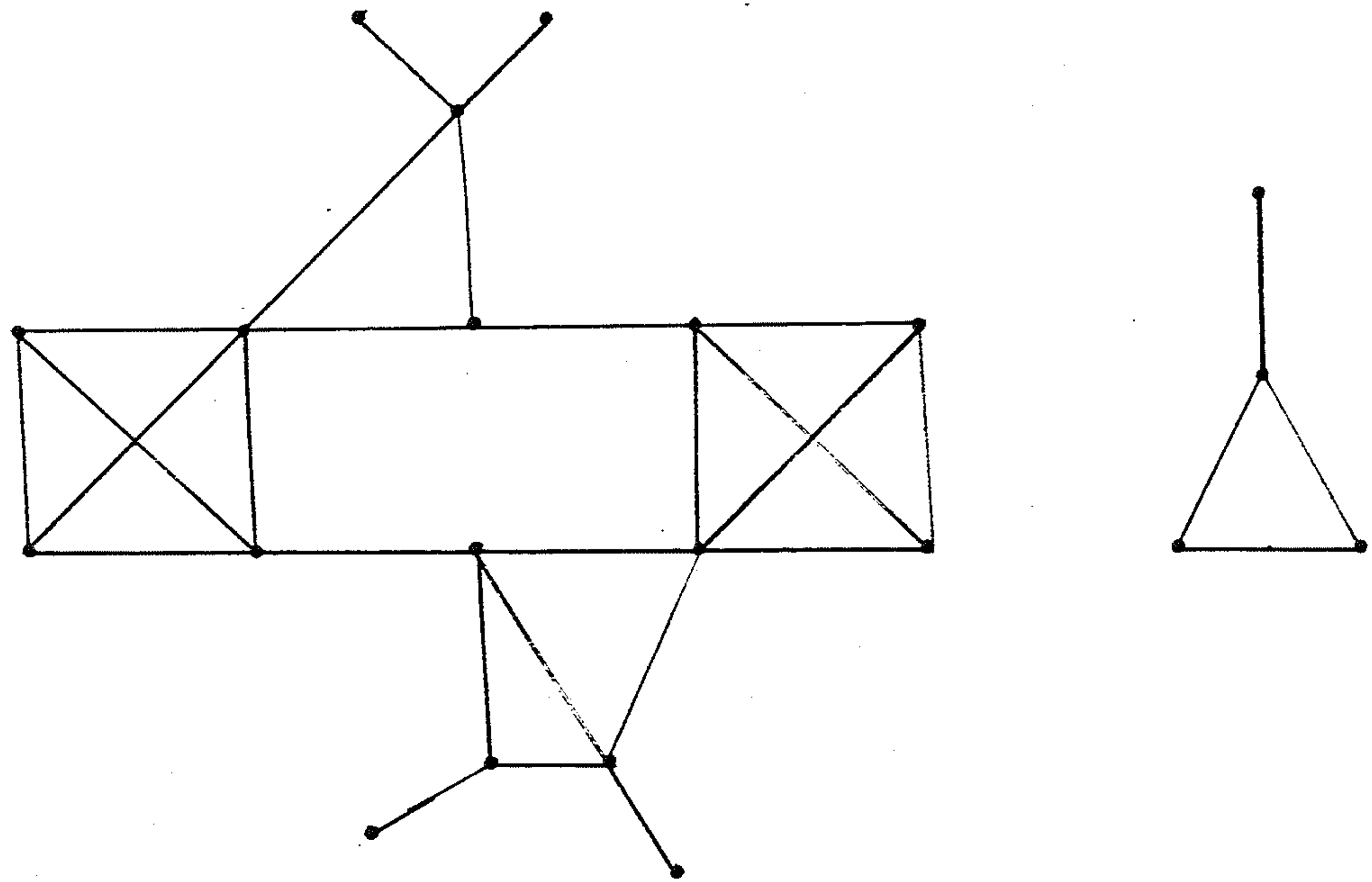

FIGURE 1A

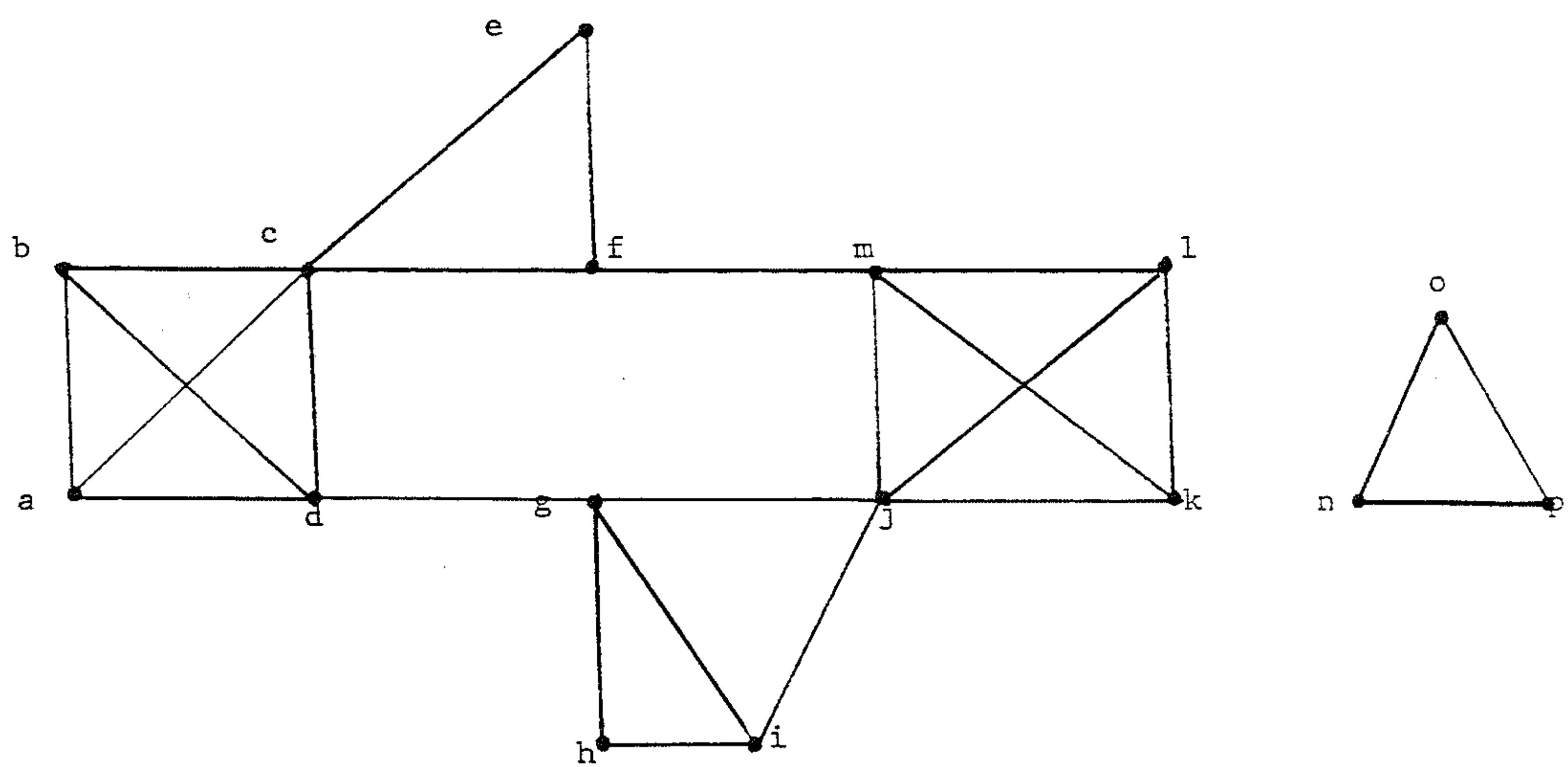

FIGURE 1B

FIGURE 1C

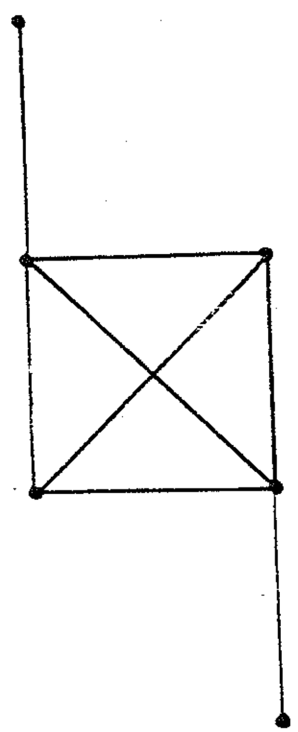

FIGURE 2A

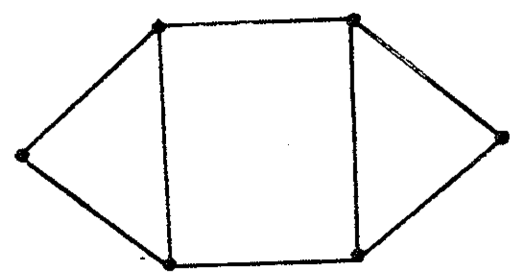

FIGURE 2B

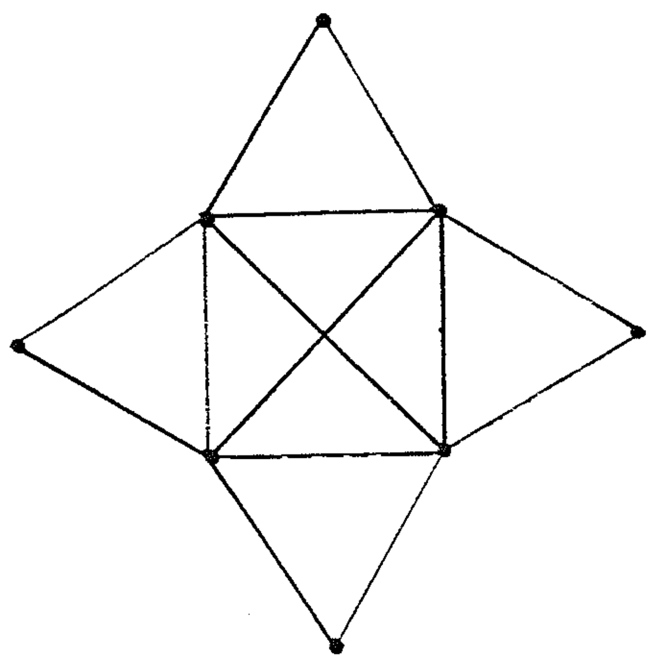

FIGURE 3A

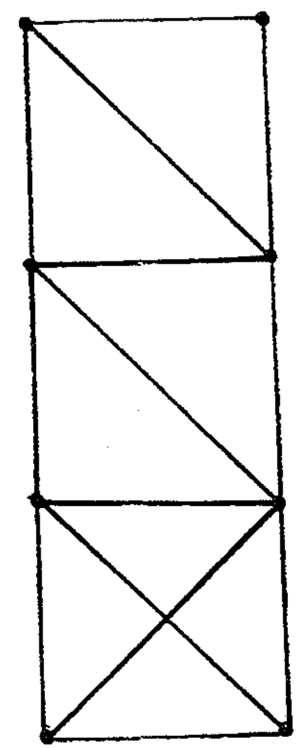

FIGURE 3B

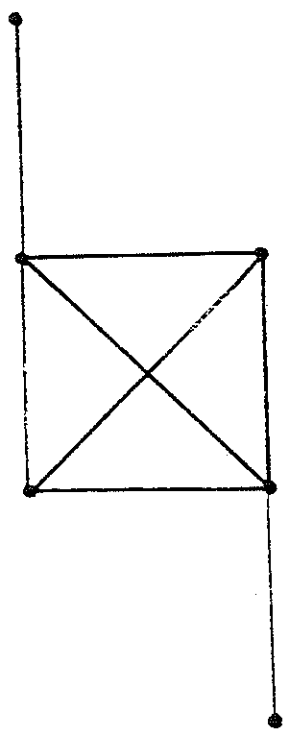

FIGURE 2A

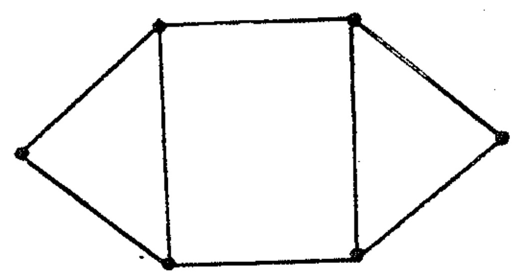

FIGURE 2B

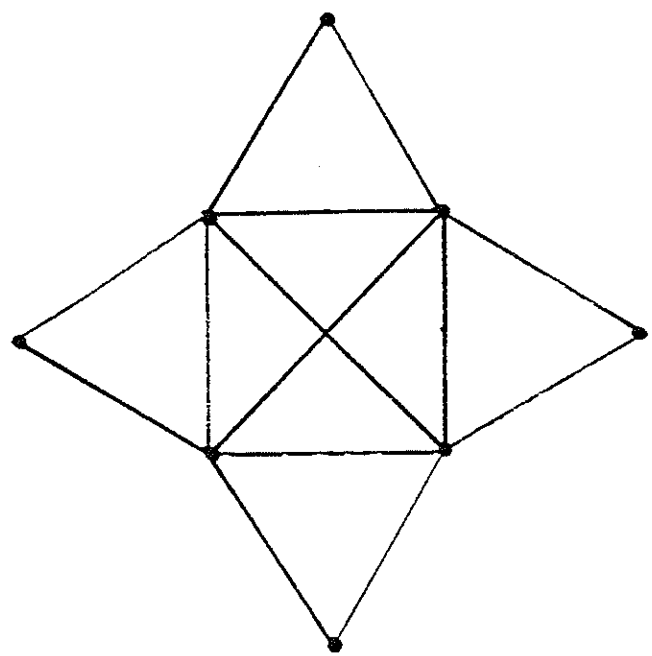

FIGURE 3A

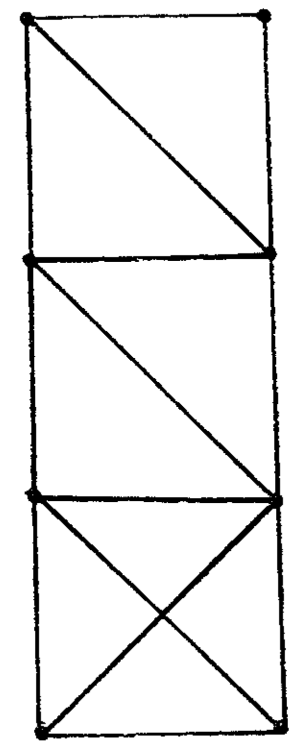

FIGURE 3B