Mathematical Social Sciences 3 (1982) 291-304
North-Holland Publishing Company


# CENTERED GRAPHS AND THE STRUCTURE OF EGO NETWORKS*

Linton C. FREEMAN
*School of Social Sciences, University of California, Irvine, CA 92717, U.S.A.*

Communicated by F.W. Roush
Received 14 December 1981
Revised 8 March 1982

A way of comparing ego networks through examining patterns among their ties is introduced. It is derived from graph-theoretic ideas about centered graphs. An illustration using data from a computer conference is provided.

*Key words:* Centered graph; ego network; clique.

## 1. Introduction: ego networks

This essay is about a special kind of social networks called *ego networks*. Like other networks, they are made up of *social units* - persons, groups or organizations - and *social relations* - like marriage or friendship - that link pairs of units into some sort of overall patterned structure.

Ego networks, however, are special in that they are built around a particular social unit designated as *ego*. To uncover an ego network you begin by choosing some social unit as ego. You then define a symmetric social relation or combination or relations and determine all the other units with whom ego has the relation. Finally, you note all pairs among those other units who are directly connected by the relation.

The result is a mini-network or immediate neighborhood surrounding ego that can, perhaps, reveal something important about the social world from ego's perspective. We might guess, for example, that if the social units are persons, those in the ego network are the ones that have the greatest impact on ego's attitudes, norms, values, goals and perceptions of the world. Moreover, those are the ones to whom ego must turn to seek information, help and support. In some sense, then, an ego network is a map of ego's personal social world. It shows something about how that ego is tied in to the larger human society.

Bott (1957) first called attention to the importance of such ego networks. She called them *social* networks and reasoned that to the extent that ego was involved

* This is a report of a computer conference on social networks supported by the National Science Foundation under Grant No. DSI77.16578.

0165-4896/82/0000-0000/$02.75 

with a set of others that were themselves tied directly together, everyone would tend "to reach consensus and ... exert consistent informal pressure on one another to conform to the norms, keep in touch with one another, and, if need be, to help one another." She suggested that such tightly linked collections of others are similar to traditional organized communities. She called them 'close-knit'.

At the other extreme, Bott reasoned that if the ties were sparse, such normative consistency would be less likely. In such a case, the focal unit, ego, would have a greater potential to choose freely among options and, at the same time, would experience less organized social support.

Bott's concern with the degree to which others in ego's network were directly tied together led her to stress the *density* (she called it *connectedness*) of links in the network. Density is simply the ratio of the number of observed links to their maximum.[1]

Other writers were quick to pick up Bott's basic idea. Many examined the relationship between density in an ego network and the support and satisfaction ego's position provides. Speck and Attneave (1973), for example, looked at the effects of increased density on ego's personal adjustment. And Mayer (1966), Boissevain (1968) and Kapferer (1969) all studied various effects of differing densities in ego networks on the ways in which egos use their networks to satisfy their personal needs and desires.

Concern with density as a structural attribute of ego networks has not, however, been free from criticism. Mitchell (1969, p. 18–19) argued:

> In sociological analysis our interest is primarily in reachability since norm enforcement may occur through transmission of opinions and attitudes along the links of a network. A dense network may imply that this enforcement is more likely to take place than a sparse one but this cannot be taken for granted. The pattern of the network must also be taken into consideration.

In seeking such patterning, Epstein (1969) suggested that the density of links might differ in different regions of ego's network. And Granovetter (1973) demonstrated the need for considering not only the existence of dense areas but of the *bridges* between them.

Thus, if we seek an indicator of the pattern or structure of ties in an ego network, we cannot – it seems – simply consider the overall density. The problem is to develop a way of looking at patterning that is sensitive to variations in local density and to linkages between dense areas. An attempt to develop such an approach is made in this essay. A set of formalisms will be introduced in the next section and, finally, they will be applied to some network data from a computer conference.

[1] The ratio should, as Granovetter (1973) has pointed out, be calculated for the rest of the network, excluding ego. This eliminates the effect of differing numbers of links to ego for ego networks of differing sizes.

## 2. Centered graphs

From a graph theoretic view, ego networks define a specific class of graphs that can be called *centered.*

Consider a graph, $G$, consisting of a set, $S$, of $k$ points and a set, $E$, of $e$ symmetrical edges linking pairs of points. Now if $k>2$ and there are $k-1$ edges such that some one point, $p^*$, is directly connected or adjacent to all of the others, $G$ is a *k-star*. A centered graph, then, is any graph of $k$ points that *contains* a $k$-star. Clearly, any ego network is, structurally, a centered graph.

There are several more or less obvious, but interesting properties of centered graphs:

(1) In terms of density, centered graphs have two extremals: the minimal $k$-star and the maximal complete graph (where all possible edges are present). A $k$-star has $k-1$ edges connecting $p^*$ to the $k-1$ other points. A complete graph has $\binom{k}{2}=\frac{1}{2}k(k-1)$ edges connecting each point to all of the others. Any density index is, of course, completely insensitive to the arrangement of those edges in intermediate forms.

(2) Centered graphs are, of course, *connected*; that is, there is a sequence of points and edges - a *path* - from any point to all others.

(3) The shortest path connecting a pair of points is called the *geodesic*. The longest geodesic linking any pair of points in a centered graph is less than or equal to 2: this is called the *diameter* of the graph.

(4) Any centered graph can be decomposed into Luce–Perry (1949) *cliques* that are the maximal subgraphs in which each point is directly connected, or *adjacent*, to all the others.

(5) Any geodesic connecting a pair of points in a centered graph has, according to (3) above, length equal to either 2 or 1. If a path has a length of 1, the points involved are adjacent and are part of a clique. If, alternatively, its length is 2, the end points are in different cliques that are linked by a mid-point. Since path lengths form a partition in a centered graph, we can examine the structure of cliques as well as the pattern of their linkages by counting either one. In this paper the linkage pattern between cliques is used. The discussion will draw on earlier work on betweenness.

Anthonisse (1971) and Freeman (1977) have independently developed measures of the degree to which a point in a network links others. A point is said to be strictly *between* two others if it falls on all the shortest paths or geodesics connecting them. If it falls on some, but not all, of the geodesics connecting a pair it may be viewed as partially between them.

Let $g_{ij}$ be the number of geodesics connecting two points, $p_i$ and $p_j$. Then if $g_{ij}(p_h)$ is the number of those geodesics that contain point $p_h$,

$$b_{ij}(p_h)=g_{ij}(p_h)/g_{ij}$$

is the probability that a randomly selected geodesic connecting $p_i$ with $p_j$ contains

$p_h$. If we sum for all unordered pairs, the betweenness-based centrality of $p_h$ is

$$C_B(p_h) = \sum_{i>j} b_{ij}(p_h).$$

Thus, whenever $P_h$ falls on every geodesic linking $p_i$ and $p_j$, that point pair contributes 1 to the sum $C_B(p_h)$. And when $p_h$ falls on some of the geodesics linking $p_i$ and $p_j$, $C_B(p_h)$ grows proportionately to the occurrence of $p_h$ on those paths. Thus, $C_B(p_h)$ is an index of the degree to which point $p_h$ stands between and therefore links pairs of others.

In centered graphs the index $C_B(p_h)$, has some special properties. We know, from (1) above, that a centered graph falls somewhere between a $k$-star with $k-1$ edges and a complete graph with $\binom{k}{2}$ edges. To understand the meaning of $C_B(p_h)$ in this context we need to develop an hypothetical growth process for a centered graph.

Consider a $k$-star centered around point $p^*$. Essentially, such a $k$-star is a family containing $k-1$ two-point cliques linked together by $p^*$ that is a member of all the cliques. Thus, $p^*$ can be conceived of as a coordinator, solely responsible for keeping the whole thing together.

Now imagine that the $k$-star grows by the addition of edges until it becomes complete. Intuitively, it would seem that there are two ways in which such a $k$-star could grow.

(1) Additional edges could be added in such a way that they contribute primarily to the increasing sizes of the cliques connected by $p^*$. In this case, the growing ego network would continue to be a family of growing cliques still linked only through $p^*$.

(2) Additional edges could be added in such a way that the growth of cliques is minimized; instead, the new edges could contribute to the creation of an alternative center or centers for the coordination of group processes. In this case, these new centers could either *assist* or *compete with* $p^*$ in performing the coordination role.

To untangle these two possibilities we need an index of the degree to which others could compete with $p^*$. Let the *total* betweenness,

$$T_B = \sum_{i=1}^{k} C_B(p_i).$$

This is an index of the degree to which all points (including $p^*$) serve as links between pairs of others. $T_B$ can be partitioned into two components. First, we already know that $C_B(p^*)$ is the linking index of $p^*$. Then $C_B(\bar{p}^*) = T_B - C_B(p^*)$ is the linking associated with all competing others. It is the sort of index we want.

In itself, however, $C_B(\bar{p}^*)$ is not enough. Differing numbers of points in the graph, or differing numbers of extra edges (beyond the $k-1$ needed for a centered graph) will obviously provide different opportunities for others to compete. We must therefore be able to calculate $\max[C_B(\bar{p}^*)]$ in order to give our competition index a baseline.

To do that, consider a $k$-star. By definition

(1) The number of edges, $e = k - 1$.
(2) $T_B = \binom{k-1}{2}$.
(3) $C_B(p^*) = \binom{k-1}{2}$.
(4) $C_B(\tilde{p}^*) = 0$.
(5) The number of cliques $= k - 1$.
Now let us add edges. If

$$e' = e - (k - 1)$$

is the number of extra edges, let $e' = 1$. We create a new path (like edge $x$ in Fig. 1) and directly connect a pair of points ($p_k$ and $p$) that previously had to 'use' $p^*$ to coordinate their efforts. In effect we create a new three-point clique that is linked to all the others by $p^*$. Thus, $p^*$ loses 1 unit of betweenness as does the total index for the graph, and

$$T_B = \binom{k-1}{2} - 1, \qquad C_B(p^*) = T_B \quad \text{and} \quad C_B(\tilde{p}^*) = 0,$$

since no other point is competing with $p^*$ as a link. Moreover, since two two-point cliques are joined to form a three point clique, the number of cliques is now $k - 2$.

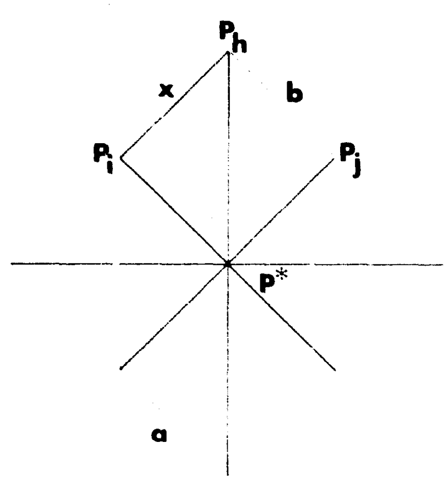


Fig. 1.

Let us, then, add a second edge;

$$e' = 2,$$

and

$$T_B = \binom{k-1}{2} - 2,$$

since *every* additional edge must connect *some* pair of points directly together that previously had to use some intermediate link or links. But what happens to $C_B(p^*)$ depends on how the new edge is placed.

If the new edge connects a pair in such a way that no new path of length $= 2$ is created (like edge $a$ in Fig. 1) it has the same effect as the first new edge. It reduces the number of cliques by one by creating a new three-point clique, and it reduces the linking of $p^*$ without creating any competing center. Thus,

$$C_B(p^*) = T_B \quad \text{and} \quad C_B(\bar{p}^*) = 0.$$

But if the new edge creates a new length = 2 path (like edge $b$ in Fig. 1), it produces a situation where a point (here $p_h$) does compete with $p^*$ by providing an alternative link (here between $p_i$ and $p_j$). In this case, still assuming as we did above that alternative links are used with equal probability,

$$C_B(p^*) = T_B - \tfrac{1}{2} \quad \text{and} \quad C_B(\bar{p}^*) = \tfrac{1}{2},$$

since $p^*$ must surrender one-half of its linking role for one pair to a competing link. Furthermore, no new clique is formed with the addition of a two-step path in this manner.

A third new edge can be added in any of three ways. Any will reduce $T_B$ by one additional unit for the reason stated above. Thus

$$T_B = \binom{k-1}{2} - 3.$$

A new edge, however, will have differing effects on $C_B(p^*)$, $C_B(\bar{p}^*)$ and the number of cliques depending upon how it is attached:

(1) Creating a new three-point clique will again reduce the number of cliques but it will have no effect on $C_B(p^*)$ except the automatic reduction of one unit resulting from adding the new edge.

(2) Creating another *new* linking point still has no effect on the number of cliques and still leads to the reduction of $C_B(p^*)$ by one-half by assigning that half unit to the new point as it did above. So, under this condition

$$C_B(p^*) = T_B - 1 \quad \text{and} \quad C_B(\bar{p}^*) = 1.$$

(3) Adding another edge to the *already established* competing link (see $p'$ in Fig. 2) creates a triple of points ($p_1$, $p_2$ and $p_3$ in Fig. 2) all of which can now reach one an-

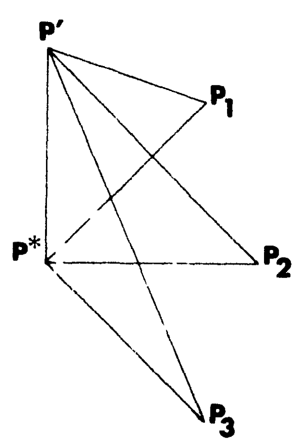


Fig. 2.

other either via $p^*$ or via the new link, $p'$. This means that for $\binom{3}{2}=3$ pairs, the new link divides the responsibility of linking with $p^*$. The betweenness associated with linking one of these was already divided when the second edge was added, but that associated with linking the additional two must be divided between $p^*$ and the competing link. In this case then, the number of cliques is constant and

$$C_B(p^*) = T_B - 1\tfrac{1}{2} \quad \text{and} \quad C_B(\bar{p}^*) = 1\tfrac{1}{2}.$$

Thus, $C_B(p^*)$ suffers the greatest loss, and $C_B(\bar{p}^*)$ experiences the greatest gain if this third strategy is used. It has a multiplier effect.

As a matter of fact, as we continue to add edges this multiplier effect becomes increasingly pronounced. A fourth edge added to an established competing linking point assigns it half of $\binom{4}{2}=6$ linking units and a fifth gives it half of $\binom{5}{2}=10$ units. It continues to grow in this way until $e' = k-2$, at which point the competitor has $k-1$ edges; it is saturated and becomes structurally identical to $p^*$.

The next series of new edges will again reduce the number of cliques by one. It will continue to generate maximum competition by allowing another new linking point to compete with the previous two so that the linking function is divided three ways. This strategy may be repeated, generating ever more competing linking points until the graph is complete. The linking function of the original $p^*$ is hacked away by competing links until, in the complete graph, all points are identical in structure and no linking at all takes place.

Generating a computer algorithm to calculate this maximum value for $C_B(\bar{p}^*)$ is straightforward. An exact solution is also possible, though somewhat complicated. One, by Karl P. Reitz is presented in Appendix A.

In any case, given $e'$ we can calculate $\max[C_B(\bar{p}^*)]$ for any centered graph. Then the degree to which points other than the designated center, $p^*$, compete for the role of coordinator can be indexed by

$$L = C_B(\bar{p}^*)/\max[C_B(\bar{p}^*)],$$

the ratio of the linking of cliques by points other than $p^*$ to its maximum. $L$ will approach one as new paths of length $=2$ using a minimal number of linking points are generated. In such cases one or more points other than $p^*$ are effectively competing for the role of coordinator, and the number of cliques is minimized.

## 3. An application to a computer conference

The data for this illustration were collected as part of a two-year study of the impact of a computer-based communications system on some members of a developing field of science (L. Freeman and S. Freeman, 1980; S. Freeman and L. Freeman, 1979; L. Freeman, 1980). The study group was a collection of specialists in social networks. Network specialists were chosen for study because they seemed to be in a position where frequent communication could facilitate their development as a collectivity. They were a mixed bag of differing kinds of people from a range of backgrounds and disciplines. Moreover, they were geographically dispersed. At the same time, however, they were engaged in the process of trying to set norms and standards for an emerging new field of study.

It is usually assumed that for a new field to develop, its founders must be in face-to-face contact (Mullins, 1973). Some sort of regular interaction seems to be necessary for the establishment of the kinds of consenses that are the foundation of a specialty. Here, the hope was that the computer could provide a viable substitute for actual physical proximity.

The computer communications system uzed was the Electronic Information Exchange System (EIES). Participants used computer terminals that were linked via phone lines to a computer in Newark, New Jersey. They could send private messages to designated others, participate in 'conferences' where the ongoing communication was permanently recorded or work in a private personal 'notebook' space. If two participants were logged on at the same time (either by chance or by arrangement) they could communicate in a line-by-line interactive mode. But most communication was asynchronous. When a participant signed on, the system typed out his or her waiting messages. A participant could respond, generate new messages, look over conference activity or sign off without fear of missing anything until he or she next signed on. Moreover, the whole communication process was monitored by the computer for the two-year duration of the experimental trials.

All in all there were 48 participants who were more or less active for some portion of the two-year trial period. They represented a range of social science disciplines and included some mathematicians and computer scientists. Some pairs had been long term colleagues and close personal friends, but many were unacquainted when the project began. Details of their interpersonal friendship ties are reported elsewhere (S. Freeman and L. Freeman, 1979; L. Freeman, 1980). In the present paper concern will be with questions involving the organization of communication in the computer milieu.

No attempt was made to impose any structure on the communication of participants in the computer trials. The principal investigator initially selected and recruited the other participants. He was assigned the tasks associated with overall project supervision. He acted as a go-between, linking participants with each other and with computer management personnel. He was responsible for allocating computer time, paying communication bills and the like. It was his job to be directly in touch with everyone else; he was the coordinator in an ego network.

The coordinator, along with four other members of the networks group started the computer communication process several weeks before the main group of participants. The main group was started in January 1978 and continued through December 1979. The data used in the present report on communication patterns were collected during the first 16 months of that period.

This analysis is restricted to data on personal exchanges by means of messages among participants. For present purposes, the 16 months are divided into four fourmonth segments. A participant was defined as active in personal communication during a segment if he or she sent at least one mesage to another participant. During the first time segment there were 30 such active communicators. In the second segment that number dropped to 25. It was 27 in the third segment and 26 in the fourth.

Rates of message sending activity varied greatly - both by individuals and by times - but here we are not concerned with rates, but with the simpler question of whether a pair of persons were in touch at all. Thus, these data show a pair of participants as symmetrically linked during a time segment if either sent one or more messages to the other during that segment. Our basic data, then, consist of four symmetrical adjacency matrices.

In their present form, these might be viewed as matrices of *non-avoidance*. Non-zero entries were assigned in a maximally generous way. Pairs were defined as linked if there was *any* indication of even attempts at communication. In this case, a zero entry in a cell of a matrix suggests an almost systematic mutual avoidance between a pair of persons.

In every period, the conference coordinator was in touch with all other active participants. Thus, all networks contained a $k$-star and were centered graphs. But, given the aim of the conference to establish group norms and consenses, success should imply a steady reduction in the importance of that role.

We are concerned here with an ego network where one person was initially designated as the coordinator or facilitator of communication among others. But the aim of the whole exercise was to help the others get organized into a problem-solving work group or groups. The coordinator's job then, was not to dominate or control, but simply to facilitate communication among others.

Results are shown in Table 1. Neither densities nor betweenness scores for $p^*$ show any systematic relation with time. Nor are they monotone with each other. Each, it seems, is confounded by extraneous effects.

Values of $L$, however, are approximately monotone decreasing with time. They show a good deal of initial competition, diminishing at the second time period and finally seeming to stabilize at times 3 and 4.

These results are consistent both with casual observations about this computer conference and with more or less standard intuitions about group processes. When the computer conference began, the four 'old timers' on the EIE System seemed to generate a 'welcome wagon' effect. As established citizens of a different kind of a

Table 1

| Time | Number of points | Density | $C_B(p^*)$ | $L$ |
|---|---|---|---|---|
| 1 | 30 | 0.428 | 0.121 | 0.45 |
| 2 | 25 | 0.282 | 0.272 | 0.40 |
| 3 | 27 | 0.252 | 0.391 | 0.31 |
| 4 | 26 | 0.350 | 0.305 | 0.32 |

'world' these participants were eager to greet newcomers and make them feel at home. Thus, in the first time period, the coordinator shared at least a part of the coordination role with these others.

This phenomenon was reduced by the start of the second time period. The remaining competition for coordination seemed to result primarily from the entry of one new - gregarious - participant. As he entered the system, this newcomer boldly announced his arrival to everyone on the system. By that time, however, most participants had settled down to a more or less consistent set of others with whom they communicated. Very few even responded, so the new arrival seemingly was quickly socialized into the standard pattern.

By the third and fourth time periods, the communication patterns had all pretty well stabilized. The coordinator was still coordinating, but most participants had formed cliques - representing either strong affective ties or common interests - and settled into a pattern of mostly within-clique communication. Some individuals other than the coordinator linked two or more cliques. This seemed to occur when an individual had an interest in the ideas being generated by more than one group. By this time, however, there was no evidence that anyone other than the official coordinator was self-consciously trying to provide coordination as such.

All this makes sense in terms of ordinary sociological intuition. The initial thrust toward broad communication seems to reflect the excitement of a new shared experience. After that the participants settled down and established regular patterns of communication leaving the coordination task to the coordinator. There was some bridging between groups exhibited by other participants, but this seemed to reflect an interest in substance rather than an interest in competing for coordination.

An overall image of the stable state that is consistent with these results is quite close to Campbell's (1969) conception of the 'fish-scale' model of scientific specialties. These network specialists have, themselves, specialized further and formed clique-like working groups. But the work groups are - to some degree - overlapping. There are joint members who link groups. As Campbell has suggested, such an arrangement is important for the systematic development of a science. These link persons who form the bridges between groups can play an important role. They are critical links in passing information between the groups. Without them, the foci of the groups could drift apart and end up as unrelated specialties.

These are the 'marginal men' in Park's (1928) terms - the ones who are subjected to the strain of living with two or more sets of ideas. Ideally, they are motivated to rationalize these separate ideas and to generate creative syntheses. Without them, there is no chance for the sort of coordination among intellectual disciplines that can lead to the development of an interrelated set of scientific specialties. Thus, it is possible that these networkers unselfconsciously found an optimum social arrangement for scientific productivity. Whether this is true or not remains to be seen.

## Appendix A[2]

Let $G$ be a graph with $k$ points and $e$ edges, and let $G$ contain a star with center $p_1$. Let $e'$ be the number of edges not contained in the star. $e'$ is the number of 'extra' edges and is given by $e'=e-(k-1)$. $C_B(p^*)$ is the betweenness associated with $p_1$ in the notation above, and $C_B(\tilde{p}^*)$ is the sum of betweenness associated with all other points. Let $B$ be the maximum $C_B(\tilde{p}^*)$ for fixed $e'$. For a fixed value of $e'$ it has been shown in the text above that the maximum betweenness $B$ occurs if the extra edges are allocated in such a manner as to saturate a sequence of points. $e'$ ranges from a minimum of 0 to a maximum of $(k-1)(k-2)/2$. This maximum value can be written in the form

$$\tfrac{1}{2}(k-1)(k-2)=(k-2)+(k-3)+\cdots+2+1.$$

Each term in the expression on the right corresponds to the number of edges needed to saturate each successive point. If the sequence of points is designated by $p_1, p_2, p_3, \ldots, p_k$ where $p_1$ is the center of the star, then it takes $k-2$ edges to saturate $p_2$, $k-3$ edges to saturate $p_3$, and so on. It only takes one edge to saturate the last two points ($p_{k-1}$ and $p_k$).

In the general case, let $\pi$ be the number of saturated points in a graph constructed as above. Since $G$ contains a star, $1\le\pi\le k$. $\pi$ is a function of $e'$. For example, if $e'=0$, then $\pi=1$ and if $e'=k-2$, $\pi=2$ and so on until $e'=(k-1)(k-2)/2$ when $\pi=k$. Define $a$ to be the number of edges that are incident to no saturated points. $a$ is also a function of $e'$. For example if $e'=1$, then $\pi=1$ and $a=1$. Similarly if $e'=(k-2)+1$, $\pi=2$ and $a=1$. In general if $e'<(k-1)(k-2)/2$ we have $e'=(k-2)+(k-2)+\cdots+(k+\pi)+a$ where $0\le a<k-\pi-1$. Using this relationship we have the following theorem.

**Theorem A.1.** *Given $k$, $e'$, $\pi$ and $a$ as defined above*

$$\pi=\operatorname{int}\left((k-\tfrac{1}{2})-\sqrt{(k-\tfrac{1}{2})^2-2(k-1+e')}\right)$$

*and*

$$a=\tfrac{1}{2}\pi^2+(\tfrac{1}{2}-k)\pi+(k-1)+e'$$

*where* int *is the greatest integer value function.*

**Proof.** If $\pi$ is the number of saturated points and $a$ the number of edges not used in the saturation of those points, then $\pi$ is the largest integer such that

(1) $e'\ge(k-2)+(k-3)+\cdots+(k-\pi)$ and

(2) $a=e'-((k-2)+(k-3)+\cdots+(k-\pi))$.

Using a closed form expression for the sums in the above relations we have

$$e'\ge((k-2)+(k-\pi))(\pi-1)/2=-\tfrac{1}{2}\pi^2+(k-\tfrac{1}{2})\pi-(k-1)$$

[2] This part of the paper has been developed by Karl P. Reitz at the School of Social Sciences, University of California at Irvine, Irvine, CA 92717.

and $\frac{1}{2}\pi^2-(k-\pi)+(k-1+e')\geq 0$. Now define $Q(x)=\frac{1}{2}x^2-(k-\frac{1}{2})x+(k-1+e')$. Then $\pi$ is the largest integer for which $Q(\pi)\geq 0$ and $0\leq\pi\leq k-1$.

$Q$ is a quadratic function in $x$ and has a minimum value when $x=k-\frac{1}{2}$. Therefore $Q$ is decreasing on the interval $0\leq x\leq k-1$. Note also that $Q(k-1)=e'-(k-2)(k-1)/2\leq 0$. Therefore $Q$ has two real roots, the smaller of which occurs in the interval $0\leq x\leq k-1$. Since $\pi$ is the largest integer such that $Q(\pi)\geq 0$, $\pi$ is the greatest integer less than the smallest root of $Q$. Therefore

$$\pi=\operatorname{int}\left((k-\tfrac{1}{2})-\sqrt{(k-\tfrac{1}{2})^2-2(k-1+e')}\right).$$

From (2) we have $a=e'-((k-2)+\cdots+(k-1))=e'-(-\frac{1}{2}\pi^2+(k-\frac{1}{2})\pi-(k-1))=\frac{1}{2}\pi^2+(\frac{1}{2}-k)\pi+(k-1+e')$.

From this theorem we now know that for a given value of $k$ and $e$ there are uniquely defined $\pi$ and $a$. Using these numbers we can write an expression for the betweenness of the center of the star.

**Theorem A.2.** *Let G be a graph with $k\geq 2$ points and at least $k-1$ edges. Let G be constructed so that the first $k-1$ edges form a star and any additional edges are added successively to form a sequence of saturated points. If e is the total number of edges, e' the number of edges in excess of the $k-1$ edges, $\pi$ the number of saturated points, and a the number of edges not included in the saturation of the $\pi$ points, then the maximum betweenness B of the center of the star is given by*

$$B(k,\pi,a)=\frac{(k-\pi-2)(k-\pi-1)}{2(\pi+1)}+\frac{((k-\pi)-(2-a))((k-\pi)-(a+1))}{2(\pi+1)\pi}+\frac{(k-\pi)-(a+1)}{\pi}$$

*for $\pi<k-1$ and*

$$B(k,\pi,a)=0$$

*for $\pi\geq k-1$.*

**Proof.** The betweeness of any point in a saturated graph is 0. An unsaturated graph must have at least two unsaturated points, so the number of saturated points in an unsaturated graph must be 2 or more less than the total number of points in the graph. Therefore $B=0$ for $\pi\geq k-1$. We may assume then in the remainder of the proof that $\pi<k-1$.

As $\pi$ decreases, the betweeness of $p_1$ increases, and for fixed $\pi$, as $a$ decreases, the betweeness of $p_1$ increases. Each missing edge therefore contributes to the betweeness of $p_1$. The formula for $B(k,\pi,a)$ can be thought of as the betweeness of $p_1$ in the event of $\pi+1$ saturated points plus the betweeness contributed to $p_1$ by edges missing in the full saturation of the $(\pi+1)$st point.

If there are $\pi+1$ saturated points, with no additional edges, there will be $k-(\pi+1)$ points with edges only to $p_1,\ldots,p_{\pi+1}$. Each pair of unsaturated points is connected by exactly $\pi+1$ geodesics of length 2. Each of these pairs contributes $1/(\pi+1)$ to the betweeness of $p_1$. There are $(k-\pi-1)(k-\pi-2)/2$ such pairs, so the total betweeness of $p_1$ is $(k-\pi-1)(k-\pi-2)/2(\pi+1)$. Note that the formula will reduce to this expression in the case that $\pi$ stays the same and $a=k-\pi-1$ or if $\pi$ is replaced by $\pi+1$ and $a=0$.

If instead of $\pi+1$ saturated points there are $\pi$ saturated points and $a$ edges added to $p_{\pi+1}$, the unsaturated points can now be partitioned into three sets. The set $S_1=\{p_{\pi+1}\}$, the set $S_2=\{p_{\pi+2},\ldots,p_{\pi+1+a}\}$ of points connected directly to $p_{\pi+1}$, and the set $S_3=\{p_{\pi+2+a},\ldots,p_k\}$ of points not connected directly to $p_{\pi+1}$. Fig. A.1. shows the different classes of missing edges.

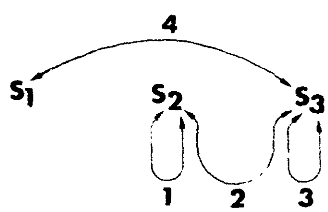


Fig. A.1.

*Case* 1. The edges missing between pairs of points from $S_2$ to $S_2$ contribute the same betweeness as if $p_{\pi+1}$ were saturated since each pair is connected by $\pi+1$ geodesics. Therefore the first term accounts for the betweeness contributed by these missing edges.

*Case* 2. There are $a$ points in $S_2$ and $k-(\pi+a+1)$ points in $S_3$. There are therefore $a(k-(\pi+a+1))$ missing edges between $S_2$ and $S_3$. The pairs of points representing these missing edges are connected by only $\pi$ goedesics of length 2. They, therefore, each contribute $1/\pi$ to the betweeness of $p_1$ instead of the amount $(1/(\pi+1))$ already included. The difference for each pair is

$$\left(\frac{1}{\pi}-\frac{1}{\pi+1}\right)=\frac{1}{\pi(\pi+1)}.$$

These pairs therefore contribute a total extra betweeness of $a(k-(\pi+a+1))/\pi(\pi+1)$ to $p_1$.

*Case* 3. As in Case 2 the missing edges from $S_3$ to $S_3$ each contribute $1/\pi$ instead of $1/\pi+1$ to the betweeness of $p_1$. There are

$$\tfrac{1}{2}(k-(\pi+a+1))(k-(\pi+a+1)-1)$$

such pairs and therefore an additional betweeness of

$$\frac{(k-(\pi+a+1))(k-(\pi+a+1)-1)}{2(\pi+1)\pi}.$$

The sum of the additional betweeness of Case 2 and Case 3 gives the second term of $B(k,\pi,a)$.

*Case* 4. If the point $p_{\pi+1}$ was saturated there would be no missing edges between $S_1$ and $S_3$. The first term does not even partially account for these missing edges. There are $k-(\pi+a+1)$ such pairs, each connected by only $\pi$ geodesics of length 2. They therefore contribute a total of $(k-(\pi+a+1))/\pi$ to the betweeness of $p_1$.

The combination of these terms results in the given formula for $B(k,\pi,a)$.

The net result of Theorems A.1 and A.2 is to give us a closed form expression for the minimum betweeness of the center of a star to which extra edges have added. This minimum betweeness is now expressed in terms of the total number of points $k$, and the number of edges $e$.

## References


J.M. Anthonisse, The rush in a graph, Mimeo, Mathematical Centrum, Amsterdam (1971).

J. Boissevain, The place of non-groups in the social sciences, Man 3 (1968) 542–556.

E. Bott, Family and Social Network (Tavistock, London, 1957).

D.T. Campbell, Ethnocentrism of disciplines and the fish-scale model of omniscience, in: M. Scherif and C. Scherif, eds., Interdisciplinary Relationships in the Social Sciences (Aldine, Chicago, 1969).

A. Epstein, The network and urban social organization, in: J.C. Mitchell, ed., Social Networks in Urban Situations (Manchester University Press, Manchester, 1969).

L.C. Feeman, A set of measures of centrality based on betweenness, Sociometry 40 (1977) 35–41.

L.C. Freeman, Q-analysis and the struture of friendship networks, Internat. J. Man-Mach. Stud. 12 (1980) 367–378.

L.C. Freeman and S.C. Freeman, A semi-visible college: structural effects on a social networks group, in: M.M. Henderson and M.J. MacNaughton, eds., Electronic Communication: Technology and Impacts, AAAS Selected Symposium 52 (AAAS, Washington, DC, 1980).

S.C. Freeman and L.C. Freeman, The networkers network: a study of the impact of a new communications medium on sociometric structure, Social Sciences Research Repts. No. 46, University of California, Irvine, CA (1979).

M.S. Granovetter, The strength of weak ties, Amer. J. Soc. 78 (1973) 1360–1380.

B. Kapferer, Norms and the manipulation of relationships in a work context, in: J.C. Mitchell, ed., Social Networks in Urban Situations (Manchester University Press, Manchester, 1969).

R.D. Luce and A. Perry, A method of matrix analysis of group structure, Psychometrika 14 (1949) 94–116.

A. Mayer, The significance of quasi-groups in the study of complex societies, in: M. Banton, ed., The Social Anthropology of Complex Societies (Praeger, New York, 1966).

J.C. Mitchell, Social Networks in Urban Situations (Manchester University Press, 1969).

N.C. Mullins, Theories and Theory Groups in Contemporary American Sociology (Harper and Row, New York, 1973).

R. Park, Human migration and the marginal man, Amer. J. Soc. 33 (1928) 881–893.

R. Speck and C. Attneave, Family Networks (Pantheon Books, New York, 1973).