Available online at www.elixirpublishers.com (Elixir International Journal)

Discrete Mathematics

Elixir Dis. Math. 45 (2012) 7927-7936

ISSN: 2229-712X

On certain centrality measures related to distance graphs

V. Yegnanarayanan[1] and V.Thamaraiselvi[2]

[1]Department of Mathematics, Velammal Engineering College, Chennai - 600 066, India.

[2]Bharathiar University, Coimbatore - 641 046, India.

ARTICLE INFO

Article history:

Received: 2 February 2012;

Received in revised form: 15 April 2012;

Accepted: 20 April 2012;

Keywords

Centrality measures,
Degree,
Closeness,
Betweenness,
Graphs,
Networks.

ABSTRACT

We have given in this paper a fairly decent introduction to the importance of the study of centrality measures of distance graphs. We have discussed at length several concepts with a number of interesting examples and highlighted interesting results. Then, the scope of practical applicability of these concepts is also indicated.

© 2012 Elixir All rights reserved.

Introduction

Centrality and Centralization

The idea of the centrality of individuals and organizations in their social networks was one of the earliest to be pursued by social network analysts. The immediate origins of this idea are to be found in the sociometric concept of the ’star’ - that person who is the most ’popular’ in his or her group or who stands at the center of attention. The formal properties of centrality were initially investigated by Bavelas (1950), and, since his pioneering work, a number of competing concepts of centrality have been proposed. As a result of this proliferation of formal measures of centrality, there is considerable confusion in the area. What unites the majority of the approaches to centrality is a concern for the relative centrality of the various vertices in the graph the question of so-called ’vertex centrality’. But from this common concern they diverge sharply. In this paper we will review a number of measures of vertex centrality, focusing on the important distinction between ‘local’ and ‘global’ vertex centrality. A vertex is locally central if it has a large number of connections with the other vertices in its immediate environment if, for example, it has a large ‘neighborhood’ of direct contacts. A vertex is globally central, on the other hand, when it has a position of strategic significance in the overall structure of the network. Local centrality is concerned with the relative prominence of a focal vertex in its neighbourhood, while global centrality concerns prominence within the whole network. Related to the measurement of vertex centrality is the idea of the overall ’centralization’ of a graph, and these two ideas have sometimes been confused by the use of the same term to describe them both. Freeman’s important and influential study (1979), for example, talks of both ’vertex centrality’ and ’graph centrality’. Confusion is most likely to be avoided if the term; centrality’ is restricted to the idea of vertex centrality, while the term ’centralization’ is used to refer to particular properties of the graph structure as a whole. Centralization, therefore, prefers not to the relative prominence of vertices, but to the overall cohesion or integration of the graph. Graphs may, for example, be more or less centralized around particular vertices or sets of vertices. A number of different procedures have been suggested for the measurement of centralization, contributing further to the confusion which besets this area. Implicit in the idea of centralization is that of the structural ‘center’ of the graph, the vertex or set of vertices around which a centralized graph is organized. There have been relatively few attempts to define the idea of the structural center of a graph, and it will be necessary to give some consideration to this.

Centrality: Local and Global

The concept of vertex centrality, originated in the sociometric concept of the ’star’. A central vertex was one which was ’at the center’ of a number of connections, a vertex with a great many direct contacts with other vertices. The simplest and most straightforward way to measure vertex centrality, therefore, is by the degrees of the various vertices in the graph. The degree, it will be recalled, is simply the number of other vertices to which a vertex is adjacent. A vertex is central, then, if it has a high degree; the corresponding agent is central in the sense of being ’well-connected’ or ’in the thick of things’. A degree-based measure of vertex centrality, therefore, corresponds to the intuitive notion of how well connected a vertex is within its local environment. Because this is calculated simply in terms of the number of vertices to which a particular vertex is adjacent, ignoring any indirect connections it may have, the degree can be regarded as a measure of local centrality. The most systematic elaboration of this concept is to be found in Nieminen (1974).

Degree-based measures of local centrality can also be computed for vertices in directed graphs, though in these situations each vertex will have two measures of its local centrality, one corresponding to its indegree and the other to its outdegree. In directed graphs, then, it makes sense to distinguish between the ’in-centrality’ and the ’out-centrality’ of the various vertices.

Tele:
E-mail addresses: prof.yegna@gmail.com

© 2012 Elixir All rights reserved

A degree-based measure of vertex centrality can be extended beyond direct connections to those at various path distances. In this case, the relevant 'neighbourhood' is widened to include the more distant connections of the vertices. A vertex may, then, he assessed for its local centrality in terms of both direct (distance 1) and distance 2 connections or, indeed, whatever cut-off path distance is chosen. The principal problem with extending this measure of vertex centrality beyond distance 2 connections is that, in graphs with even a very modest density, the majority of the vertices tend to be linked through indirect connections at relatively short path distances. Thus, comparisons of local centrality wares at distance 4, for example, are unlikely to be informative if most of the vertices are connected to most other vertices at this distance. Clearly, the cutoff threshold which is to be used is a matter for the informed judgment of the researcher who is undertaking the investigation, but distance 1 and distance 2 connections are likely to be the most informative in the majority of studies.

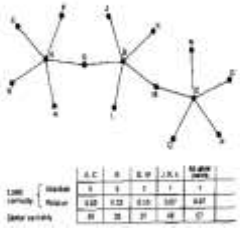

Figure 1

It is important to recognize that the measurement of local centrality does not involve the idea that there will be any unique,'central' vertex in the network. In Figure 1, for example, vertices A, B and C can each be seen as local centers: they each have a degree of 5, compared with degrees of 1 or 2 for all other vertices. Even if vertex A had many More direct connections than vertices B and C it would not be 'the' center of the network: it lies physically towards one 'side' of the chain of vertices, and its centrality is a purely 'local' phenomenon. The degree, therefore, is a measure of local centrality, and a comparison of the degrees of the various vertices in a graph can show how well connected the vertices are with their local environments. This measure of local centrality has, however, one major limitation. This is that comparisons of centrality mores can only meaningfully be made the members of the same graph or between graphs which are the same size. The degree of a vertex depends on, among other things, the size of the graph, and so measures of local centrality cannot be compared when graphs differ significantly in size. The use of the raw degree score may, therefore, be misleading. A central vertex with a degree of 25 in a graph of 100 vertices, for example, is not as central as one with a degree of 25 in a graph of 30 vertices, and neither can be easily compared with a central vertex with a degree of 6 in a graph of 10 vertices, In an attempt to overcome this problem, Freeman (1979) has proposed a relative measure of local centrality in which the actual number of connections is related to the maximum number which it could sustain. A degree of 25 in a graph of 100 vertices, therefore, indicates a relative local centrality of 0.25, while a degree of 25 in a graph of 30 vertices indicates a relative centrality of 0,86, and a degree of 6 in a graph of 10 vertices indicate a relative centrality of 0.66.' Figure 1 shows that relative centrality can also t-e used to compare vertices within the same network, It should also be clear that this idea can be extended to directed graphs. A relative measure, therefore, gives a far more standardized approach to the measurement of local centrality.

The problem of comparison which arises with raw degree measures of centrality is related to the problem of comparing densities between different graphs, which was discussed in the previous chapter. Both are limited by the question of the size of the graphs. It will be recalled, however, that the density level also depends on the type of relation that is being analyzed. The density of an 'awareness' network, I suggested, would be higher than that of a 'loving' network. Because both density and vertex centrality are computed from degree measures, exactly the same considerations apply to measures of vertex centrality. Centrality measured in a loving network, for example, is likely to be lower, other things being equal, than centrality in an awareness network. Relative measures of vertex centrality do nothing to help with this problem. Even if local centrality scores are calculated in Freeman's relative terms, they should be compared only for networks which involve similar types of relations. Local centrality is, however, only one conceptualization of vertex centrality, and Freeman (1979, 1990) has proposed a measure of global centrality based around what he terms the 'closeness' of the vertices. Local centrality measures, whatever path distance is used, are expressed in terms of the number or proportion of vertices to which a vertex is connected. Freeman's measure of global centrality is expressed in terms of the distances among the various vertices. It will be recalled that two vertices are connected by a path if there is a .sequence of distinct lines connecting them, and the length of a path is measured by the number of lines which make it up. In graph theory, the length of the shortest path between two vertices is a measure of the distance between them. The shortest distance between two vertices on the surface of the earth lies along the geodesic which connects them, and, by analogy, the shortest path between any particular pair of vertices in a graph is termed a 'geodesic'. A vertex is globally central if it lies at short distances from many other vertices. Such a vertex is 'close' to many of the other vertices in the graph.

The simplest notion of closeness is, perhaps, that calculated from the 'sum distance', the sum of the geodesic distances to all other vertices in the graph (Sabidussi, 1966). If the matrix of distances between vertices in an undirected graph is calculated, the sum distance of a vertex is its column or row sum in this matrix (the two values are the same). A vertex with a low sum distance is 'close' to a large number of other vertices, and so closeness can be seen as the reciprocal of the sum distance. In a directed graph, of course, paths must be measured through lines which run in the same direction, and, for this reason, calculations based on row and column sums will differ. Global centrality in a directed graph, then, can be seen in terms of what might be termed 'in-closeness' and 'out-closeness'. The Table 1 compares a sum distance measure of global centrality with degree-based measures of absolute and relative local centrality. It can be seen that A, B and C are equally central in local terms, but that B is more globally central than either A or C. In global terms, G and M are less central than B, but more central than the locally central vertices A and C. These distinctions made on the basis of the sum distances measure; therefore, confirm the impression gained from a visual inspection of the graph. This is also apparent in the measures for the less central vertices. All the remaining vertices have a degree of 1, indicating low local centrality, yet the sum distance measure clearly brings out the fact that J, K and L are more central in global terms than are the other vertices with degree 1.

Freeman (1979) adds yet a further concept of vertex centrality, which he terms the betweenness. This concept measures the extent to which a particular vertex lies ’between’ the various other vertices in the graph: a vertex of relatively low degree may play an important ’intermediary’ role and so be very central to the network. Vertices G and M in Figure 1, for example, lie between a great many pairs of vertices. The betweenness of a vertex measures the extent to which an agent can play the pan of a ’broker’ or ’gatekeeper’ with a potential for control over others. G could, therefore, be interpreted as an intermediary between the act of agents centered on B and that centered around A, while M might play the same role for the sets of B and C. Freeman’s approach to betweenness is built around the concept of ’local dependency’. A vertex is dependent upon another if the paths which connect it to the other vertices pass through this vertex. In Figure 1, for example, vertex E is dependent on vertex A for access to all other parts of the graph, and it is also dependent, though to a lesser extent, on vertices G, B, M and C. Betweenness is, perhaps, the most complex of the measures of vertex centrality to calculate. The ’betweenness proportion’ of a vertex Y for a particular pair of vertices X and Z is defined as the proportion of geodesics connecting that pair which pass through Y it measures the extent to which Y is ’between’ X and Z 3 ’Me ’pair dependency’ score of vertex X on vertex Y is then defined as the sum of the betweenness proportions of Y for all pairs which involve X. The ’local dependency matrix’ contains these pair dependency scores, the entries in the matrix showing the dependence of each row element on each column element. The overall ’betweenness’ of a vertex is calculated as half the sum of the values in the columns of this matrix, i.e., half the sum of all pair dependency scores for the vertices represented by the columns. Despite this rather complex calculation, the measure is intuitively meaningful, and it is easily computed with the UCINET and GRADAP programs. In Freeman’s work, then, can be found the basis for a whole family of vertex centrality measures: local centrality (degree), betweenness and global centrality (closeness), I have shown how comparability between different social networks can be furthered by calculating local centrality in relative rather than absolute terms, and Freeman has made similar proposals for his other measures of centrality. He has produced his own relative measure of betweenness, and he has used a formula of Beauchamp (1965) for a relative closeness measure.

All these measures, however, are based on raw scores of degree and distance, and it is necessary to turn to Bonacich (1972, 1987) for an alternative approach which uses weighted scores. Bonacich holds that the centrality of a particular vertex cannot be assessed in isolation from the centrality of all the other vertices to which it is connected. A vertex which is connected to central vertices has its own centrality boosted, and this, in turn, boosts the centrality of the other vertices to which it is connected (Bonacich, 1972). There is, therefore, an inherent Circularity involved in the calculation of centrality. According to Bonacich, the local centrality of vertex *i* in a graph, *ci*, is calculated by the formula rijcj, where rij is the value of the line connecting vertex *i* and vertex *i* and cj is the centrality of vertex j. That is to say, the centrality of i equals the sum of its connections to other vertices, weighted by the centrality of each of these other vertices.

Bonacich (1987) has subsequently generalized his initial approach, did Freeman, to a whole family of local and global measures. The most general formula for centrality, he argued, is ci = rijcj (+ cj). In this formula, the centrality weighting is itself modified by the two parameters; and, is introduced simply as an arbitrary standardizing constant which ensures that the final centrality measures will vary around a mean value of 1, on the other hand, is of more substantive significance. It is a positive or negative value which allows the researcher to set the path distances which are to be used in the calculation of centrality. Where is set as equal to zero, no indirect links are taken into account, and the measure of centrality is a simple degree-based measure of local centrality. Higher levels of increase the path length, so allowing the calculation to take account of progressively more distant connections. Bonacich claims that measures based on positive values of correlate highly with Freeman’s measure of closeness. A major difficulty with Bonacich’s argument, however, is that the values given to 0 are the results of arbitrary choices made by researchers. It is difficult to know what theoretical reasons there might be for using one 0 level rather than another.

While the original Bonacich measure may be intuitively comprehensible, the generalized model is more difficult to interpret for values of P which are greater than zero. On the other hand, the suggestion that the value of 0 can be either positive or negative does provide a way forward for the analysis of signed graphs. Bonacich himself suggests that negative values correspond to ’zero-sum’ relations, such as these involved in the holding of money and other financial resources. Positive values, on the other hand, correspond to ’non-zero-sum’ relations, such as those involving access to information. I have discussed centrality principally in terms of the most central vertices in a graph, but it should be clear that centrality scores also allow the least central vertices to be identified. Those vertices with the lowest centrality, however this is measured, can be regarded as the peripheral vertices of the graph. This is true, for example, for all the vertices in Figure 1 which have degree 1. They are locally peripheral in so far as they are loosely connected into the network. The global centrality scores in Figure 1, however, show that vertices J, K and L are not as globally peripheral as the other vertices with degree 1.

Centralization and Graph Centers

I have concentrated, so far, on the question of the centrality of particular vertices. But it is also possible to examine the extent to which a whole graph has a centralized structure. The concepts of density and centralization refer to differing aspects of the overall ’compactness’ of a graph. Density describes the general level of cohesion in a graph; centralization describes the extent to which this cohesion is organized around particular focal vertices. Centralization and density, therefore, are important complementary measures.

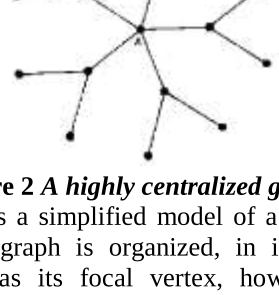

Figure 2 ***A highly centralized graph***

Figure 2 shows a simplified model of a highly centralized graph: the whole graph is organized, in important respects, around vertex A as its focal vertex, how is this level of centralization to be measured? Freeman (1979) has shown how measures of vertex centrality can be converted into measures of the overall level of centralization which is found in different graphs. A graph centralization measure is an expression of how tightly the graph is organized around its most central vertex. Freeman's measures of centralization are attempts to isolate the various aspects of the simplified notion of centralization. On this basis, he identifies three types of graph centralization, rooted in the varying conceptions of vertex centrality which Freeman has defined. The general procedure involved in any measure of graph centralization is to look at the differences between the centrality scores of the most central vertex and those of all other vertices. Centralization, then, is the ratio of the actual sum of differences to the maximum possible sum of differences. The three different ways of operationalizing this general measure which Freeman discusses follow from the use of one or other of the three concepts of vertex centrality. Freeman (1979) shows that all three measures vary from 0 to I and that a value of I is achieved on all three measures for graphs structured in the form of a 'star' or 'wheel'. He further shows that a value of 0 is obtained on all three measures for a 'complete' graph. Between these two extremes lie the majority of graphs for real social networks, and it is in these cases that the choice of one or other of the measures will be important in illuminating specific structural features of the graphs. A degree-based measure of graph centralization, for example, seems to be particularly sensitive to the local dominance of vertices, while a betweenness-based measure is rather more sensitive to the 'chaining' of vertices.

Assessing the centralization of a graph around a particular focal vertex is the starting vertex for a broader understanding of centralization. Measures of centralization can tell us whether a graph is organized around its most central vertices, but they do not tell us whether these central vertices comprise a distinct set of vertices which cluster together in a particular part of the graph. The vertices in the graph which are individually most central, for example, may be spread widely through the graph, and in such cases a measure of centralization might not be especially informative. It is necessary, therefore, to investigate whether there is an identifiable 'structural center' to a graph. The structural center of a graph is a single vertex or a cluster of vertices which, like the center of a circle or a sphere, is the pivot of its organization, This approach to what might be called 'nuclear centralization' has been outlined in an unpublished work of Stokman and Snijders,,' Their approach is to define the set of vertices with the highest vertex centrality scores as the 'center' of the graph. Having identified this set, researchers can then examine the structure of the relations between this set of vertices and all other vertices in the graph. A schematic outline of the Stokman and Snijders approach is shown in Figure 3.

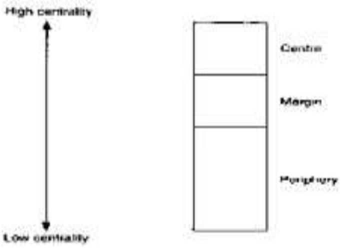

Figure 3

If all the vertices in a graph are listed in order of their vertex centrality - Stokman and Snijders use local centrality then the set of vertices with the highest centrality is the center. The boundary between the center and the rest of the graph is drawn wherever there appears to be a 'natural break' in the distribution of centrality scores. The decrease in the centrality score of each successive vertex may, for example, show a sharp jump at a particular vertex in the distribution, and this is regarded as the boundary between the center and its 'margin'. The margin is the set of vertices hich clusters close to the center and which is, in turn, divided from the 'peripheral' vertices by a further break in the distribution of centrality scores. The Stokman and Snijders concept applies only to highly centralized graphs. In a graph such as that in Figure 2, which is centralized around a particular set of central vertices, as measured by one of Freeman's indicators. It may be cry informative to try to identify the sets defined by Stokman and Snijders, though there will be an inevitable arbitrariness in identifying the boundaries between center, margin and periphery.

A solution to both of these problems, though not one pursued by Stokman and Snijders, is to use some kind of clique or cluster analysis to identify the boundaries of the structural center if the most central vertices, for example, constitute a clearly defined and well-bounded 'clique', then it may make sense to regard them as forming the nuclear center of the graph. But not all graphs will have such a hierarchical structure of concentric sets. Where the central vertices do not cluster together as the nucleus of a centralized graph, the Stokman and Snijders 'centre' will constitute simply a set of locally central, though dispersed, vertices. In such circumstances, it is not helpful to use the term 'center'. It is possible to extend the analysis of centralization a little further by considering the possibility that there might be an 'absolute center' to a graph. The absolute center of a graph corresponds closely to the idea of the center of a circle or a sphere; it is the focal vertex around which the graph is structured. The structural center, as a set of vertices, does not meet this criterion. The absolute center must be a single vertex. The center of a circle, for example, is that unique place which is equidistant from all vertices on its circumference. By strict analogy, the absolute center of a graph ought to be equidistant from all vertices in the graph. This idea is difficult to operationalize for a graph, and a more sensible idea would be to relax the criterion of equidistance and to use, instead, the idea of minimum distance. That is to say, the absolute center is that vertex which is 'closest' to all the other vertices in terms of path distance.

Christofides has suggested using the distance matrix to conceptualize and compute the absolute center of a graph. The beat step in his argument follows a similar strategy to that used by Freeman to measure 'closeness'. Having constructed the distance matrix, which shows the shortest path distances between each pair of vertices, he defines the eccentricity, or 'separation', of a vertex as its maximum column (or row) entry in the matrix,' The eccentricity of a vertex, therefore, is the length of the longest geodesic incident to it. Christofides's first approximation to the idea of absolute centrality is to call the vertex with the lowest eccentricity the absolute center. Vertex B in sociogram (i) of Figure 4 has an eccentricity of 1, and all the other vertices in the graph have eccentricity 2. In this sociogram, then, vertex B, with the lowest eccentricity, is the absolute center. In other graphs, however, there may be no single vertex with minimum eccentricity. There may be a number of vertices with equally low eccentricity, and in these circumstances a second step is needed. This second step in the identification of the absolute center involves searching for an imaginary vertex which has the lowest possible eccentricity for the particular graph. The crucial claim here is that, while the absolute center of a graph will be found on one of its constituent paths, this place may not correspond to any actual vertex in the graph. Any graph will have an absolute center, but in some graphs this center will be an imaginary rather than an actual vertex.

This claim is not so strange as it might at first seem. All the points in sociogram (ii) in Figure 5.4 base eccentricity 2, and so all are equally 'central'. It is possible, however, to conceive of an imaginary point, Z, which is mid-way between points A and B, as in sociogram (iii), 'Point' Z is distance 0.5 from both A and B, and it is distance 1.5 from points C, D and E. The artificial point Z is more central than any of the actual points, as its eccentricity is 1.5.

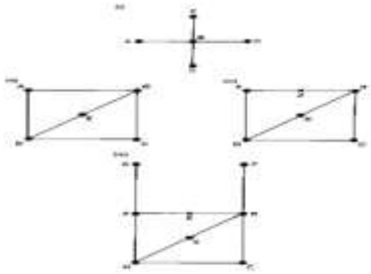

Figure 5.4 ***The absolute center of a graph***

But it is still not possible to find a single absolute centre for this sociogram. The imaginary point Z could, in fact, have been placed at the midpoint of any of the lines in the sociogram with the same results, and there is no other location for the imaginary point which would not increase its minimum eccentricity. The best that can be said for this graph, therefore, is that there are six possible locations for the absolute centre, none of which corresponds to an actual point. Moving to the second step of searching for an imaginary point as the absolute centre, then, will reduce the number of graphs for which there is no unique absolute centre, but it does not ensure that a single absolute centre can be identified for all graphs. Thus, some graphs will have a unique absolute centre, while others will have a number of absolute centers. Christofides provides an algorithm which would identify, through iteration, whether a graph contains a mid-point or actual point which is its unique absolute centre. “In sociogram (iv) of Figure 5.4, for example, there is a unique absolute centre. Its 'point' Z has an eccentricity of 1.5, compared with eccentricity scores of 2.5 for any other imaginary mid-point, 2 for points A and B, and 3 for points C, D, E, F and G.

“A Digression on Absolute Density"

The problem with the existing measures of density, as I showed in the previous chapter, is that they are size-dependent. Density is a measure which is difficult to use in comparisons of graphs of radically different sizes. Density is relative to size. This raises the question of whether it might not be possible to devise a measure of absolute density which would be of more use in comparative studies. I cannot give a comprehensive answer to that question here, but the idea of the absolute centre of a graph does raise the possibility that other concepts required for a measure of absolute density might be formulated along similar lines. A concept of density modelled on that used in physics for the study of solid bodies, for example, would require measures of 'radius', 'diameter' and 'circumference', all of which depend on the idea of the absolute centre. The radius of a circular or spherical object is the distance from its center to its circumference, on which are found its most distant reachable points. Translating this into graph theoretical terms, the eccentricity of the absolute centre of a graph can be regarded as the 'radius' of the graph. The 'diameter' of a graph, as will be shown in the following chapter, is defined as the greatest distance between any pair of its points. In sociogram (iv) of Figure 5.4, for example, the radius is 1.5 and the diameter is 3. In this case, then, the diameter is equal to twice the radius, as would be the case in the conventional geometry of a circle or a sphere. This ill not, however, be true for all graphs.

In geometry there is a definite relationship between the area and the volume of a body, these relationships being generalizable to objects located in more than three dimensions. The area of a circle is wr' and the volume of a sphere is 4,r3/3, where w is the ratio of the circumference to the diameter. 'Me general formula for the area of a circle, therefore, is c,21d, and that for the volume of a sphere is 4cr'13d, where c is the circumference, r is the radius and d is the diameter. Applying this to the simple sociogram (iv) of Figure 5.4 would show that it has a volume of 4c(I.5)/9, or 1.5c.'3 But what value is to be given to c in this formula? If the diameter of a graph is taken to be the length of the geodesic between its most distant points (the longest geodesic), the circumference might most naturally be seen as the longest possible path in the graph. In sociogram (iv), this is the path of length 5 which connects point G to point F. Thus, the 'volume' of the example sociogram is 7.5. Relatively simple geometry has, therefore, enabled us to move a pan of the way towards a measure of the absolute density of a graph in three dimensions. Density in physics is defined as mass divided by volume, and so to complete the calculation a measure of the 'mass' of a graph is required. Mass in physics is simply the amount of matter that a body contains, and the most straightforward graph theoretical concept of mass is simply the number of lines that a graph contains. In sociogram (iv) there are eight lines, and so its absolute density would be 817.5, or 1.06.

Generalizing from this case, it can be suggested that the absolute density of a graph is given by the formula 11(4c,'13d), where I is the number of lines. Unlike the relative density measure discussed in the previous chapter, this formula gives an absolute value which can be compared for any and all graphs, regardless of their size. But one important reservation must be entered: the value of the absolute density measure is dependent on the number of dimensions in which it is measured. The absolute density measure given here has been calculated for graphs in three dimensions. The concept could be generalized to higher dimensions, by using established formulae for 'hyper-volumes', but such an approach would require some agreement about how to determine the dimensionality of a graph.

Bank Centrality in Corporate Networks

Studies of interlocking directorships among corporate enterprises are far from new, but most of the studies which had been carried out prior to the 1970s had made little use of the formal technique, of social network analysis. Despite some limited use of density measures and cluster analysis, most of these studies twit a strictly quantitative approach, simply counting the numbers of directorships and interlocks among the companies. Levine's influential paper (1972) marked a shift in the direction f this research while, at about the same time, Mokken and his associates in the Netherlands began a pioneering study in the systematic use of graph theory to explore corporate interlocks (Helmets et at., 1975). The major turning point, however, occurred in 1975, when Michael Schwartz and his students presented their major conference paper which applied the concept of centrality to corporate networks (Bearden et a]., 1975). This long paper circulated widely in cyclostyled form and, despite the fact that it remains unpublished, it has been enormously influential. The work of Schwartz's group, and that which it has stimulated, provides a compelling illustration of the conceptual power of the idea of point centrality

Michael Schwartz and Peter Mariolis had begun to build a database of top American companies during the early 1970s, and their efforts provided a pool of data for many subsequent studies (see, for example, Marioiis, 1975; Sonquist and Koenig, 1975). They gradually extended the database to include the top 500 industrial and the top 250 commercial and financial companies operating in the United States in 1962, together with all new entrants to this 'top 750' for each successive year from 1963 to 1973. The final database included the names of all the directors of the 1131 largest American companies in business during the period 1962-73; a total of 13,574 directors. This database is, by any standard, that for a large social network, As such, it lends itself to the selection of substantial sub-sets of data for particular years. One such sub-set is the group of the 797 top enterprises of 1969 which were studied by Mariolis (1975).

The path-breaking paper of Schwartz and his colleagues (Bearden et at., 1975) drew on the Schwartz Marietta database, and it analysed the data using Granovetter's (1973) conceptual distinction between strong and weak ties. The basis of their argument was that those interlocks which involved the full-time executive officers the enterprises could be regarded as the 'strong' ties of the corporate network, while those high involved only the part-time non-executive directors were its 'weak' ties. The basis of this theoretical claim was that the interlocks which were carried by fulltime executive officers were the most likely board-level links to have a strategic salience for the enterprises concerned. For this reason, they tended to be associated with intercorporate shareholdings and trading relations between the companies." Interlocks created by non-executive directors, on the other hand, involved less of a time commitment and so had less strategic significance for the enterprises concerned,

The top enterprises were examined for their centrality, using Bonacich's (1972) measure. This, it will be recalled, is a measure in which the centrality of a particular point could be measured by a combination of its degree, the value of each line incident to it, and the centrality of the other points to which it is connected. This is a ,recursive', circular measure which, therefore, requires a considerable amount of computation. A network containing 750 enterprises, for example, will require the solution of 750 simultaneous equations. The first step in Bearden et al.'s analysis was to decide on an appropriate measure for the value of the lines which connected the enterprises. For the weak, undirected lines, Bearden et al. held that the value of each should be simply the number of separate interlocks, weighted by the sizes of the two boards. This weighting rested on the supposition that having a large number of interlocks was less significant for those enterprises with large boards than it was for those with small boards. The formula used in the calculation was bij/SQRT(didj), where bij is the number of interlocks between the two companies i *and* j, and di and dj are the sizes of their respective boards. This formula allows Bonacich's centrality measure to be calculated on the basis of all the 'weak ties' in the graph.

A more complex formula was required to measure centrality in terms of the strong ties. In this case, the measure of the value of each line needed to take some account of the direction which was attached to the lines in the graph. For those companies which were the 'senders' of lines (the 'tails', in the terminology of the GraDAP program) the value of the lines was calculated by the number of directors 'sent', weighted by the board size of the 'receiving' company. The attempt in this procedure was to weight the line by the salience of the interlock for the receiving board. Conversely, for those companies which were the 'receivers' of interlocks (the 'heads'), the number of directors received was weighted by the sender's board size, 16 For the final calculation of centrality scores, Bearden el al. introduced a further weighting. Instead of taking simply the raw weighted scores for the tails and the heads, they took 90 per cent of the score for the senders and 10 per cent of the score for the recipients. The reasoning behind this weighting of the scores was the theoretical judgement that, in the world of corporate interlocking, it is 'more important to give than to receive': the sending of a director was more likely to be a sign of corporate power than was the receiving of a directorship. Thus, the arbitrary adjustment to the centrality scores was introduced as a way of embodying this judgement in the final results.

The Bonacich measure of centrality which was calculated for the companies in the study correlated very highly, at 0.91, with the degrees of the companies. Bearden et al. held, however, that the more complex Bonacich measure was preferable because it had the potential to highlight those enterprises which had a low degree but which were, nevertheless, connected to highly central companies. Such a position, they argued, may be of great importance in determining the structural significance of the companies in the economy.

Schwartz and his colleagues also used a further approach to centrality, which they termed 'peak analysis'. This was later elaborated by Mizruchi (1982) as the basis for an interpretation of the development of the American corporate network during the twentieth century. A point is a peak, it was argued, if it is more central than any other point to which it is connected. Mintz and Schwartz (1985) extend this idea by defining a bridge as a central point which connects two or more peaks (see Figure 5.5). They further see a 'cluster' as comprising all the direct contacts of a peak, except for those which have a similar distance I connection to another peak. Thus, peaks lie at the hearts of (heir clusters.,,

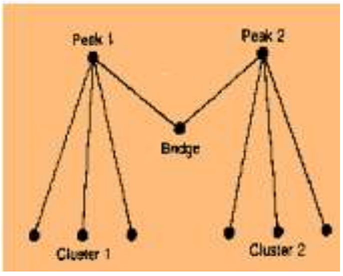

Figure ***5.5 Peaks and Bridges***

The results which were arrived at through the use of these techniques for the measurement of point centrality have become widely accepted as indicating some of the most fundamental features of intercorporate networks. In summary, Bearden et al. argued that the American intercorporate network showed an overall pattern of 'bank centrality': banks were the most central enterprises in the network, whether measured by the strong or the weak ties. Bank centrality was manifest in the co-existence of an extensive national interlock network (structured predominantly by weak ties) and intensive regional groupings (structured by the strong ties). Strong ties had a definite regional base to them. The intensive regional clusters were created by the strong ties of both the financial and the non-financial enterprises, but the strong ties of the banks were the focal centres of the network of strong ties. The intercorporate network of 1962, for example, consisted of one very large connected components two small groupings each of four or five enterprises, and a large number of pairs and isolated enterprises. Within the large connected component, there were five peaks and their associated clusters, The dominant element in the network of strong ties was a regional cluster around the Continental Illinois peak, which, with two other Chicago banks, was connected with a group of II mid-Western enterprises with extensive connections to a larger grouping of 132 enterprises. The remaining four peaks in the network of strong ties were Mellon National Bank, J.P. Morgan, Bankers Trust and United California Bank, their clusters varying in size from four to ten enterprises.

Overlying this highly clustered network of strong, regional ties was an extensive national network created by the weak ties which linked the separate clusters together. This national network, Bearden et al. argued, reflected the common orientation to business art airs and a similarity of interests which all large companies shared. Interlocks among the non-executive directors expressed this commonality and produced integration, unity and interdependence at the national level (see also Useem, 1984). The great majority of the enterprises were tied into a single large component in this network, most of the remainder being isolates. Banks were, once more, the most central enterprises, especially those New York banks which played a 'national' rather than a 'regional' role. It was the nonexecutive directors of the banks who cemented together the overall national network.

The graphs considered in this paper are mostly finite, simple and undirected. Many network structures in real life are not assigned by central authorities Instead; they are formed by autonomous agents who often have selfish motives. Typical examples of such networks include the Internet, where autonomous systems linked together to achieve global connection, peer-to-peer networks where peers connect to one another for online file sharing, and social networks where individuals connect to one another for information exchange and other social functions.

A key measure of importance of a node is its betweenness centrality, which is introduced originally in social network analysis. Betweenness with path length constraint is reasonable in real world scenarios. In peer-to-peer networks, query requests are searched only on nodes with a short graph distance away from the query initiator. In social networks, researches show that short connections are much more important than long range connections. In a decentralized network with autonomous agents, each agent may have incentive to maximize its betweenness in the network. For example, in computer networks and peer-to-peer networks, a node in the network may be able to change the traffic that it helps relaying, in which case the revenue of the node is proportional to its betweenness in the network. So the maximization of revenue is consistent with the maximization of the betweenness. In social network, an individual may want to gain or control, the most amount of information traveling in the network by maximizing her betweenness. So it would be interesting to suggest a formation of a network, in which every node is a selfish agent who decides which other nodes to be connected to in order to maximize its own betweenness. Building connections with other nodes incur costs. Each node has a budget such that the cost of building its connections cannot exceed its budget. Keeping this in mind, we have developed or suggested some new constructions out of a graph with given number of central vertices. Several interesting properties are derived and some of them may appear to be of mere academic interest. We are quite inclined to apply them to real-life networks and hopefully it will be reported as a continuation of this elsewhere.

Measures of Centrality

The centrality of a node in a network is a measure of the structural importance of the node. A person’s centrality in a social network affects the opportunities and constraints that they face. There are three important aspects of centrality: degree, closeness and betweenness.

Degree Centrality

It is simply the number of nodes that a given node is connected to. In general, the greater a person’s degree, the more potential influence they have on the network and vice-versa. For example in a gossip network, a person who has more connections can spread information more quickly and will also be more likely to hear more stuff. This can be both good and bad. In a sexual network, high degree centrality implies higher risks of disease. The greater a people degree, the greater the chance that they will catch whatever is flowing through the network, whether good or bad.

Closeness Centrality

Closeness centrality is defined as the total graph theoretic distance to all other nodes in the network. The larger the number the less central they are (because they are farther away from everyone). When a node has a low closeness score (i.e., is highly central), it tends to receive anything flowing through the network very quickly. This is because the speed with which something spreads in a network is a function of the number of links in the paths traversed. Since nodes with low closeness scores are close to all nodes, they receive things quickly. Once again, whether this is good or bad depends on the situation. In the case of information about what is happening in the company, this is usually good. In the case of a new disease that is spreading, it is very bad to be one of the first people to get it (because doctors have not worked out a treatment yet).

Betweenness Centrality

It is defined as the number of geodesic paths that pass through a node. It is the number of ’times’ that any node needs to pass through a given node to reach any other node by the shortest path. In a diffusion process, a node that has betweenness can control the flow of information, acting as a gatekeeper. This is the classic role played by the executive secretary, who can acquire a great deal of unofficial power this way. In a network of friendship relations, say, among top players in a personal computer field, the node with high betweenness can serve as a liasion between disparate regions of the network. For example, Bill Gates, head of Microsoft, is part of a certain cycle of friends. Larry Ellison, head of Oracle, is a part of a different cycle of friends Bill and Larry violently do not get along. Whenever cooperation is needed between Microsoft and Oracle, as in developing standards for network computers it has to be arranged by third party that has tie up with both camps.

We can think of betweenness as a measure of the extent to which a node is in a position to exploit many structural holes. A structural hole in a graph is a network with lack of connection between two nodes. A third party that is connected to the two unconnected nodes can sometimes exploit the situation. There are two generic benefits to being in the middle. One is the information benefit from being plugged into different camps or regions of the network. If all your ties are to one group of persons who are all interconnected (a clique), all you ever hear is the same information being recirculated. The other is the control benefit of being able to play one person against the other. If ego is a woman that alter 1 and alter 2 are counting, ego can let alter 1 know that alter 2 is buying her an expensive ring for her birthday. This may lead to alter 1 trying to top that with an even better gift.

Interpretation in terms of Graph Concepts

Moxley and Moxley[4] raised an important problem with respect to the measurement of centrality in social networks. They were concerned with measuring centrality in the large, often unconnected networks encountered in natural settings. The problem, as they defined it, was that the classical centrality measures of Bavelas[1,2], Beauchamp[3] and Sabidussi[5] could not be used for unconnected network. In each of these measures the centrality of a vertex is a function of the sum of the minimum distances between that vertex and all others. Since all distance sums are infinite in unconnected networks, these measures are useful only in settings where connectivity can be assured. Moxley's proposed solution for this problem was both arbitrary and ad hoc. They suggested that unconnected vertices be connected by an imaginary path with a length greater than that linking any pair of connected vertices in the network. The result is a crude ranking of the centrality of vertices and no index whatsoever of the overall centrality of the entire network. Moreover, since the rankings themselves are an artifact of a series of nonexistent connections, it is difficult to imagine what they might mean in terms of human communication.

The earliest intuitive conception of vertex centrality in communication was based upon the structural property of betweenness.

According to this view, a vertex in a communication network is central to the extent that it falls on the shortest path between pair of other vertices. This idea of vertex centrality was introduced by Bavelas[1,2] in his first paper on the subject. He suggested that when a particular person in a graph is strategically located on the shortest communication path connecting pairs of others, that person is in a central position. Other members of the network were assumed to be responsive to persons in such central positions who could influence the group by "withholding information".

A vertex is considered to be central here, to the degree that, it falls between other vertices on their shortest paths. A vertex falling between two others can facilitate, block, distort or falsify communication between the two; it can more or less completely control their communication.

But if it falls on some but not all of the shortest paths connecting a pair of vertices, its potential for control is limited. All this suggests that we need to generalize the graph theoretical notion of betweeness. Given a vertex, *vk* in a graph and an unordered pair of vertices, $\{v_i; v_j\}$ where $i \neq j \neq k;$ we can define the partial betweeness $b_{ij}(v_k)$, of v_k with respect to $(v_i; v_j)$in the following way.

If *vi* and v_j are not reachable from each other, v_k is not between them, so in that case let $b_{ij}(v_k) = 0:$ If *vi* and v_j are reachable, assume that they are indifferent with respect to the routing of their communication among alternative shortest paths. Thus, the probability that a message passes through any particular shortest path among alternatives is equal to $\{1=g_{ij}\}$ where g_{ij} = the number of shortest path links between *vi* and *vj* . The potential of a vertex v_k for control of information passing between *vi* and *vj* then may be defined as the probability that *vk* falls on a randomly selected shortest path connecting *vi* and *vj* . If $g_{ij}(v_k)$ = the number of shortest paths linking *vi* and v_j that contain v_k, then $b_{ij}(v_k) = (g_{ij}(v_k))=(g_{ij})$is the probability we seek. $b_{ij}(v_k)$ is the probability that a vertex v_k falls on a randomly selected shortest path linking *vi* with *vj* .

To determine the overall centrality of a vertex, v_k, we need merely to sum its partial betweeness values for all unordered pairs of vertices where $i \neq j \neq k$. $C_B(v_k) = \Sigma^n{}_{i<j}\Sigma n b_{ij}(v_k)$, where n = the number of vertices in the graph. The sum $C_B(v_k);$ is an index of the overall partial betweenness of a vertex v_k. Whenever v_k falls on the only shortest path connecting a pair of vertices, $C_B(vk)$ is increased by 1. When there are alternative shortest paths, $C_B(v_k)$ is increased in proportion to the frequency of occurrence of v_k among those alternatives. Both locating and counting shortest paths becomes tedious and difficult as the networks increase in size. $C_B(v_k)$ indexes the potential of a vertex for control by counting its opportunities for control.

It is the simplest and in many cases probably the most useful betweenness based measure of centrality. Since $C_B(v_k)$ is essentially a count, its magnitude depends upon two factors. (i) the arrangement of edges in the graph that define the location of v_k with respect to shortest paths linking pairs of vertices; and ii) the number of vertices in the graph. For certain classes of substantive problems it is desirable to create a measure that eliminates the impact of the number of vertices from the measure. For example, consider a vertex v_i in a graph containing five vertices. Let us say v_i has a value, $C_B(v_i) = 6$. On the other hand, assume a vertex, v_j, in a graph of 25 vertices, where $C_B(v_j) = 6$. Both v_i and v_j have the same potential for control in absolute terms, they can facilitate or inhibit the same number of communications. However, they differ markedly in their relative potential for control within their respective networks. v_i can dominate more than half of the communications between pairs of vertices in its graph, while v_j can control only slightly more than one percent.

To the degree that this potential for control is perceived as relative by participants in networks, v_i and v_j are in quite difficult positions with respect to centrality. What is needed in this context is a measure that is relative to its maximum value in terms of the number of vertices in the graph. Consider *S* a totally disconnected graph with n = the number of vertices ($n \geq 3$) and $m = 0$, the number of edges. For such a graph let $r = 0$, the number of unordered pairs, $\{v_i; v_j\}$, where *vi* and v_j are mutually reachable and $C_B(v_k) = 0$ the centrality index of a vertex v_k. Now if we add an edge to $S; m = 1$ and $r = 1;$ but still $C_B(v_k) = 0$, since with only one edge, no vertex can fall on a path between any others. However, when we add second edge and let $m = 2;$ it can be added either such that $r = 3$ and $C_B(v_k) = 1$ for a vertex if there is a connection with the previous edge as in the graph P_3 U $K_1;$ where $P_3 = v_i v_k v_j$ and $K_1 = v_h$ or such that $r = 2$ and $C_B(v_k) = 0$ for all vertices as in in the graph $2K_2;$ where the first copy of K_2 is, $K_2 = v_i v_j$ and the second copy of K_2 is $K_2 = v_h v_k:$ The former case, then shows a vertex v_k, that falls on a path between v_i and v_j .

This is the most central graph possible with $m = 2$. When successive new edges are added, maximum centrality is maintained only if all new edges are connected to the center vertex, v_k. This will be true until there are (n-1) edges linking v_k with every other vertex in *S*. Under these conditions each vertex is reachable from all others either directly (in the case of v_k itself) or through v_k, *S* is connected. Since all vertices are reachable there are nC_2 paths connecting the unordered pairs in

S. Out of these nC_2 paths (n-1) are connected to vk, so the number of paths connecting pairs of vertices where v_k falls on the path between them is $C(v_k) = nC_2 - (n - 1) = (n^2 - 3n + 2)=2$. Any new edge added to S after this stage must directly link two vertices that previously were connected only through v_k. Each new edge will, therefore, define a new shortest path that will reduce $C(v_k)$ by one.

Thus, maximum vertex centrality can be obtained only when the number of edges equals $n - 1$ and there exists a vertex v_k, that falls on all shortest paths of length greater than one. The relative centrality of any vertex in a graph, then, may be expressed as a ratio, $C'_B(v_k) = 2C_B(v_k)=(n^2-3n+2)$. Values of $C'_B(v_k)$ may be compared between graphs. Both $C_B(v_k)$ and $C'_B(v_k)$ may be determined for any symmetric graph whether connected or not. Thus, these measures solve the problem raised by Moxley and Moxleys[4] of determining the centrality of vertices in unconnected graphs. However, both measures take their maximum values only for vertices that are the centers of stars or the hubs of wheels. There are two distinct views on the meaning of the term centrality, when it refers to a property of a whole graph. One of these, based on graph theory, views a graph as exhibiting centrality to the degree that all of its vertices are central. They have limited utility and may be applied only to problems like the design of maximally efficient communication networks. The alternative view leads to the development of measures of graph centrality based upon the dominance of one vertex. In this conception, a network is central to the degree that a single vertex can control its communication.

They turn out to be related empirically to a wide range of behavioral characteristics of communicating groups including perception of leadership, frequency of error, rate of activity, speed of organization and personal satisfaction or morale. What is needed here is a graph centrality measure of the second, more general, and type. We can define $C_B(v_k)$ as the largest centrality value associated with any vertex in the graph under investigation. Then a natural measure of the dominance of the most central vertex is $C_B = \Sigma^n_{i=1}[C'_B(v_k) - C_B(v_i)]=(n - 1)$, which is the average difference in centrality between the most central vertex and all others. C'_B varies between 0 and 1. Its value is 0 for all graphs of any size where the centralities of all vertices are equal. Its value is 1 only for the wheel or star. Thus C'_B is an expression of the natural prescription that: "A communication network is considered structurally centralized to the degree that the network approaches that of a wheel network, and decentralized to the degree that the graph is an all-channel (complete)". The original application of the centrality idea was in the study of communication in small groups.

The study related to speed activity and efficiency in solving problems and personal satisfaction and leadership in small group settings. All of these variables were demonstrated to be related to centrality in some way. The range of applications, for the concept of centrality has been very wide, for instance, its impact on urban growth, study of the diffusion of a technological innovation in the steel industry, in inter organizational relations, to explain political integration in Indian civilization, to name a few. These several studies used perhaps a dozen different measures of centrality. While many are related, it is clear that there is little consensus on the solution to the problem of measuring centrality. Where then do these three new measures fit into this picture? The three numbers, $C_B(v_k);C'_B(v_k)$ and C'_B are more generally applicable than most of the alternatives. They are not limited to use in connected networks. The important question in considering applications and the one that is most often neglected-involves considering the relevance of the particular structural attribute measured to the substantive problem being studied. Thus, the use of these three measures is appropriate only in networks where betweenness may be viewed as important in its potential for impact on the process being examined. Their use seems natural in the study of communication networks where the potential for control of communication by individual vertices may be substantively relevant.

Scope for Practical Applications

The study of various parameters related to the concept of distance graphs has tremendous scope for applications. For example, consider the problem of placing sensors in a building or a utility network in order to detect contamination of the air or water supply. Practical motivation for this problem derives from recent world events including Tokyo's subway incident, London's poison gas bomb plot and various government warnings. While more effective sensors are currently being developed to address the increasing threats of contamination, these new sensors are likely to be expensive. Hence we resort to algorithmic techniques to place sensors in a network in such a way that cost is minimized and contamination can still be quickly detected. Two main goals are 1) Contamination detection and 2) Source identification. Two natural constraints for sensor placement are: (i) sensor constrained, that is, allowing only a fixed number of sensors and (ii) time constrained, that is, requiring contamination detection or source identification within a given time limit. These two goals and two constraints define various sensor placement problems.

Formal Description of Sensor Placement Problem

We model the network as a directed weighted graph $G = (V,E)$. V is a set of vertices representing possible locations for the sensors. There may be rooms in a building or pipe junctions in a water network. E is a set of edges representing flow between the vertices. These may be airways or hallways in a building or pipes in a utility network. In the sensor constrained variant of our problem, we are given a maximum number of sensors, S_{max}, and we want to minimize the time from contamination detection or source identification.

Role of Shortest Paths

Each edge (i, j) has weight $r_{ij} \in R^+$ which is the time it takes for contaminant to pass from vertex i to vertex j. One uses the shortest path metric to define the actual translocation rates in the graph. That is., for any two vertices i, j the translocation rates r_{ij} is defined as the length of the shortest path between i and j; r_{ij} is infinite if no such path exists. Since the shortest path metric defines the actual translocation rates in the graph, the triangle inequality is valid. One can construct an adjacency matrix representation of the effective translocation rates in the graph by using any all-paths shortest paths algorithm for directed weighted graphs with no negative weights.

Thus, the input graph is a weighted complete graph (possibly with some infinite weight edges), where some edges represent the actual flow conduits and the rest are inferred to represent the effective translocation rates. Given this input, the goal is to place sensors on the vertices so that it is always possible to detect the contamination and identity the vertex that is the source of contamination. A sensor s can detect contamination at vertex v if there exists a directed path from v to s. s detects contamination at v within time t if the length of the $v - s$ path is at most t.

Importance of finding the center of a graph

In the sensor-constrained contamination detection problem, we are given a weighted digraph $G = (V,E)$ with positive weights r_{ij} and a positive integer S_{max}, which is the maximum number of sensors to be used. The aim is to place the sensors onto the vertices of the graph in such a way that minimizes the maximum contamination detection time. This problem is equivalent to the Asymmetric K-CENTER Problem, which is well known to be NP-hard.

Asymmetric K-CENTER

Given a complete digraph $G = (V,E)$ of shortest (weighted) path distances between the vertices that satisfies triangle inequality, and a positive integer k. Find a subset of vertices S, $|S| = k$, which minimizes the longest distance from a vertex in S and any vertex in the graph. That is, find appropriate S that minimizes, *cost(S)*, defined as: $Cost(S) = Max_{v \in V}\ min_{s \in S} d(s, v)$.

Theorem 3.4 The sensor-constrained contamination detection problemis equivalent to the Asymmetric K-CENTER Problem.

Proof : To show the reduction in either direction, we equate $K = S_{max}$ and retain the underlying digraph G with the edge weights reversed $d(i, j) = r_{ij}$. Since the edge weights use the shortest path metric, the triangle inequality is satisfied. A solution to the Asymmetric K-CENTER is a subset of at most k vertices, S, that minimizes the distance from a vertex in S to a vertex in the graph. That is, this is a subset of at most S_{max} vertices such that the maximum time from any vertex to a vertex in S is a minimum among all such subsets. That is, this is a subset of vertices such that a placement of sensors at these vertices minimizes the contamination detection time.

Rooted Trees as Model for Water and Other Distribution Networks

Let G be any graph with one central vertex. The graph G^* is a rooted (graph) tree. A rooted tree is a tree graph with a special vertex designed as a root. All the edges are oriented away from the root. That is, for any edge (i, j), it is oriented from i to j. If i is on the path from the root to j. Water and other distribution networks can in some cases be modeled as rooted trees. In such a network there is a single supply source and it is delivered along a unique path to each destination. This model does not take into account possible back flow, this is, and it assumes the flow moves from the source to the destination only.

Detection There must be sensor in each leaf of the tree (a vertex with no outgoing edge), otherwise there is no way to detect contamination at that vertex. For a detection time limit T, the following procedure can be used to find the minimum set of sensors that ensures contamination detection within time T.

Procedure

1. Place a sensor at each leaf.
2. Follow the edges in the reverse direction from the current sensors and mark all the vertices that have distance at most T to a sensor as”covered”.
3. Put a sensor at each uncovered vertex that is first on the path from a sensor to the root.
4. Repeat steps 2 and 3 until all vertices are covered.

Theorem 3.6 In a rooted tree graph, for any vertex at most two sensors are sufficient to uniquely identify contamination at that vertex. (The two sensors are not necessarily the same for all the vertices).

Proof In a tree with no vertices of degree 2 every internal vertex has a set of descendants different from the set of descendants of its child (besides the child itself) and each internal vertex has at least one pair of descendants such that it is their least common ancestor. Thus, for any internal vertex it is sufficient to have sensors at some two vertices whose least common ancestor it is. These are two sensors identity their least common ancestor as the contamination source in such a tree. Placing a sensor at any vertex of degree 2 uniquely identities contamination at that vertex and effectively converts the tree to a tree with no vertices of degree 2.

References:

1. Bavelas.A., *A mathematical model for a graph structure,*Human Organization, 7, 16-30, 1948.
2. Bavelas.A., *Communication patterns in task oriented groups,* Journal of the Acoustical Society of America, 22, 271-288, 1950.
3. Beauchamp.A. M., *An improved index of centrality,* Behavioral Science, 10, 161-163, 1965.
4. Moxley.R. L., and Moxley.N.F., *Determining point centrality on contrived social network,* Sociometry, 37, 122-130, 1974.
5. Sabidussi.G., *The Centrality index of a graph,* Psychometrika, 31, 581-603,1966.