

# The Effect of Graph Layout on Inference from Social Network Data

Jim Blythe<sup>1</sup> and Cathleen McGrath<sup>2</sup> and David Krackhardt<sup>2</sup>

<sup>1</sup> School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213

<sup>2</sup> Heinz School of Public Policy and Management, Carnegie Mellon University

**Abstract.** Social network analysis uses techniques from graph theory to analyze the structure of relationships among social actors such as individuals or groups. We investigate the effect of the layout of a social network on the inferences drawn by observers about the number of social groupings evident and the centrality of various actors in the network. We conducted an experiment in which eighty subjects provided answers about three drawings. The subjects were not told that the drawings were chosen from five different layouts of the same graph. We found that the layout has a significant effect on their inferences and present some initial results about the way certain Euclidean features will affect perceptions of structural features of the network. There is no “best” layout for a social network; when layouts are designed one must take into account the most important features of the network to be presented as well as the network itself.

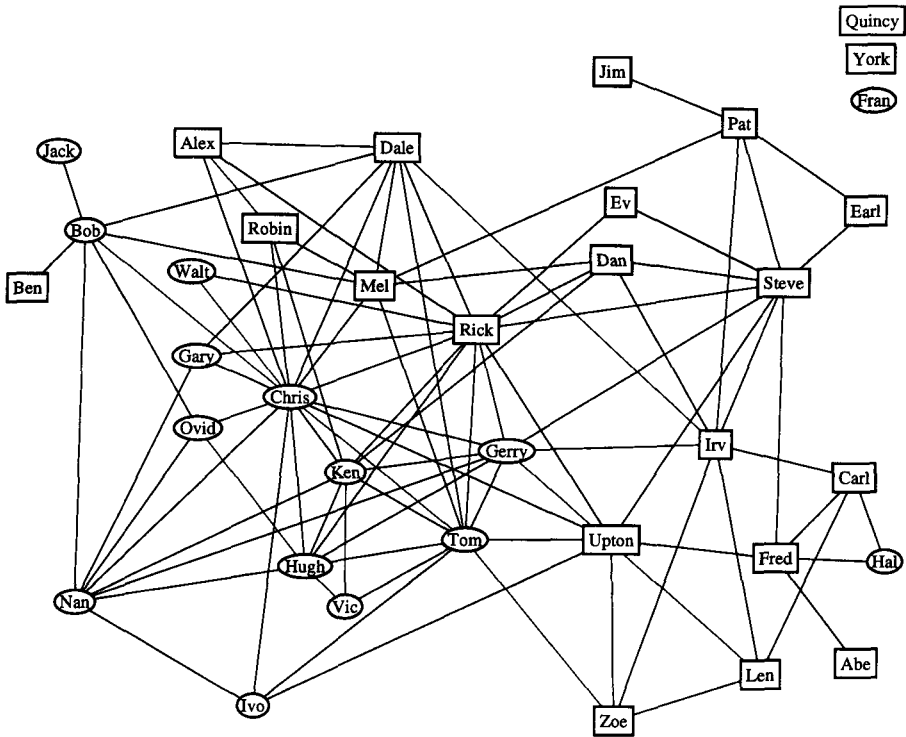
## 1 Introduction and Problem Statement

Social network analysis is a fast-growing field of social science that uses graph theory to analyze the structure of relationships among a set of individuals or groups. In this field, the vertices of a graph represent actors in a community and the edges or arcs typically represent patterns of relationships, such as communication, trade or friendship, within the community. In the remainder of this paper, we shall refer to a graph interpreted in this way as a “social network”, and to a drawing of a social network as a “sociogram”. We have been designing layout and visualization tools specifically aimed at social network analysis that include automatic layout facilities as well as the graphical presentation of some measures used in the field (Krackhardt et al 1994).

In this paper we explore the influence of network layout on individuals’ perceptions of common social network measures, in particular “prominence” and “bridging”, which are two facets of centrality, as well as the number of groups present in a network. Prominence deals with how visible or involved an actor is based on his or her position in the network. Importance as a bridge deals with an actor’s strategic positioning between groups, and most closely relates to betweenness centrality which we define below. Grouping most closely relates to the number of maximal cliques present in the graph.

### 1.1 Social Networks

Focussing on a particular application area allows for a more concrete metric of a good drawing than the general case, because of the more restricted kinds of information to be



**Fig. 1.** A network of friendship ties in a small U.S. high-tech firm with 36 employees. This sociogram was laid out by hand.

presented. We present a brief example of how social network data is often collected and analyzed. Krackhardt (1992) studies the role of close friendship or “philos” in organizational change in a small U.S. high-tech firm with 36 employees. The 33 employees who participated in the study were presented with a list of all other employees and asked to mark those they considered to be “personal friends”. The question was repeated for the other employees so that each respondent reported who all the employees considered to be personal friends. A “philos” relation is said to exist from  $i$  to  $j$  if  $i$  and  $j$  agreed that  $i$  considers  $j$  a personal friend. Figure 1 shows the sociogram for the philos relationship. Arrows are omitted because most relations are symmetric.

After the study was completed, a union attempted certification at the firm — to be voted an official union by the workers of the firm — and failed. A network analysis can help explain this relatively rare occurrence (Krackhardt 1992). In Fig. 1, the actors who are eligible to vote on the certification are depicted as ovals, and the managers as squares. Hal was the person who originally contacted the union, and he was chosen as their representative leading up to the vote on certification. His position in the network is far from central, and his influence is correspondingly low. Chris, a considerably more central actor, had originally supported the union but was increasingly ambivalent as the

vote approached, in part because of his ties to several managers. Because of his position in the network, Chris played a pivotal role in the final rejection of union certification.

As can be seen from this example, node centrality is an important concept in social network analysis. Two graph theoretical measures commonly used for centrality are “degree centrality”, and “betweenness centrality”. Degree centrality measures the number of nodes that are adjacent to the focal node. Betweenness centrality measures the number of times a node is on the shortest path between two other nodes, following Freeman (1978). For a detailed description of these and other measures of centrality used in social network analysis see Wasserman and Faust. (1994).

## 1.2 Social Network Layout

We are interested in how well a layout of a social network conveys information about the centrality of actors and the groupings of the network. We conducted an experiment in which eighty subjects were presented with different drawings of the same graph and asked to rate certain nodes on their prominence and importance as a bridge, two common features used in social network analysis, as well as estimate the number of distinct social groupings evident from the drawing. We found that both graph theoretical and purely Euclidean elements of the drawing were statistically significant in predicting the responses of the subjects. These results may lead to aesthetics for drawing social networks which are aimed at making human perception of these measures more closely aligned to the graph theoretical determinants typically used in the field.

A large body of work exists on computer programs to draw graphs according to certain “aesthetics” (see Battista et al. (1994) for a survey of this work). However, almost all of this work considers aesthetics that attempt to improve graph readability from a very general point of view without considering specific applications, and uses general aesthetics such as the regular spacing of the nodes and the minimization of bends and edge crossings. In addition only a few studies explicitly question the aesthetics in use and there has been very little work on analyzing how well such aesthetics actually improve the information conveyed in a graph drawing. Ding and Mateti (1990) consider the subjective factors that go into the drawing of a diagram intended to explain data structures in computer programs. They produce their factors by examining the pictures that appear in a number of text books in that field. Batini et al (1985) make an experimental study of the aesthetics used in Entity Relationship diagrams from the field of software engineering. We are not aware of any studies that consider the information one might want to convey in a social network drawing.

In the rest of this section we describe the features of a drawing that are used to predict the values a human might put on social network measures while looking at a drawing. In the next section we describe our experimental design and the results of the study.

Our experiment was conducted using 5 different drawings of one particular network having 12 actors and 24 ties. This framework allows us to hold structural relationships among nodes constant while varying their spatial relationships. The drawings vary in the proximity of nodes to each other and the positioning of nodes in the center or periphery of the graph. We investigate the influence of two different kinds of factors on perceptions from sociograms: those based purely on the structure of the graph and those

based on the spatial properties of the layout. After the structural factors are taken into account, we explore how spatial factors are found to influence respondents' perceptions of prominence, bridging, and grouping. While controlling for structural characteristics we study the influence of:

1. The proximity of a node to the center of the layout on the perception of its prominence.
2. The positioning of a node between pairs of other nodes on the perception of its importance as a bridge.
3. The spatial clustering of groups of nodes on the perception of grouping.

## 2 Experimental Design

80 graduate students who had just completed a course in organizational theory (a course that emphasized the importance of understanding networks in organizations) volunteered to be subjects in the experiment. The subjects were given a questionnaire containing three of the five graph drawings and told the following:

The following three graphs are modelled after networks of communications observed in three different merger and acquisition teams of an investment banking firm. A connection between two team members means that they discuss work related matters with each other. If no line exists between two team members then they never discuss work with each other.

All nodes were labelled with first names in the drawings presented to the subjects. By providing a context, the investment banking firm and group member names, we attempt to focus the subjects' attention on the social aspect of communication networks. In every drawing, each node was mapped to a new name. For analysis and discussion purposes, we have renamed the nodes of interest with letters from A to E. The five drawings are shown in Fig. 2.

We used an incomplete balanced latin square design to control for order in the presentation of the sociograms by displaying each drawing first, second, and third exactly once. Each drawing is preceded by every other drawing except one exactly once.

We asked respondents three questions about the same five focal nodes in each sociogram: 1) how many subgroups were in the sociogram; 2) how "prominent" was each player in the sociogram; and 3) how important a "bridging" role did each player occupy in the sociogram. The subject rated the prominence and importance of the same five nodes (with different first name labels) for each drawing by circling a number from 1 to 7 on a Likert scale that went from not prominent (or not important as a bridge) to most prominent (or most important as a bridge). The question about prominence was worded as follows: *"Some individuals have a more prominent role in their team than other individuals. Please rate the following people according to how prominent within their team they appear to you by circling the appropriate number next to each name."* The question about importance as a bridge was worded in the following way: *"Some individuals are important because they form a bridge between subgroups. Please rate the following people according to how important they appear to be as bridges between*

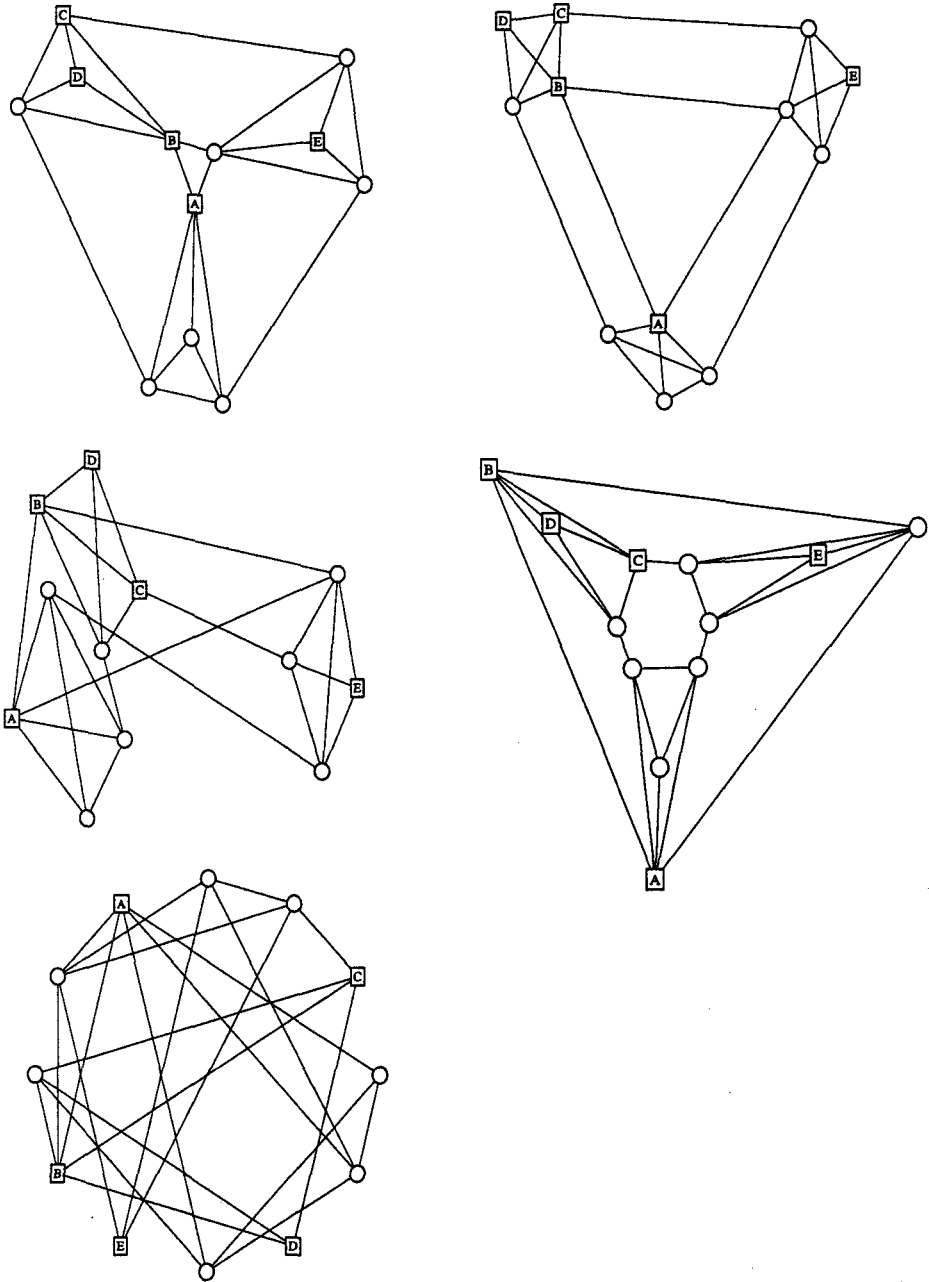


Fig. 2. The five different drawings of the graph.

*subgroups by circling the appropriate number next to each name.”* Finally, we created the questionnaire using a format that allowed one drawing and questions about that drawing to be viewed at the same time, without viewing any other drawing.

## 2.1 The Network

**Table 1.** Centrality Measures

	Degree Centrality	Freeman Betweenness Centrality
A	5	8.67
B	5	8.67
C	4	4.67
D	3	0.00
E	3	0.00

The network itself is symmetric, with 12 actors and 24 ties. We used a small graph so that the subjects would not be overwhelmed by the amount of information presented to them. The overall measure of network density is 36 percent. Table 1 shows the values for the two measures of centrality defined in section 1.1 for nodes A through E. Nodes A and B are automorphically equivalent as are nodes D and E. The network has four cliques.

We adhered to the general standards of aesthetics discussed earlier when laying out the graphs, that is, we tried to avoid line crossings, and we tried to avoid nodes landing on top of lines. Drawing 5, the circle, did not adhere to the standards. It is included because of its general acceptance as a means for presenting social network data.

## 3 Results

The results of our analysis of individuals' reports of each node's prominence and importance as a bridge and of the overall number of groups in the network support our hypotheses that the spatial relationships influence perception in the graph layouts. We use analysis of variance to test whether the average prominence or importance as a bridge for each node changes with each layout. While the ordering of prominence and importance as a bridge does not change across drawings, the relative values assigned to prominence and importance as a bridge do change.

We also use regression analysis on all nodes combined to test the influence of our spatial measures when structural measures are included. We find that both structural and spatial characteristics of the nodes predict the respondents' reports of nodes' prominence and nodes' importance as a bridge. For grouping, we compare the distribution of groups across all five layouts and find some differences in the shapes of the distributions across the five layouts.

The following sections give details of the analysis for prominence, importance as a bridge, and grouping. In the tables that follow, one star (\*) denotes significance at the 0.05 level, two stars (\*\*) at the 0.01 level and three stars (\*\*\*) at the 0.001 level.

**Table 2.** Individual mean centered scores for prominence

	Drawing 1	Drawing 2	Drawing 3	Drawing 4	Drawing 5	F Value
A	0.65 (1.79)	1.50 (1.16)	1.07 (1.25)	1.10 (1.60)	0.36 (1.91)	3.63 **
B	0.60 (1.86)	1.35 (1.38)	0.95 (1.46)	1.08 (1.60)	0.27 (1.77)	3.07 *
C	-0.31 (1.31)	-0.12 (0.93)	0.23 (1.02)	0.26 (1.44)	-0.39 (1.61)	2.42 *
D	-0.07 (1.36)	-0.98 (1.30)	-1.18 (1.56)	-0.25 (1.49)	-0.94 (1.34)	5.63 ***
E	-0.11 (1.65)	-1.60 (1.31)	-1.16 (1.57)	-0.41 (1.50)	-1.10 (1.66)	6.99 ***

### 3.1 Prominence

Before we analyzed the prominence data, we converted the prominence that was reported on a 7-point Likert scale to a mean centered prominence score. We did this by subtracting each respondent's average prominence score across all three layouts from each node's prominence score assigned by the respondent. This controls for individuals' tendencies to rate high or low in general. If a node has a negative mean centered prominence score, the respondent rated that node less prominent than the average.

We use analysis of variance to test the null hypothesis that each node's average mean centered prominence score is the same across all five layouts. The average mean centered prominence scores and results of the analysis of variance are reported in Table 2. When we compare the prominence scores for all five drawings, the null hypothesis that the nodes' average mean scores are all the same can be rejected. Thus, respondents' evaluation of nodes' prominence is not constant across different drawings, rather the degree of prominence assigned by the respondents changes as the drawing changes.

We further estimate a linear equation using Ordinary Least Squares (OLS) to predict prominence using degree centrality (the number of edges adjacent to each node) and the Euclidean distance from the center of each node (normalized across drawings). We estimate this equation on one data set made up of all 5 nodes A through E in all 5 drawings combined, with a total of 1127 observations after accounting for missing responses. Each node's degree centrality remains constant across drawings because it is a structural characteristic of the node, while the node's normalized Euclidean distance from the center of the graph varies across drawings. We find that (1) degree centrality has a positive and significant relationship to reported prominence and (2) normalized Euclidean distance from the center of the graph has a negative and significant relationship

to reported prominence. Both effects are significant at the 0.001 level. These two results support our initial hypothesis that both the structural characteristics of the network as well as the spatial representation of nodes in a sociogram influence individuals' perceptions of prominence.

We included dummy variables for individual respondents to control for individual effects, so we did not use mean centered prominence scores, but rather the reported prominence scores. More details about the regression analysis for prominence and importance as a bridge are found in a longer version of this paper (McGrath et al. 1995).

### 3.2 Importance as a Bridge

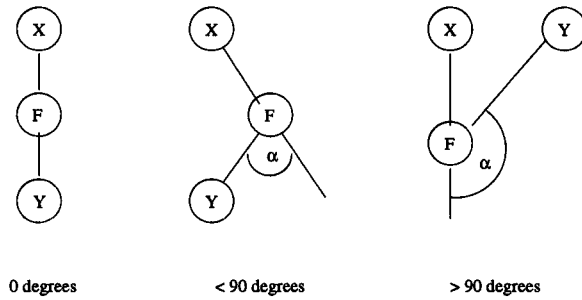
**Table 3.** Individual mean centered scores for importance as bridge

	Drawing 1	Drawing 2	Drawing 3	Drawing 4	Drawing 5	F Value
A	1.70 (1.44)	1.99 (0.90)	1.45 (1.28)	1.31 (1.84)	0.65 (1.66)	5.73 ***
B	1.73 (1.41)	1.85 (1.25)	1.47 (1.14)	1.48 (1.66)	0.75 (1.58)	4.15 **
C	0.26 (1.26)	0.24 (1.03)	0.58 (1.24)	0.83 (1.38)	-0.42 (1.48)	5.77 ***
D	-1.72 (1.34)	-2.05 (0.99)	-1.73 (1.18)	-1.31 (1.67)	-0.86 (1.46)	5.31 ***
E	-1.61 (1.49)	-2.09 (0.97)	-1.68 (1.41)	-1.56 (1.42)	-1.04 (1.79)	3.09 *

We perform the same analysis of variance on measures of importance as a bridge. Results are shown in Table 3. Again, we calculate a mean centered bridging score for each node. The null hypothesis that the average mean centered bridging score is equal across drawings can be rejected.

We also estimate an equation to predict the reported importance as a bridge of each node. We estimate it as a linear function of the node's betweenness centrality, using Freeman's betweenness centrality measure (1978), and the average angle value between the node and every other pair of nodes in the drawing. The betweenness angle for a node  $F$  with respect to a pair of other nodes  $X$  and  $Y$  is the angle of deviation between the line from  $X$  to  $F$  and the line from  $F$  to  $Y$ . It is independent of the ordering for  $X$  and  $Y$ . If  $F$  lies directly between  $X$  and  $Y$ , this angle is zero, and as  $F$  move further from  $X$  and  $Y$  the angle increases to a maximum value of 180 degrees, as shown in Fig. 3. We sum the betweenness angle for each node between all other pairs of nodes, and normalize this value.

From linear regression, we find that the coefficient of betweenness centrality is positive and statistically significant suggesting (as expected) that betweenness centrality has a positive relationship with reported importance as a bridge. Also, the coefficient of normalized angle between pairs of nodes is negative and statistically significant.



**Fig. 3.** Angle of deviation as a spatial measure of betweenness.

This supports our hypothesis that spatially positioning a node between pairs of other nodes will increase viewers' perception of that node's importance as a bridge. Again, the effects are significant at the 0.001 level.

### 3.3 Grouping

**Table 4.** Number of groups reported in each drawing

Graph Drawing	Mean Number of Groups Mean (Standard Deviation)	
1	5.09	(3.44)
2	4.29	(3.73)
3	3.58	(1.62)
4	4.79	(3.94)
5	3.49	(2.92)

We investigate the effect of the spatial clustering of nodes on the subjects' perceptions of the number of groups in the network. Table 4 shows the mean number of groups reported for each drawing.

We consider the shape of the distributions of the number of groups reported. Figure 4 shows the distribution of responses for the number of groups for all 5 drawings. It can be seen that drawing 5, the circle, has a distribution that is much flatter than the rest, suggesting that respondents could not determine groupings with consistency in the circle drawing. Drawing 3, which presented the nodes clumped in two groups, one large and one small was not interpreted as two groups by many people. We used the Kolmogorov-Smirnov test to determine if each pair of distributions of reported number of groups could be considered statistically significantly different. The Kolmogorov-Smirnov Test for two samples tests the null hypothesis that two random samples are taken from the same distribution. Table 5 shows that the distribution of number of groups for Drawing 5 differs from the distribution of number of groups for every other drawing. In addition, Drawing 1 differs from Drawings 2 and 3.

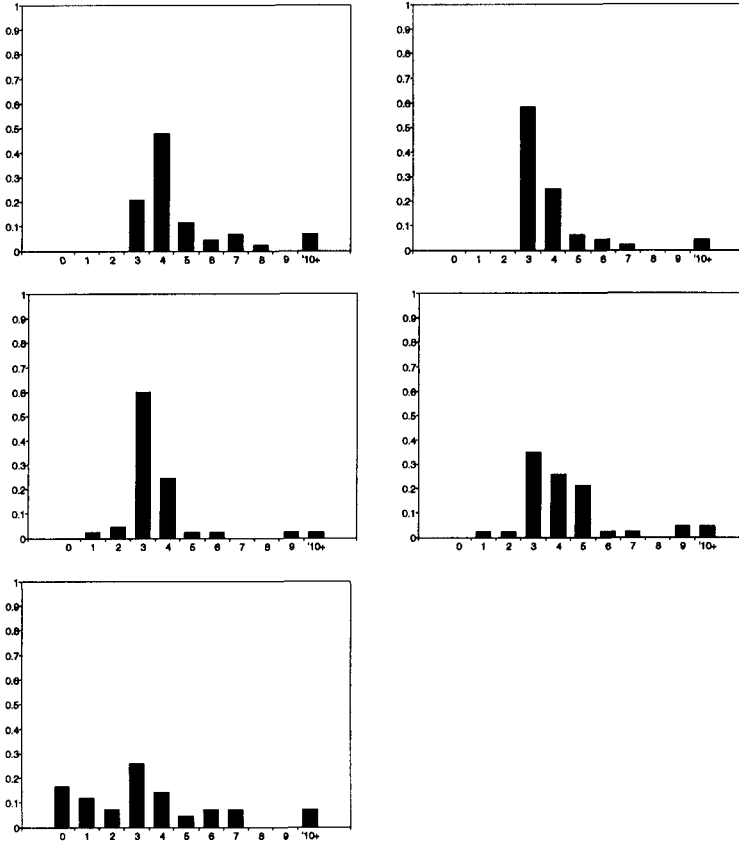


Fig. 4. Distributions of the number of groups reported for each drawing

## 4 Discussion

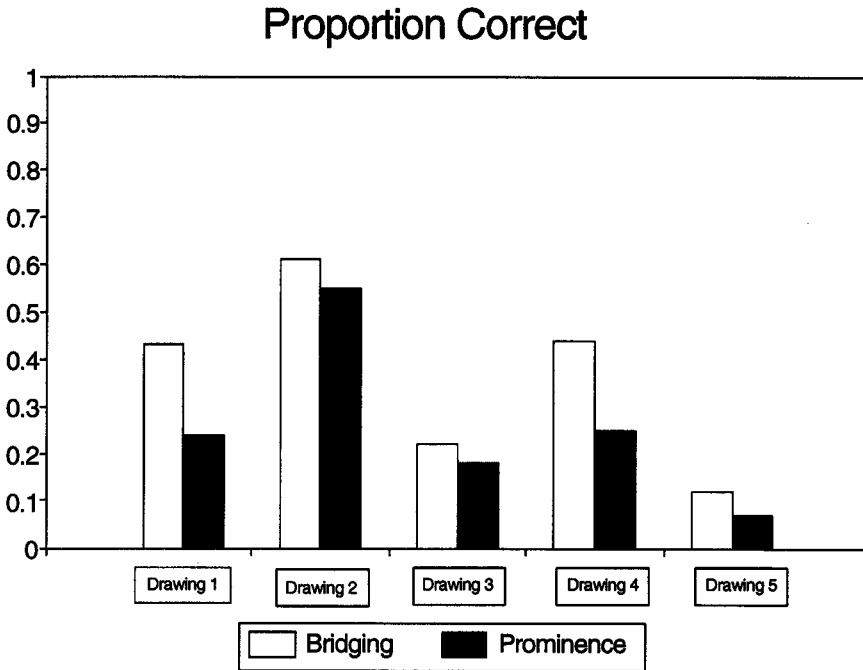
We have shown that both structural and spatial factors influence individual's perceptions of prominence, importance as a bridge and grouping. We conducted the first empirical study of the influence of these factors, and while it is dangerous to draw strong conclusions from this small experiment, it would appear that in many cases the influence of spatial layout on the drawing can be predicted to some degree. In particular we found that (1) the perceived prominence of an actor decreases as the corresponding node moves away from the center of the social network, (2) the number of groups perceived by an individual can be altered by the relative proximity of nodes in the sociogram and (3) an actor's perceived importance as a bridge between two groups can be decreased by moving the node further from the center of the bisector of the two groups.

In general it is not possible to say that one drawing is the "best" drawing of a particular social network. Often the best drawing is the one that highlights the characteristic of the network that is being discussed, and there may be no single drawing that best highlights

**Table 5.** Kolmogorov-Smirnov Test for Difference of Distribution of Number of Groups D stat (p-value)

	Drawing 1	Drawing 2	Drawing 3	Drawing 4	Drawing 5
Drawing 1	X	0.38 (**)	0.46 (***)	0.19	0.40 (**)
Drawing 2		X	0.08	0.19	0.35 (**)
Drawing 3			X	.27	.28
Drawing 4				X	0.30 (*)
Drawing 5					X

every characteristic. For our example network, a “good” drawing should highlight the order of prominence and bridging for the five nodes of interest, and clearly display the group structure.



**Fig. 5.** The proportion of responses with correct ordering of nodes for each layout.

We compare the number of times respondents report the correct order of nodes for each drawing. For both prominence and bridging, the correct order is  $A = B > C > D = E$ . Figure 5 shows that Drawing 2 has the highest proportion of correctly ordered responses for both prominence and bridging. This suggests that Drawing 2 is

the “best” drawing to convey prominence and bridging. On the other hand, Fig. 4 shows that Drawing 1 has the highest proportion of correct responses for the number of groups, 4, in the network.

The results of our study suggest objective measures that can be applied to a sociogram to estimate how well it conveys important structural information about the network. The measures allow us to compare two drawings of the same network, given a set of structural characteristics they are desired to convey, and also suggest a new set of drawing aesthetics that can be used to help produce good drawings of sociograms automatically. However, this is a preliminary study and there is clearly much more work to be done in this area. While we have focussed on social networks in this paper, we believe the same approach can be used to provide more concrete metrics in many application areas of graph drawing.

## References

- C. Batini, L. Furlani, and E. Nardelli. What is a good diagram? a pragmatic approach. In *4th International Conference on the Entity Relationship Approach*, Chicago, 1985.
- G. Di Battista, P. Eades, R. Tamassia, and I. Tollis. Algorithms for drawing graphs: an annotated bibliography. *Computational Geometry: Theory and Applications*, 1994.
- C. Ding and P. Mateti. A framework for the automated drawing of data structure diagrams. *IEEE Transactions on Software Engineering*, 16(5):543–557, 1990.
- L.C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1:215–239, 1978.
- David Krackhardt. *The Strength of Strong Ties: the Importance of Philos in Organizations*, in Nohria and Eccles (eds) “Networks and Organizations: Theory and Practice”, chapter 3. Harvard Business School Press, Boston, 1992.
- David Krackhardt, Jim Blythe, and Cathleen McGrath. Krackplot 3.0: An improved network drawing program. *Connections*, 17(2):53–55, 1994.
- Cathleen McGrath, Jim Blythe, and David Krackhardt. The effect of spatial arrangement on inference from graphs. Working note, Heinz School of Public Policy and Management, Carnegie Mellon University, 1995.
- Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.