

# Seeing Groups in Graph Layouts<sup>1</sup>

Cathleen McGrath, Jim Blythe, David Krackhardt  
*Carnegie Mellon University*

## Introduction

Social networkers frequently make use of drawings to communicate information and ideas about networks. However, the impact of the layout of a network on the conclusions that a viewer is likely to draw has so far received very little scrutiny. In this paper, we extend work begun in (Blythe et al, 1995) and (Mcgrath et al, 1996) to understand how the layout of graphs depicting social network data influences the inferences viewers draw about social networks. Our previous work focused on the perception of prominence or bridging of a particular node. Here we focus on perceptions of clustering among nodes.

Previous empirical work studying graph layout and social networks has shown that layout influences viewers' perception of the prominence, or importance of individuals in the network (Blythe et al, 1995). Purchase et al. (1995) report on experimental work validating general graph layout aesthetics. Both of these empirical studies of human perception of graphs build on earlier work on graph drawing aesthetics (see Battista et al.(1994) for a survey of this work).

Finding groups in networks of people is an important part of social network analysis. According to Scott (1991):

One of the most enduring concerns of those working in social network analysis has been the attempt to discover the various 'cliques' and cohesive sub-groups into which a network can be divided.

We extend experimental work testing viewers' understanding of graphs based on layout by using an interactive system that allows us to closely track the responses and response time of people answering questions about the graphs.

## Experimental Design

### The Study

This paper reports results of a larger study of network perception in which sixty-one graduate students who had just completed a course in organizational theory emphasizing networks in organizations participated. Participants evaluated one of five different orderings of five

---

<sup>1</sup>Address correspondence to Cathleen McGrath (cm3t+@andrew.cmu.edu)

layouts chosen at random. The layouts presented to the participants had nodes labeled with first names which differ with each layout.

Our test platform is a modified version of KrackPlot 3.0 (Krackhardt et al, 1994), a social network drawing package, which participants use to interactively assign nodes to groups. Participants were able to assign nodes to one of (at most) five different groups by clicking on a square to activate a group and then clicking on the nodes they believed to be in that group. Participants were able to change a node's group assignment by clicking on the node while a different group was activated. This system allows us to track the exact order of assignment of nodes to groups as well as the time spent assigning each individual node to a group.

We brought participants together in a computer lab as we demonstrated the program. Before they began the exercise we told them :

*In the following exercise, you will practice finding groups when the social network is presented as a graph. After we do two warm-up exercises together, you will have a chance to identify groups for five different graphs that show communication patterns among people, that is, a line between two people means they talk often.*

To ensure that they participants were comfortable with the system we led the participants through two exercises: one in finding a particular node and clicking on it and one in activating a group and then assigning nodes to that group. Before each of the five graphs appeared, the following message appeared on the screen:

*In the next screen, you will be shown a graph and asked to divide its nodes into separate groups. Each time you click on a node, its shape and color will change. You can select the shape and color by selecting the appropriate node in the menu bar at the top of the screen. Please make use of different shapes and colors to divide the nodes into the groups that seem appropriate.*

Since each node could belong to only one group, respondents assigned nodes to the group in which they fit "best".

### The Network

We report on the results of three different layouts representing an interaction among bank employees. These data were obtained by Krackhardt as part of a research project on networks in banking. The ethnography of the site suggests that there are two groups, each with their own subculture within the bank. The network is dense with 26 nodes and 93 edges. Figure 1 shows the three layouts. The first layout depicts two spatial groups with two edges connecting the groups. In the second layout, a node labeled "N" from the right spatial group is positioned in the left spatial group. Six edges connect the spatial groups. Four of those edges are connected only to N which is not itself connected to any nodes in its spatial group. Finally, the third layout has the nodes clustered in three spatial groups with many edges connecting the spatial groups.

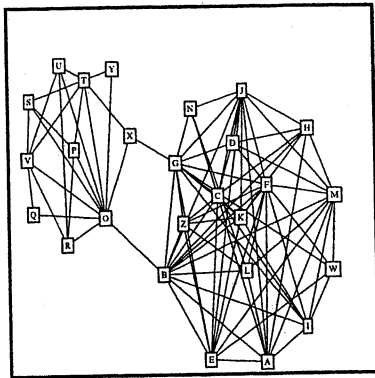


Figure 1: Layout 1

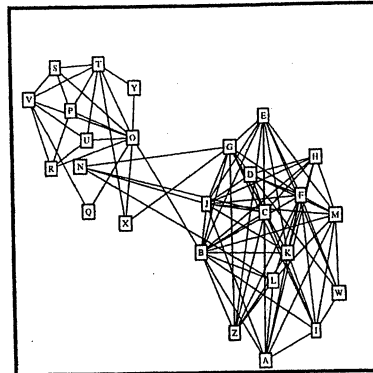


Figure 1: Layout 2

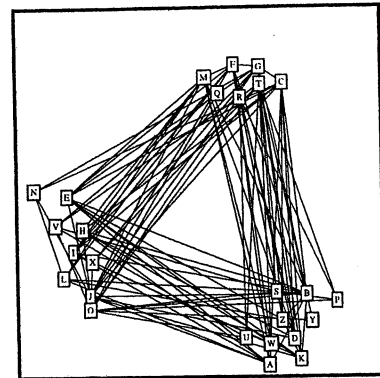


Figure 1: Layout 3

## Results

Because the purpose of graphical presentation of data is to convey information quickly in a clear and correct form, we will consider a "good" layout to be one that allows viewers to draw inferences about the information presented quickly and correctly. We evaluate graphs based on the amount of time viewers take to finish assigning nodes to groups. We show that different layouts of the same graph suggest different numbers of groups to viewers. The following sections will compare each layout based on viewers' time to complete assignments, perception of the number contained in the graph, and perceptions of group co-membership for pairs of nodes. We define a dyadic relationship between nodes called "co-membership" as the proportion of times viewers place two nodes in the same group. We test a model that predicts the proportion of times two nodes,  $i$  and  $j$ , will be placed in the same group based on structural properties of the graph, the adjacency matrix and the path distance between  $i$  and  $j$ , as well as spatial properties of the layout, the Euclidean distance between  $i$  and  $j$ .

### Time

The more quickly viewers can look at a layout and make inferences about the underlying structure of the graph, the better the layout is at conveying information. We compare time to complete group assignments for all three layouts.

**Table 1: Time to Complete Group Assignment**

Layout	N	Time in seconds (St. Dev)
Layout 1	31	72.5 (32.48)
Layout 2	24	100.7 (43.89)
Layout 3	33	132.8 (98.54)

Table 1 shows that on average viewers took the least amount of time assigning groups for Layout 1. The average amount of time for Layout 1 was 72.5 seconds. Viewers took 100.1

seconds on average to assign nodes for Layout 2 and 132.8 seconds on average to assign nodes for Layout 3. Of the three layouts of the interaction data, Layout 1 presents the information in a way that is most quickly understood by the viewers.

### Perceived Number of Groups

Manipulating the layout of the graph can change viewers' perceptions of grouping even though the underlying structure of the graph is unchanged. The number of groups that viewers find for a given layout is a good first estimate of how their perception of the graph changes when layout changes.

Figure 2 shows the proportion of viewers reporting one, two, three, four, or five groups, respectively. Forty-five percent of respondents who evaluated Layout 1 divided the nodes into two groups. Of those respondents evaluating Layout 2, 38% divided the nodes into two groups.

For Layout 3 only 9% divided the nodes into two groups; instead 49% divided the nodes into three groups. This suggests that the spatial clustering of nodes, not simply the structure of the graph, is influencing respondents' perceptions of the group structure. A chi-square test of independence supports the hypothesis that the number of groups reported by respondents is not independent of layout.

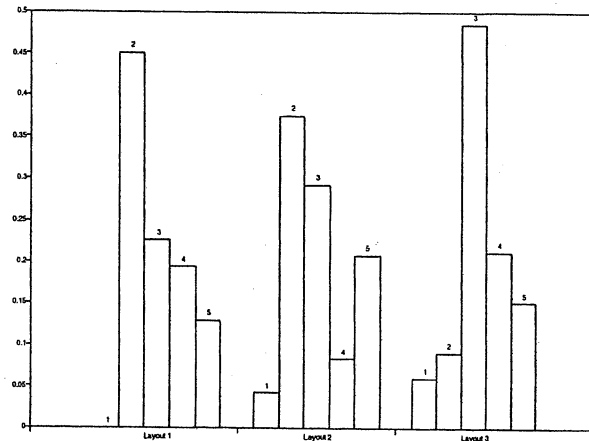


Figure 2: Number of Groups Assigned by Layout

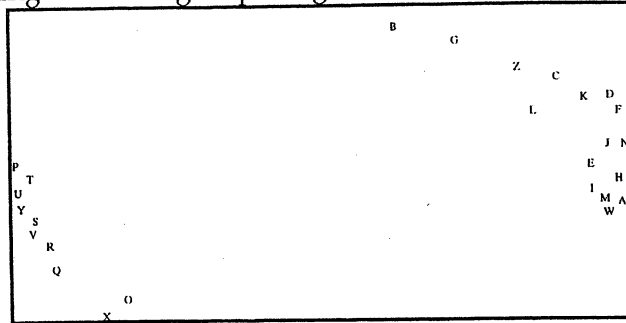
### Group Co-Membership

We can describe participants' perceptions of grouping by showing how often they assign pairs of nodes to the same group. Viewers will assign pairs of nodes,  $i, j$ , to the same group based on  $i$  and  $j$ 's structural relationship and  $i$  and  $j$ 's spatial relationship. We measure  $i$  and  $j$ 's structural relationship by adjacency and path distance. We measure  $i$  and  $j$ 's spatial relationship by the Euclidean distance between  $i$  and  $j$  in each of three layouts. For each layout we calculate co-membership as the proportion of times  $i$  and  $j$  are assigned to the same group. Using multi-dimensional scaling (MDS) we can represent viewers' perceptions of co-membership. We can further test the effect of adjacency, path distance, and Euclidean distance using QAP to model group co-membership.

### MDS of Group Co-Membership

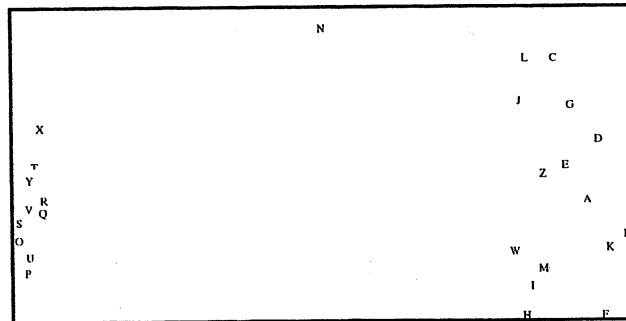
The MDS representation of the proportion of times  $i$  and  $j$  are assigned to the same group shows some interesting differences between Layouts 1, 2, and 3. Figure 3 shows the MDS of

Layout 1 shows two clear clusters of nodes. Nodes X, O and B, G are pulled more toward the middle of the suggesting that their group assignments sometimes coincide.



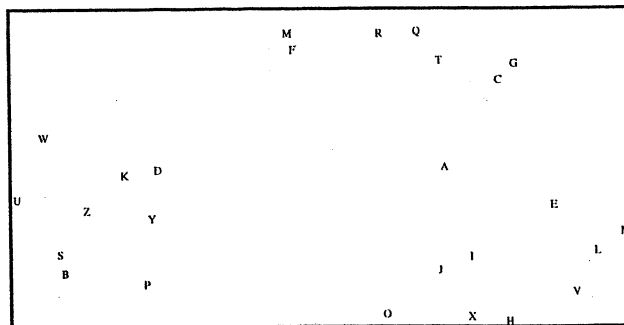
**Figure 3: MDS: Co-Membership for Layout 1**  
(Stress=.094)

The MDS of Layout 2, Figure 4, shows a different picture. In the second layout, N is pulled from the spatial cluster to which it is connected structurally and is positioned in the spatial cluster to which it has no ties. While there are still two main groups, N has moved to the middle of the display, suggesting that N shares co-membership with nodes in both of the main groups. The nodes X, O, B and G are completely integrated into their groups, suggesting that their bridging role is no longer noticed. A third group, {F, H, I, M, W}, emerges on the right. In Layout 1, {F, H, I, M, W} was more integrated into the larger group.



**Figure 4: MDS of Co-Membership for Layout 2**  
(Stress=0.111)

Finally, in Figure 5, the MDS of Layout 3 is more disperse than the previous two MDS representation. However, three groups can be distinguished.



**Figure 5: MDS of Co-Membership for Layout 3**  
(Stress=.196)

### QAP Analysis of Group Co-Membership

For the QAP analysis, we ran three separate models for each layout. The Euclidean distance matrix is the distance (in pixels) between each pair of nodes, the adjacency matrix is a binary matrix where 0 means that there is no edge between a pair of nodes and 1 means that there is an edge connecting the pair of nodes. The path distance matrix shows the number of links in the shortest path between each pair of nodes. QAP analysis allows us to test the independent effects of these three matrices on individuals' perceptions of grouping. We see that Euclidean distance has an important effect on perception of grouping no matter what the graph layout is.

Tables 2 through 4 show that, as Euclidean distance between a pair of nodes increases, the proportion of times that pair of nodes will be assigned to the same group decreases. This effect remains significant through all three layouts. Path distance between pairs of nodes is also negatively related to the proportion of times the pair will be assigned to the same group. The adjacency matrix is positively related to co-membership, but this effect is weakened with the addition of path distance. Path distance is negatively correlated with the adjacency matrix. Table 5 shows the correlations between all independent variables.

**Table 2: QAP analysis predicting proportion of times  $i$  and  $j$  are assigned to the same group: Layout 1**

Independent Variable	1	2	3
Euclidean Distance/100	-0.14 ***	-0.08 **	-0.08 **
Adjacency Matrix	0.26 ***	0.10	
Path Distance		-0.28 ***	-0.24
R <sup>2</sup>	0.568	0.690	0.685

**Table 3: QAP analysis predicting proportion of times  $i$  and  $j$  are assigned to the same group: Layout 2**

Independent Variable	1	2	3
Euclidean Distance/100	-0.14 ***	-0.11 ***	-0.11 ***
Adjacency Matrix	0.23 ***	-0.02	
Path Distance		-0.19 ***	-0.18 ***
R <sup>2</sup>	0.615	0.683	0.683

