Perceptual organization in user-generated graph layouts

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Perceptual Organization in User-Generated Graph Layouts

Frank van Ham and Bernice E. Rogowitz, *Senior Member, IEEE*

Abstract— Many graph layout algorithms optimize visual characteristics to achieve useful representations. Implicitly, their goal is to create visual representations that are more intuitive to human observers. In this paper, we asked users to explicitly manipulate nodes in a network diagram to create layouts that they felt best captured the relationships in the data. This allowed us to measure organizational behavior directly, allowing us to evaluate the perceptual importance of particular visual features, such as edge crossings and edge-lengths uniformity. We also manipulated the interior structure of the node relationships by designing data sets that contained clusters, that is, sets of nodes that are strongly interconnected. By varying the degree to which these clusters were "masked" by extraneous edges we were able to measure observers' sensitivity to the existence of clusters and how they revealed them in the network diagram. Based on these measurements we found that observers are able to recover cluster structure, that the distance between clusters is inversely related to the strength of the clustering, and that users exhibit the tendency to use edges to visually delineate perceptual groups. These results demonstrate the role of perceptual organization in representing graph data and provide concrete recommendations for graph layout algorithms.

Index Terms— Network layout visualization, perceptual organization, graph layout, user studies.

◆

1 INTRODUCTION

Recent years have seen a meteoric growth in the availability of data about networks, ranging from neural networks to communication networks to social networks. Since these networks are typically large and complex, it is often difficult to represent the information in a way that allows users to understand patterns and relationships in the data. Graph drawing algorithms allow us to automatically generate visual representations of these networks, and make it easier for users to reason about network properties. These algorithms often optimize particular visual characteristics of the layout, such as minimizing edge crossings, maintaining uniform edge length or keeping edges orthogonal.

In most cases, these characteristics are based on the algorithm designer's intuitions about proper network layout and not on empirical data about how human observers perceive and organize networks. Although a number of studies have been performed to see how various properties compare on different user tasks [21][11][7], these all analyze a predetermined set of visual characteristics. They do not address the question of how end users would spontaneously arrange a graph, which might reveal visual characteristics that have not previously been considered or controlled for.

The experiments in this paper fill this gap by allowing the users themselves to arrange the nodes in a graph, providing more direct and explicit insight into how human observers perceive and organize network information. In particular, we ask users to arrange the nodes in graph layouts in order to best represent the structure in the data and we then analyze the graphs they create. We measure edge-crossings, node positions, edge length uniformity and examine their ability to discover and represent cluster structures in the data with different levels of internal complexity. We also examine their layouts for patterns that we had not anticipated. By measuring the users' behaviour directly, as they discover and represent information in a connected graph, we hope to improve our knowledge about what spatial parameters should be emphasized for network layout.

Consistent with intuitions derived from the literature on perceptual organization, we found that users have a strong preference for creating layouts that perceptually group clusters together and that the distance between these groups is inversely related to the number of edges connecting them. Furthermore, we found that users often visually delineate their cluster layouts by using the cluster's edges to create a bounding hull. Both of these results can offer guidance to network layout algorithm design and suggest potential avenues for future research.

2 RELATED WORK

At the beginning of the twentieth century, visual psychology research was dominated by an empirical tradition that sought to describe higher-level perception in terms of hierarchical low-level operators. The Gestalt school was a reaction to this approach, pointing out, famously, that "the whole is greater than the sum of the parts." In a very convincing set of experiments, they demonstrated that perception is largely driven by top-down processes [5]. In their view, human perception is an active process, and we are hard-wired to perceive visual grouping and organization. Later work in cognitive psychology has showed how we actively seek to organize information into meaningful groups, or "clusters", and that information can be more easily remembered if it is organized into meaningful "chunks" [8].

The Gestalt principles of Proximity and Closure [22] bear explicit exploration. The law of Proximity states that objects near each other will be perceived as belonging to a common group. The Law of Closure states that objects that are bound together by a common edge will be perceived as belonging together. Recent studies have empirically demonstrated the operation of Gestalt principles in visual organization [6][10]. To test the operation of these principles in network organization, we have devised an experimental paradigm in which users dynamically adjust the nodes in a network diagram. The goal is to see whether users are sensitive to clusters in a network structure and if so, how they surface these groupings perceptually. We hypothesize that users will organize nodes that belong within a cluster near each other, according to the principle of Proximity, and that edges may function to reinforce cluster boundaries.

Previous work in evaluation of visual characteristics in graph layouts is mostly aimed at evaluating a particular aesthetic's fitness for certain tasks. Purchase's early work performed a number of experiments to rank layout aesthetics based on the effect they had on a user's understanding of the graph [11]. Minimizing edge overlap emerged as the most important aesthetic for graph comprehension, followed by minimizing the number of bends and maximizing symmetry. Later research [21] also looked at edge continuity and its effect on path finding in graphs.

• *Frank van Ham is at IBM Research in Cambridge, MA and can be reached by e-mail at fvanham@us.ibm.com*
• *Bernice Rogowitz is at IBM Research in Hawthorne, NY and can be reached by e-mail at rogowtz@us.ibm.com*

C1-Force (1 connecting edge)

C2-Force (2 connecting edges)

C3-Force (3 connecting edges)

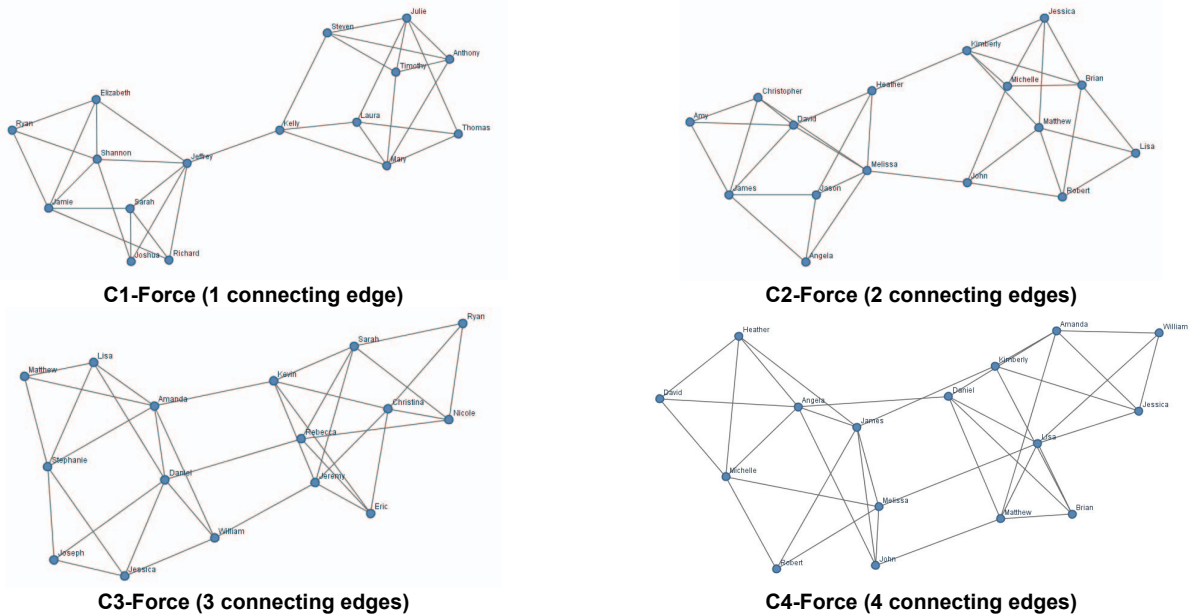**C4-Force (4 connecting edges)**

Figure 1. The observer dynamically moved the nodes of a graph to best represent the relationships between nodes. This table shows the four force-directed graph initial layouts. Clusters in the data are connected by 1, 2, 3, or 4 "masking" edges

In a study of UML-like software engineering diagrams, Purchase [12] showed that the influence of diagram semantics could have a much larger impact than the influence of aesthetics. In this study, users did not care so much that their diagrams had non-uniform edge lengths, but found the perceptual grouping of related classes much more important. In a study of social networks, McGrath et al [7] used an online questionnaire to evaluate how people perceive groups in networks. Their study shows people prefer graphs with fewer edge crossings where nodes sharing social connections were grouped together. Huang et al [3], however, have shown that simply preferring a layout doesn't necessarily correlate with the best task performance and pointed out that more empirical work was needed to understand the effectiveness of visual network layouts. In dynamic graph drawing, where the structure of the network varies over time, the trade-off is different and designers have to balance optimizing a set of aesthetics with minimizing changes to the user's mental map of the structure [13]. Finally, other studies have pointed out that in particular cases it may not make sense to use node link diagrams at all, and other representations are preferable for a specific task set [2].

All of the above studies evaluate known graph layout aesthetics on particular well-defined tasks. Although it's very useful to know which types of layouts perform well for which type of tasks, the results are limited by the particular conditions set by the experimenters. Studying how users organize network diagrams, without imposing any constraints on their structure should reveal interesting information about the importance of edge crossings, node uniformity and the importance of representing clusters, and might provide guidance to network layout algorithm design

3   EXPOSITION

The goal of this paper is to explore the importance of certain visual characteristics to users whose task it is to organize nodes and edges of a graph into a coherent picture representing the structure in the data. First, we examine the degree to which users actually minimize edge crossings and the degree to which they try to keep edge lengths uniform. Secondly, we examine the ability of human observers to discover connected groups of nodes, or clusters, in a network, and examine how this ability depends on the degree to which this information is masked by an increasing number of inter-cluster edges.

In these experiments, users viewed node-link graphs on a computer screen and manipulated the node positions until they felt that the structure best captured the structure in the data. The nodes were presented to the users in one of two different initial configurations. In the circular configuration, nodes were randomly placed in a circular arrangement and in the force-directed condition nodes were place according to an algorithm that tries to maintain uniform edge lengths and optimal node separation. Comparing user layouts with those produced by the lay-out algorithms allowed us to examine which visual characteristics are important for the perceptual organization of network data.

We manipulated the internal structure of the data, creating "clusters" of nodes that were highly interconnected. By manipulating the number of edges connecting these clusters, we varied the prominence of the clusters. This allowed us to examine how sensitive users were to the organizational structure in the data. Our initial hypotheses in this setup were the following:

- Based on previous findings [11], we hypothesized that human observers would seek to reduce edge crossings, and would produce layouts with fewer crossings than the basic force-directed algorithm, which did not optimize for this particular feature.
- We knew from the perceptual organization literature that human observers are responsive to spatial groupings and predicted that they would be able to recover these structures in the graph. Accordingly, we expected that the observers would organize nodes belonging to the same cluster preferentially into spatially-localized regions, especially in the conditions with fewer between-cluster connecting edges.
- Finally, we hypothesised that the distance between these regions would decrease if we increased the number of between-cluster edges. If true, this would provide us with a welcome design guideline for clustered graph drawing algorithms and diminish the importance of the uniform edge length aesthetic.

4   EXPERIMENTAL DESIGN

Partly inspired by the popularity of the online 'Planarity' game [17] we used a web-based interactive environment to present our participants with their tasks instead of a more traditional laboratory setup. This allowed us to test a large number of users in a relatively

short period of time [4]. During a period of 14 days all users visiting the homepage of the Many Eyes collaborative visualization site [19] were invited to participate in the experiment by means of a banner. If they choose to do so, a mouse click brought them to a webpage outlining the experiment. The instructions explained that we were measuring how people perceive and organize networks, and that they were going work with several network diagrams showing relationships between people. For each, they were asked *"to rearrange the nodes in a network in a way that you think best reflects their interconnections"*. We then briefly explained the controls and finally presented the user with four links to four experimental conditions. The order of the links was varied randomly from subject to subject to prevent order bias, and the users had no way of knowing what condition they were going to be shown when clicking a link.

4.1 Conditions

In each condition of the experiment, the user was presented with an existing graph layout of a network consisting of 16 nodes. Each data set was specifically designed to include two "clusters". That is, in each 16-node data set, there were two groupings of 8 nodes which were highly interconnected. Within a cluster, each node was connected to several (3-5) other cluster members. The within-cluster node connections were constant across the various conditions in the experiments.

The eight conditions in the experiment are shown in Table 1. We varied the degree to which the nodes within the first cluster were connected to nodes in the other cluster. In conditions C1, C2, C3 and C4 there were 1,2,3, and 4 edges connecting to nodes in the other cluster, respectively. These additional edges served to obfuscate or "mask" the cluster structure. We also varied the starting configuration of the nodes to explore how the initial presentation of the data influences organizational strategy. In half of the conditions, the starting structure for the experiment was a circular arrangement of nodes in random order; in the other half, the starting structure was computed using a popular force-directed layout algorithm that tries to minimize edge-length variability. In this case we used a stress majorization variant [1] for its ability to compute good optima. We did not perform any post processing such as crossing reduction, better spacing or label overlap removal. Both layouts were selected to provide starting points that did not emphasize any structural features by design. We did not test more sophisticated (e.g., clustered) layout algorithms because the point was to see whether users would recover the cluster structure. Also, our main objective was to measure how human observers would lay out a network, not to compare the perceptual quality of various network lay-out algorithms. Figure 1 shows layouts of the four conditions generated by the force-directed algorithm. Figure 2 shows the two starting positions for the data set with two masking edges.

4.2 Procedure

For each condition, the observers' task was to move the nodes of the network until they thought the graph best represented the structure of the data. The user was allowed to enter commentary on the submitted layout if they so desired and then saved their layout by clicking a button at the bottom of the page. There was no preset time limit and users were told they could take as long as they wanted.

We choose to label every node with an arbitrary male or female name, taken from a set of randomized first names to make the problem more concrete to users. In early pilots which did not include labels, we found that some users had difficulty performing the test because they had no concept of an abstract network, or what the optimal layout of such a network should be. To prevent users from using the labels to remember the structure over different tasks, we randomly assigned new labels to nodes whenever a new condition was shown.

Table 1. Factors and conditions

	Initial Layout	
	Force directed	Circular
1 intercluster edge	*C1-Force*	*C1-Circular*
2 intercluster edges	*C2-Force*	*C2-Circular*
3 intercluster edges	*C3-Force*	*C3-Circular*
4 intercluster edges	*C4-Force*	*C4-Circular*

4.3 Participants

The participants in this experiment were mostly self-selected from the pool of users of the Many Eyes visualization service, so the population tested did had at least some sophistication in data visualization [17], possibly limiting the ability to generalize our experimental results. However, the pool also included a small number of friends and colleagues, and where it was possible to compare, no systematic differences between the two groups was observed. Although some of our participants had created Many Eyes profiles for themselves on a prior visit to the site, most of them (68%) did not and, as a consequence, their results were completely anonymous. Although this increases the risk for duplicate submissions, we estimate it as low, given the time ($\mu = 409$ sec; almost 7 minutes) and effort required to create and submit a layout. To make sure, we hand pruned the data to discard submissions that were obvious duplications due to users clicking the submit button twice (i.e. two identical layouts submitted within a short timeframe) and found one such duplicate. This left us with 73 submitted layouts.

5 METRICS

We designed metrics to capture characteristics of user performance when organizing graph layouts. These metrics allowed us to examine the layout strategies of human observers compared with a force-directed layout algorithm, gave us insight into the degree to which observers responded to cluster structure in the data, and gave us a first indication of how their visual attention was directed.

5.1 Layout Metrics

The first set of metrics examines the strategies employed by users when organizing graph lay-outs. In particular, we measured their use of edge crossings, the spatial positioning of nodes and edges and the orientation of the final configuration, and compared these with the initial layout.

Edge crossings: To evaluate the degree to which human observers tolerated edge crossings, we counted the number of edge crossings in each submitted layout.
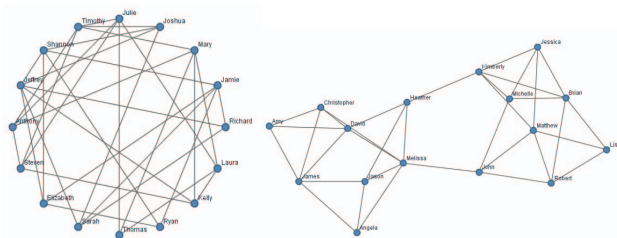

Figure 2. The circular and force-directed starting layouts for the 2-edge conditions: C2-Circular (left) and C2-Force (right).
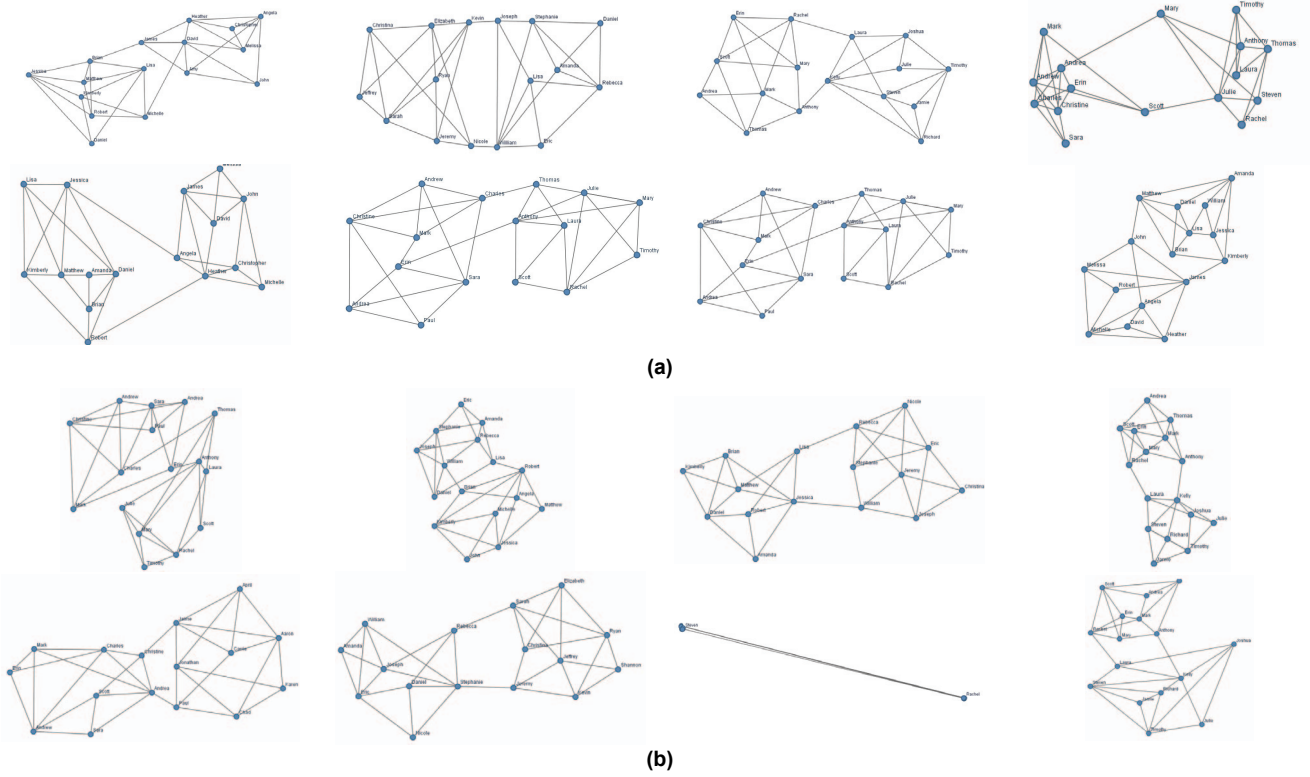
Figure 3. A number of user-generated layouts with two connecting edges between clusters for **(a)** the circular starting layout (C2-Circular) and **(b)** the force-directed layout (C2-Force).

Edge length distribution: To measure the degree to which observers sought to keep equal distances between the nodes we measured the distribution of edge lengths. For a graph with uniform edge lengths the variance in edge lengths will be low.

Orientation: To measure whether the graph is oriented on the page, or is isotropic, we computed a convex hull for the graph layout, approximated it by an ellipse, and then measured the orientation of the ellipse's major axis.

5.2   Cluster Metrics

The second set of metrics looked at whether observers were sensitive to the cluster structure in the data, and how they represented this structure.

Cluster Separability: To evaluate the degree to which human observers spatially separated clusters, we counted how many of a cluster's nodes fell inside the bounding polygon defined by the other clusters' outer nodes.

Cluster Extraction: To estimate the degree to which users identified and isolated sets of interconnected nodes in the dataset we compared the average inter-node distance within a cluster with the average inter-node distance for the whole graph. If users manage to isolate clusters correctly we should see a distinct difference between these two values.

Cluster Distance: To evaluate the separation between clusters we measured both the minimum and average inter-node distance between nodes in the two clusters. This gives us an estimate of the relative positioning of the two clusters.
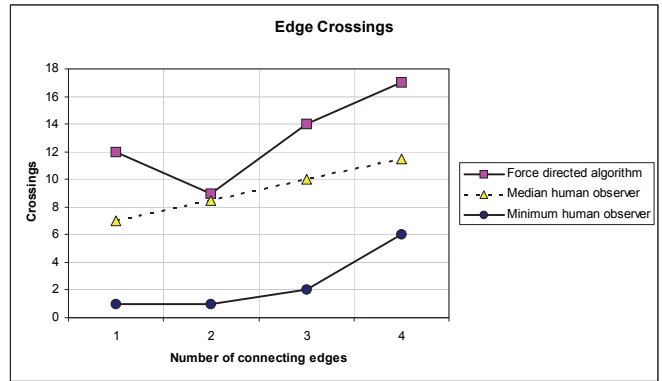

Figure 4. Edge Crossings. Human observers produced graph layouts with fewer edge crossings than the force-directed graph algorithm.
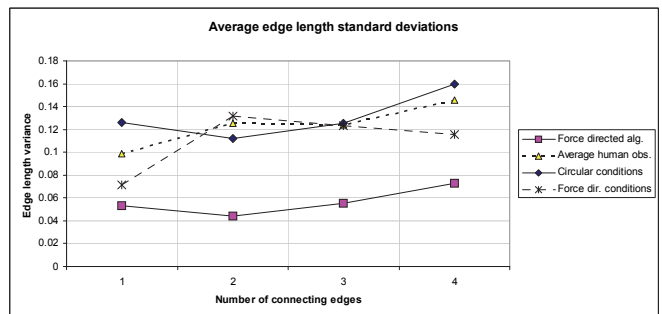

Figure 5. Edge Length Distribution. Human observers did not focus on maintaining equal edge length as much as the force directed algorithm.

Cluster Delineation by Hulls: To evaluate the perceptual organization of clusters we examined whether the set of nodes in each cluster were visually bounded by a closed segment of edges. We measured the number of clusters per dataset that were bounded by a closed segment of edges.

To account for differences in layout scales we normalize all node distances, by dividing them by the maximum node-node distance found in the graph. This allows us to directly compare distances among different layouts.

6 RESULTS

The experiment ran for two weeks, during which time 73 network diagrams were submitted. Observers were encouraged to participate in all four test conditions, but there is no way to know whether twenty anonymous observers submitted on all four conditions or whether eighty anonymous observers each submitted one. The randomization algorithms, however, produced an even distribution of observers over the eight experimental manipulations (four levels of cluster connection and two starting conditions for each). Overall, inter-subject reliability was very high, although, as to be expected, there were several cases where user behaviour was unique and far from the average. These data have all been included in the analysis. The only data values that were edited were two measurements of completion time, where the user supposedly finished the tasks more than 3 hours after initiating the test. In this case, we substituted the mean completion time for this task for these values.

Figure 3 shows some of the solutions produced by observers for conditions C2, with two connecting edges. Figure 3a shows solutions obtained from condition C2-Circular; Figure 3b shows solutions resulting from condition C2-Force. Note that one of the configurations in figure 3b is actually a degenerate case where the user either did not understand the purpose of the experiment or went out of their way to do something creative.

6.1 Perceptual Factors in Network Layout

Edge crossings. To begin this analysis we compare the performance of human observers with that of the force-directed layout algorithm. Figure 1 showed the four starting configurations produced by the force-directed algorithm. The number of edge crossings in these graphs is displayed as a solid line in Figure 4. The median human response is shown in dotted lines. Some observers created graphs that had very few crossings. In conditions C1, two observers created solutions with just one edge crossing; two observers did the same for condition two; the best solution for C3 was two edge crossings, achieved by 2 observers; and the best solution for C4 had six edge crossings.

At the other end of the spectrum, there were one or two observers per condition whose graphs were highly crossed, in two cases in excess of 100 line crossings. To avoid having these extreme values obscure the behaviour of the pool of observers, we represented these as the median. The interesting observation from this graph is that human observers strive to reduce edge crossings, as much as possible, which is consistent with results from [11]. On average, they produce layouts with fewer crossings than the force-directed graph algorithm, which prioritizes uniform edge lengths and separation between unconnected nodes. On average 67% of observes created embeddings with fewer edge crossings than the algorithmic layout, even when they were presented with the algorithmic layout to start with, which confirmed our first hypothesis.

Edge length distribution. When we examined the standard deviation in edge lengths we found that in all but 2 of the 73 of the user generated embeddings, the variance in edge lengths was bigger than the variance in the optimal force directed layout. Figure 5 plots the variance in edge lengths for the force directed layout and the average variance in the user generated layouts. Overall, all user generated diagrams have a significantly ($p < 0.05$) bigger variance in edge lengths than the force directed layouts. Looking at the diagrams
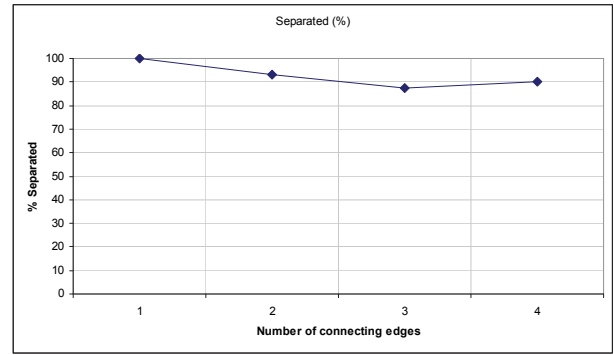

Figure 6. Cluster Separability. Even with four connecting edges between clusters, observers are able to isolate the clusters in the data, with at most 1 node from one cluster (12%) positioned in the polygon defining the other.
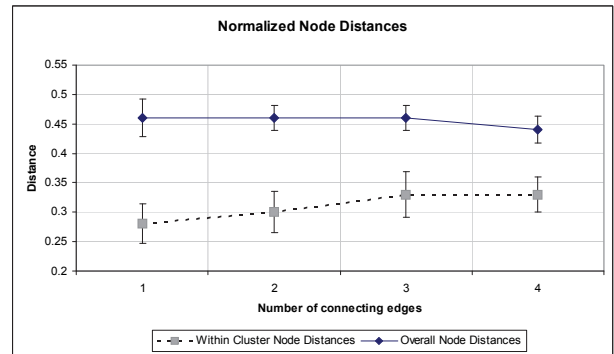

Figure 7. Cluster Extraction. For all levels of masking, the distance between nodes within a cluster is significantly smaller than the overall inter-node distance, demonstrating perceptual grouping. Error bars show 95% confidence intervals
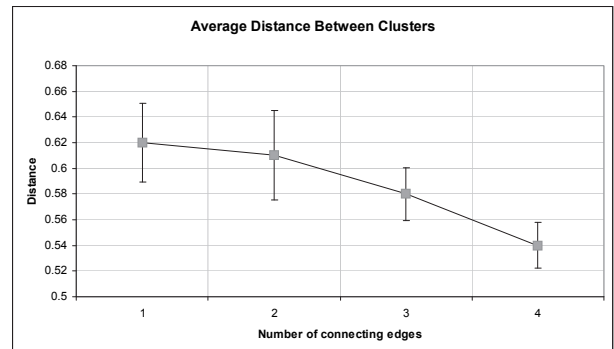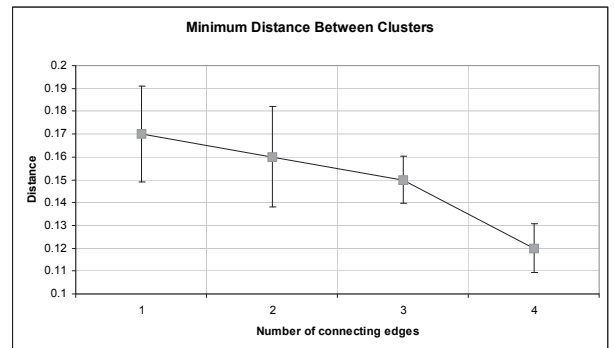

Figure 8. Cluster Distance. As the number of connecting edges between clusters increases, both the minimum distance (top) and average distance (bottom) between the clusters decreases, although only significantly for the 4-edge condition. Error bars show 95% confidence intervals

and posted comments we speculate that this is due to users putting a lot of effort into generating symmetrical or crossing-free diagrams. To see if users could actually be primed to generate drawings with uniform edge lengths if they were initially shown a force directed layout, we also examined the results by initial layout. For the C1 conditions the edge length variance was significantly smaller (p < 0.01) when the observer started from a force directed layout while we did not find a significant difference for the other conditions. For the C2-Force condition the average variance was actually bigger, mainly due to one of the degenerate cases displayed in Figure 3b. These results suggest that users are not that sensitive to edge length uniformity and would rather trade it off for fewer crossings or more symmetrical layouts.

Orientation. The starting orientation for all the force-directed layouts was roughly 60 degrees. For conditions C1-Force and C2-Force over 70% of the observers maintained this orientation in their layouts. However, as the amount of masking between clusters was increased, the proportion of horizontal and vertical orientations increased, and with four edge crossings, 90% of the user-generated graphs placed clusters either side-by-side or one above the other. With the circular starting layout, users systematically preferred horizontal orientations, especially in the two-and three-edge conditions, where 70 and 90% of the user-constructed layouts were oriented horizontally. This may suggest users have a preference for horizontal and vertical oriented layouts, although our horizontal screen aspect ratio may have been a contributing factor here.

Finally, we looked at the time it took user to complete a layout. For the circular conditions we computed a mean of 620 seconds with a standard deviation of 52. For the force-directed condition the mean was 322 with a standard deviation of 53 seconds. As expected, users took significantly longer on the circular layouts, although the difference in results was not that big.

6.2   Identifying and Representing Clusters

In these experiments, two highly-interconnected sets of nodes were present in each of the data sets. In the four test conditions, additional edges were added which connected a node in one cluster to a node in the other. This section analyzes the degree to which observers organized their visual representations in a way that preserved the cluster structure in the data.

Cluster Separability. To examine the salience of the cluster structure, we measured how the nodes in the two clusters were represented visually. To do so, we measured the degree to which the clusters were spatially separated in the users' solution. If all nodes within a cluster are organized near each other, with no node crossing into the spatial region occupied by nodes from the other cluster, we say that the clusters are separable. Figure 6 shows the percentage of user solutions where the clusters were separable. User solutions are 100% separable when there is only one connecting edge, and separability drops slightly over the range. Even with four connecting edges, however, only 2 of the 20 user solutions exhibited overlap of nodes from one cluster within the region defined by the other. Not surprisingly, these were the same solutions that exhibited high numbers of edge crossings, described above.

Cluster Extraction. Another way to look at the degree to which users identified and spatially isolated clusters is to look at the distribution of distances between nodes. If users arranged nodes randomly, ignoring the inherent cluster structure, the nodes that were part of a cluster would not be placed especially close to each other in the network. That is, the inter-node distance between nodes within the same cluster would be the same as the overall inter-node distance. To evaluate this, we computed the inter-node distance for each user solution, and averaged the result across observers. Figure 7 plots the mean inter-node distance for all nodes (solid lines) vs. the inter-node distance for nodes within a cluster (dashed lines).
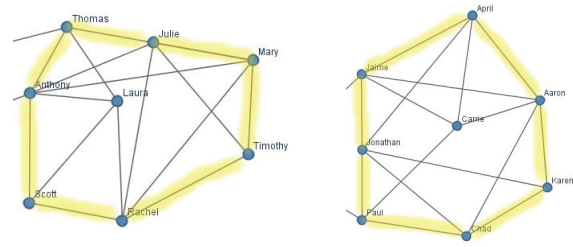

Figure 9. Cluster Hulls. Two examples of user-generated layouts where cluster edges formed a hull enclosing the cluster, organizing it into a single perceptual group.

Independent of how many inter-connecting edges there were, nodes within a cluster were organized significantly closer together than the norm. The difference between these measures is greatest when there is only one edge, and decreases monotonically as the number of connecting edges is increased. Post-hoc normality analysis and pairwise t-testing showed no significant differences between the overall node distances over the four different conditions (which was expected). Conform our second hypothesis, the distance between nodes in a single cluster indeed grows if the number of connecting edges is increased. However, our analysis showed that this effect was not significant at the $p = 0.05$ level.

Cluster Distance. Figure 8 shows two additional measures of cluster salience as the number of interconnecting edges is increased. The first measure, shown in dashed lines, identifies the node from each cluster that is closest to a node from the other cluster and computes the distance between them. The distance between clusters measured this way is largest when there is only one edge connecting clusters and drops monotonically as the number of connecting edges is increased. The second measure, shown in solid lines, calculates the average position of nodes that belong to each cluster, and computes the difference between these average positions. Subsequent Welch t-testing showed a significant (p<0.05) difference in both measures for the C4 conditions only. The greater the number of interconnecting edges, the closer the clusters, which confirmed our third hypothesis.

Cluster delineation by hulls. One interesting observation that we were not expecting at the start of this experiment is that a lot of users took special care to visually bound the clusters in the data by creating a continuous 'silhouette' of edges around their clusters, with all of the other nodes in the cluster positioned inside this hull. (See Fig. 9) Slightly over 81% of our users created hulls for one or more clusters in the data, with an average of 1.34 hulls per dataset over all conditions. If we compare this with the force directed layouts we find a total of 0.16 (1 out of 8) hulls per dataset.

From a perceptual standpoint this behaviour makes perfect sense as the human visual system is highly efficient at detecting and interpreting shape contours. In this case our human subjects are trying to create clearly delineated visual entities to indicate a cluster. Previous experiments that only measured layout algorithm effectiveness would not have surfaced this type of insight, as the input stimuli are generated by known layout algorithms. Although layout algorithms exist that start from a convex hull and then render the entire graph inside this hull [16][18] we are unaware of any layouts that render individual clusters in this manner.

7   DISCUSSION AND CONCLUSION

We have developed a direct-manipulation method for exploring perceptual factors at work in representing network information. In each condition of the experiment, the observer was presented with a 16-node directed graph and was asked to manually adjust the positions of the nodes until he or she felt the structure best captured the relationships in the data. The structure of the data was designed to include two sets of highly-interconnected nodes, or clusters, and

the number of inter-cluster "masking" edges was varied. Two popular layout algorithms defined the starting configuration for the nodes. As expected, the random configuration, in which nodes were evenly spaced on a circle, took the most time to complete. The directed graph algorithm, which organized nodes according to their connectedness, while minimizing the variability in edge length, provided a better starting point for the observers, and afforded consistently shorter completion times. However, despite this enormous difference in solution times, the actual solutions were not substantively affected by starting position.

The solutions the human observers created, however, were significantly different from the starting configurations in several ways. First, they were highly structured and did not resemble the random configuration at all. In many ways, the Force-directed algorithm provided a very good model of human behaviour, in terms of the spatial distribution of the nodes and the separation of the clusters. Human performance, however, did differ notably in two respects. First, human solutions contained 60% fewer edge crossings on average and secondly, humans did not value uniform edge length as much as the algorithm did.

One of the main findings in this study is that users are able to perceive and operate on the cluster structure in the data, even if it is obscured by extraneous edges. In almost all cases users constructed layouts that distinctively grouped clusters in a spatial region that did not overlap with the spatial region occupied by the other cluster. More importantly both the minimum and average distance between the two clusters is inversely related to the strength of their clustering, supporting layout techniques like Lin-Log layouts [9] or clustered layouts. In all cases but one, the mean edge length of edges connecting two different clusters was bigger than the mean edge length connecting nodes inside a single cluster. This leads us to believe that maintaining uniform edge lengths has less value when it comes to organizing network information that contains cluster structure. A second important finding from this study is that over 80% of observers used the edges in a cluster to visually delineate the cluster itself, creating a convex hull around the nodes in the cluster. Although this might seem obvious in retrospect, it would have been much harder to surface these types of layouts had we just looked at evaluating known layout techniques or aesthetics.

In evaluating these results, it is important to keep in mind that because these experiments were conducted over the internet through the Many Eyes visualization service [20], they may not be representative of the general population. This has previously been identified as general issue with Internet based experimentation [15]. To mitigate this potential bias, we did invite people who were not Many Eyes users to participate, but we did not collect basic population statistics.

Another potential confounding factor was the inclusion of names labelling each node. Although we were careful to randomize names over different tests to avoid learning effects, the tendency to avoid overlapping labels could potentially have influenced the resulting layouts. The labels may also have influenced some of the more "unique" graph layouts, and some users may have believed that their task was more related to the names associated with the nodes than the relationships between the nodes. For example, one user created a layout pairing male and female names.

In conclusion, these experiments show that although human observers create unique layouts, they very reliably converge on representations that share important similarities. They seek to reduce edge crossings, they enclose organizations within bounding hulls, and they are extremely competent at identifying and spatially segmenting clusters in the data. In terms of automatic layout algorithms, this suggests focus on algorithms that minimize edge crossings, suggests introducing methods that emphasize structures by manipulating the nodes so that these regions are bounded by edges, and it supports research on algorithms that are designed to reveal clustered structures. These results demonstrate the importance of research in visual perception and cognition in developing better methods for visualization.

REFERENCES

[1] E. Gansner,Y. Koren, and S. North, Graph Drawing by Stress Majorization, *Proceedings of 12th Int. Symp. Graph Drawing (GD'04)*, LNCS 3383, pp. 239-250, 2004.

[2] M. Ghoniem, J.-D. Fekete and P. Castagliola, A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations, *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04)*, pp. 17-24, 2004.

[3] W. Huang, S.-H. Hong and P. Eades, Layout Effects on Sociogram Perception, *Proc. Symp. Graph Drawing GD '06*, pp. 262-273, 2006.

[4] R. Kohavi, R. Henne, and S. Sommerfield, Practical Guide to Controlled Experiments on the Web: Listen to your customers not the HiPPO, Proceedings KDD07, 2007.

[5] W. Kohler, Physical Gestalten. *Ellis, W.D. (ed) A Sourcebook of Gestalt Psychology*, The Humanities Press, pp. 17-54, 1950 (originally published in 1920).

[6] M. Kubovy and S. Gepshtein. Perceptual Grouping and Space and in Space-Time: An Exercise in Phenomenological Psychophysics, *Perceptual Organization in Vision*, pp 45 – 86, 2003.

[7] C. McGrath, J. Blythe and D. Krackhardt, Seeing Groups in Graph Layout. *Connections*, 19(2) pp. 22-29, 1996.

[8] G. Miller, The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information, *Psychol Rev. 63(2)*, pp. 1-97, 1956.

[9] A. Noack, An Energy Model for Visual Graph Clustering, *Proceedings of the 11th International Symposium on Graph Drawing (GD 2003)*, In G. Liotta, (editor) LNCS 2912, pp. 425-436, 2004.

[10] S. Palmer, Perceptual Organization and Grouping, *Perceptual Organization in Vision*, pp 3-44, 2003.

[11] H.C. Purchase, "Which Aesthetic Has the Greatest Effect on Human Understanding?" *Proc. Symp. Graph Drawing GD '97,* pp. 248-261, 1998.

[12] H.C. Purchase, M. McGill, L. Colpoys and D. Carrington, Graph drawing aesthetics and the comprehension of UML class diagrams: an empirical study, *Proceedings of the Australian Symposium on Information Visualisation,* Eades, P. and Pattison, T. (eds), Australian Computer society, pp. 129-137, 2001.

[13] H.C. Purchase, E. Hoggan. and C. Görg. How Important is the "Mental Map"? - An Empirical Investigation of a Dynamic Graph Layout Algorithm. *Proceedings of 14th International Symposium on Graph Drawing,* 2006.

[14] H.C. Purchase, J.-A. Allder, and D. Carrington, D. Graph Layout Aesthetics in UML Diagrams: User Preferences, *Journal of Graph Algorithms and Applications* vol. 6, no. 3, pp. 255-279, 2002.

[15] U.-D. Reips, Standards for Internet-based Experimenting, *Experimental Psychology, 49 (4)*, pp. 243-256, 2002.

[16] R. Tamassia, On Embedding a Graph in the Grid with the Minimum Number of Bends, *SIAM J. Comput. 16(3)*, pp. 421-444, 1987.

[17] J. Tantalo, Planarity, *http://www.planarity.net/* (retrieved June 08)

[18] W.T. Tutte, How to Draw a Graph. *Proc. London Math. Soc. 13*, pp. 743-768, 1963.

[19] F.B. Viégas, M. Wattenberg, F. van Ham, J. Kriss and M. McKeon, ManyEyes: a Site for Visualization at Internet Scale, *IEEE Transactions on Visualization and Computer Graphics* Vol. 13, No. 6, pp. 1121-1128, 2007.

[20] F.B. Viégas, M. Wattenberg and J. Kriss, Your Place or Mine? Visualization as a Community Component, *Intl Conf on Computer Human Interaction (CHI 08)*, pp. 275-284, 2008.

[21] C. Ware, H.C. Purchase, L. Colpoys, and M. McGill, Cognitive measurements of graph aesthetics. *Information Visualization*, vol. 1, no. 2. pp 103-110, 2002.

[22] M. Wertheimer, Untersuchungen zur Lehre von der Gestalt, *Psychology Forschung, 4*, pp. 301-350, 1950. (originally published in 1923).