

# CrowdLayout: Crowdsourced Design and Evaluation of Biological Network Visualizations

Divit P. Singh<sup>1,2</sup>, Lee Lisle<sup>1</sup>, T. M. Murali<sup>1</sup>, Kurt Luther<sup>1</sup>

<sup>1</sup>Department of Computer Science, Virginia Tech, Blacksburg, VA, USA

<sup>2</sup>Amazon.com, Seattle, WA, USA

{divit52, llisle, murali, kluther}@vt.edu

## ABSTRACT

Biologists often perform experiments whose results generate large quantities of data, such as interactions between molecules in a cell, that are best represented as networks (graphs). To visualize these networks and communicate them in publications, biologists must manually position the nodes and edges of each network to reflect their real-world physical structure. This process does not scale well, and graph layout algorithms lack the biological underpinnings to offer a viable alternative. In this paper, we present CrowdLayout, a crowdsourcing system that leverages human intelligence and creativity to design layouts of biological network visualizations. CrowdLayout provides design guidelines, abstractions, and editing tools to help novice workers perform like experts. We evaluated CrowdLayout in two experiments with paid crowd workers and real biological network data, finding that crowds could both create and evaluate meaningful, high-quality layouts. We also discuss implications for crowdsourced design and network visualizations in other domains.

## ACM Classification Keywords

H.5.3 Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces

## Author Keywords

Crowdsourcing; design; computational biology; networks; graphs; graph drawing; visualization; citizen science.

## INTRODUCTION

Working with network (graph) data is extremely common in modern life sciences research. In the interdisciplinary field of network biology, scientists use graphs to represent a wide range of biological phenomena, from the effects of chemicals on human cells to the flow of electricity among neurons in the brain. Biologists often create visualizations of their networks to analyze them and communicate them to other scientists in publications.

However, creating a meaningful visualization of a biological network remains a difficult challenge. Displaying the nodes and edges in a way that minimizes occlusion and intersection is part of the problem. Even a relatively small network of 50 nodes can produce an indecipherable “hairball,” and many biological networks contain hundreds or thousands of nodes and edges. Adding to this challenge, biologists often also want to create a meaningful *layout* of the network, where the positions of nodes and edges relative to one another mimic the underlying biological knowledge.

To overcome these challenges, biologists typically rely on two techniques: (i) automatically generating a layout using a graph drawing algorithm (Figure 1a and b), and/or (ii) manually creating a layout (Figure 1d). Both techniques have significant shortcomings. While graph drawing continues to be an area of active research, most algorithms do not take advantage of the biological information in the underlying network data, so their results are too generic to be meaningful. Biologists may instead manually create a layout, or improve an automatically generated one. However, this process does not scale well because it is time-consuming, lacks specialized tools, and typically requires scientific expertise.

In this paper, we propose and evaluate CrowdLayout, a system that uses crowdsourcing to provide scalable, high-quality, meaningful layouts of biological network data. With CrowdLayout, we frame the goal of meaningful visualizations of biological networks as a visual design problem, and leverage the human intelligence and creativity of crowd workers to find solutions within that design space. CrowdLayout recruits paid novice crowd workers from Amazon Mechanical Turk and provides flexible design guidelines, abstractions, and specialized tools to help them create new layouts and review the quality of other workers’ layouts. As a result, biologists can leverage CrowdLayout to visualize many more networks than they previously could, potentially leading to new scientific breakthroughs. Our techniques may also generalize to support network analysis in other domains.

We evaluated CrowdLayout with two experiments. The first ( $n = 30$ ) found that crowdsourced layouts of real biological networks were significantly better than two popular graph drawing algorithms, and nearly as good as layouts designed by an expert computational biologist. A second experiment ( $n = 330$ ) found that crowds provided quality ratings of network layouts that were similar (within 2–10%) to another expert

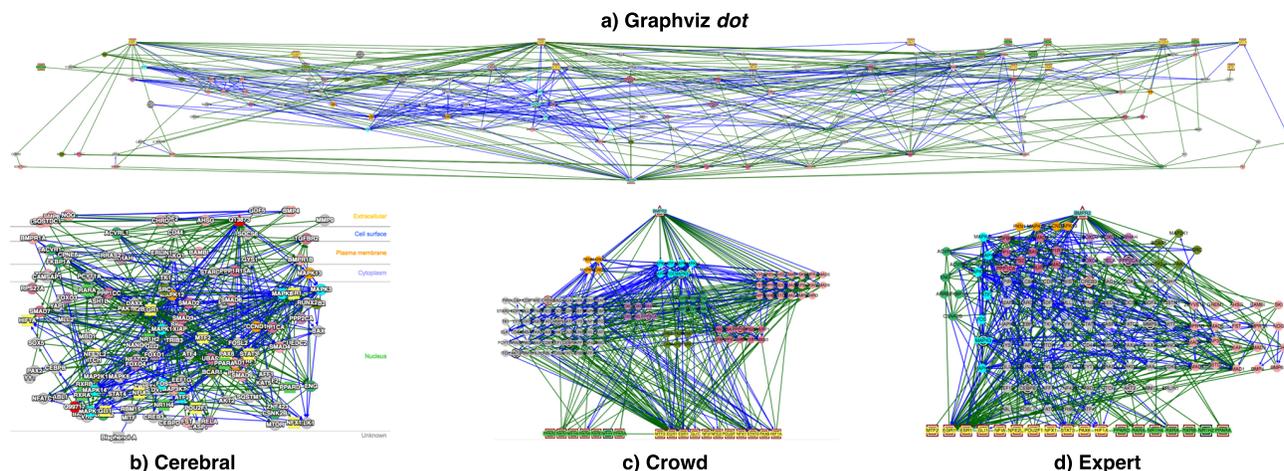
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI 2018 April 21–26, 2018, Montreal, QC, Canada

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5620-6/18/04.

DOI: <https://doi.org/10.1145/3173574.3173806>



**Figure 1.** Four layouts of the same biological network, all representing the response of cells to the compound Bisphenol-A, produced by different mechanisms. Layout (a) uses the popular open-source GraphViz *dot* algorithm. Layout (b) uses Cerebral, an algorithm that incorporates knowledge of cellular locations of proteins. Novice crowd workers used CrowdLayout to create layout (c). An expert computational biologist created layout (d).

computational biologist, suggesting that CrowdLayout can reliably create and identify high-quality layouts.

Our contributions include: (i) novel techniques for generating and evaluating scalable, high-quality biological network visualizations via novice crowdsourcing, (ii) experiments providing empirical evidence of the benefits of these techniques compared to expert gold standards and algorithmic baselines, (iii) implications for crowdsourcing complex design work and network visualization, and (iv) the CrowdLayout system itself, released as open-source software.

## RELATED WORK

### Automatic Graph Visualization

Biologists use graphs with directed or undirected connections between pairs of molecules: a protein is a node, a physical interaction between two proteins is an undirected edge, and a directed edge represents a control relationship [48, 1]. Almost every publication in this field includes a visualization of a network in which the nodes and edges are laid out in two dimensions. This layout is usually performed by an automated algorithm that assigns  $x$  and  $y$  coordinates to each node while ensuring that nodes do not overlap and that nodes connected by an edge are close to each other. Popular methods include force directed and spring embedded layouts. These methods are implemented in several systems such as BiNA [22], Cytoscape [44], Gephi [3], Graphviz [18], NetBioV [45] Pajek [12], and VisANT [25].

A major drawback with these visualizations is that layout algorithms incorporate almost no knowledge of the biological information underlying the networks, making the layouts difficult to decipher even for moderately sized networks (Figure 1a and b). Consequently, researchers annotate the networks with additional information, e.g., cellular functions performed by the proteins. Moreover, researchers often use their biological knowledge and intuition to modify the node and edge positions manually to bring out salient features. For example, the manual layout in Figure 1d suggests that a flow of information

from a receptor (a protein located at the cell’s outer membrane to sense signals, red triangle) through interactions with intermediate proteins (colored and grouped according to specific functions they perform) to transcription factors (green and yellow rectangles at the bottom, which execute the cell’s response to the chemical). This situation persists despite decades of research in graph drawing and the wide use and maturity of layout algorithms. Some recent efforts have improved the interpretability of network visualizations, e.g., zoomable layouts based on grouping nodes together [25], methods to draw graphs with grouped or “compound” nodes [13], and network organization by the location of proteins in the cell [20, 62] (see also Figure 1b).

### Humans and Graph Visualization

While much graph layout research focuses on automated techniques, a smaller set focuses on manual or mixed-initiative graph layouts. For example, Apollo combines rich user interaction with machine learning to help users explore networks of related research papers [8], and Van Ham and Perer developed a visualization tool that leverages Degree of Interest functions to support analysis of very large networks of legal citations [46]. While most of these systems are geared towards domain experts and focus on laying out networks based on semantic relatedness (e.g. clusters of related documents), CrowdLayout focuses on novice crowds and laying out networks based on protein functions and locations in the cell. Few attempts have been made to construct graph layouts using crowdsourcing. A notable exception is Yuan et al. [60], whose interactive technique decomposes a complex graph into subgraphs which are laid out by crowd workers and then merged using an algorithm. We build on this work by investigating how unaided crowds can follow high-level design guidelines and assess layout quality.

Prior work has investigated how humans, including non-experts, make sense of network visualizations [17, 47, 42, 40, 26]. This work has established, for example, that network layout design is an acquired skill and most people do not per-

form well at first, but become better with practice. We extend this work by providing design guidelines, abstractions, and specialized editing tools to help non-expert crowds become effective layout designers. This work also reports mixed results in quality analyses of human- vs. algorithm-generated network layouts. Our research sheds new light on these analyses through comparisons of layouts generated by crowds, experts, and algorithms.

A related thread of research considers what individual or synergistic properties increase the communicative potential of network visualizations. While it is widely acknowledged that aesthetics are a key consideration in effective network layouts [43], aesthetics are difficult to translate into equations [60], and strict rule-based approaches are often conflicting and brittle [27]. Recognizing this inherent subjectivity, we draw on the rich body of literature around creative practice and design critique and apply these findings to crowdsourced network layout. For example, we leverage empirical findings like the importance of parallelizing early stage design and considering diverse alternatives [15, 16]. We also push forward the growing research on crowdsourced design critique [23, 38, 59, 56], by extending these concepts beyond graphic design (posters, flyers, slide decks, etc.) to network layout in general, and biological networks in particular.

### Citizen Science and Crowdsourced Data Analysis

Citizen science refers broadly to the idea of non-expert volunteers contributing data or analysis that supports scientific research [10]. The proliferation of internet access and mobile devices has greatly increased the scale and diversity of citizen science projects [6, 51, 31, 37, 35]. For example, the online citizen science portal Zooniverse [36] boasts more than one million volunteers who have contributed to nearly 50 projects, including several in biological domains. FoldIt [29], a citizen science project designed as a game in which human players fold proteins — a difficult task for algorithms — has attracted more than 200,000 players. We take inspiration from these successes and investigate the potential of biological network visualizations as a citizen science project. (While this paper demonstrates feasibility with paid crowd workers, we plan to eventually target volunteers.) CrowdLayout’s layout review techniques also contribute to the nascent literature on community-based data validation [52].

Beyond citizen science, other research leverages crowds to perform collaborative analysis of visualizations [53, 54]. For example, ManyEyes [49, 11] allowed expert and non-expert users to create their own data visualizations and share them with others, and systems like Sense.us [24] and CommentSpace [55] provided users with structured support for collaboratively analyzing visualizations of economic statistics and other datasets. However, none of these projects focuses on network data or network layouts except Van Ham and Rogowitz [47], who focus on individual users rather than collaboration or crowds.

### CROWDLAYOUT SYSTEM DESCRIPTION

CrowdLayout is a web-based system for crowdsourcing the design and review of layouts for biological network visualizations. It was developed as an extension of GraphSpace [5], an

online collaboration space where biologists can share network data. As of this writing, 100+ GraphSpace users have uploaded more than 20,000 graphs containing a total of over 1.4 million nodes and 3.8 million edges, across 40+ research projects. The system was implemented in Python and uses the Django framework, a PostgreSQL database, and Cytoscape.js [19], an open-source network visualization library written in JavaScript. GraphSpace allows users to upload and download network data in JSON formats, visualize and interact with networks, share them with a group or the public, search for networks by title or tag, and search within graphs by node or edge label. It also implements a RESTful API. CrowdLayout and GraphSpace are both available as free, open-source software.<sup>1</sup>

CrowdLayout extends GraphSpace by allowing users to crowdsource layouts of their networks. After uploading a network, a biologist views the network and sees three buttons under the Layout panel: Manual, Auto, or Crowd. The Manual button provides access to saved layouts created manually by the biologist. Alternatively, the biologist can click the Auto button to choose from five standard layout algorithms implemented by Cytoscape.js: circle, concentric, grid, spring embedder, and tree (breadth first). The biologist can instead click the Crowd button to launch five *design tasks* to the Amazon Mechanical Turk (MTurk) crowdsourcing platform. A unique crowd worker completes each task by designing a layout, and the system sends the result back to the biologist via the API. The Layout panel shows each crowd layout in a timestamped list, and the biologist can easily view each one by clicking it.

When a crowd worker completes a layout, CrowdLayout launches five new *review tasks* to assess its quality. Five new workers provide reviews of the layout, and the reviews are stored in CrowdLayout’s database via the API. To maximize scalability, we do not train workers, recruit for or require biology expertise, or filter by other qualifications (e.g. location, HIT approval) for any CrowdLayout tasks.

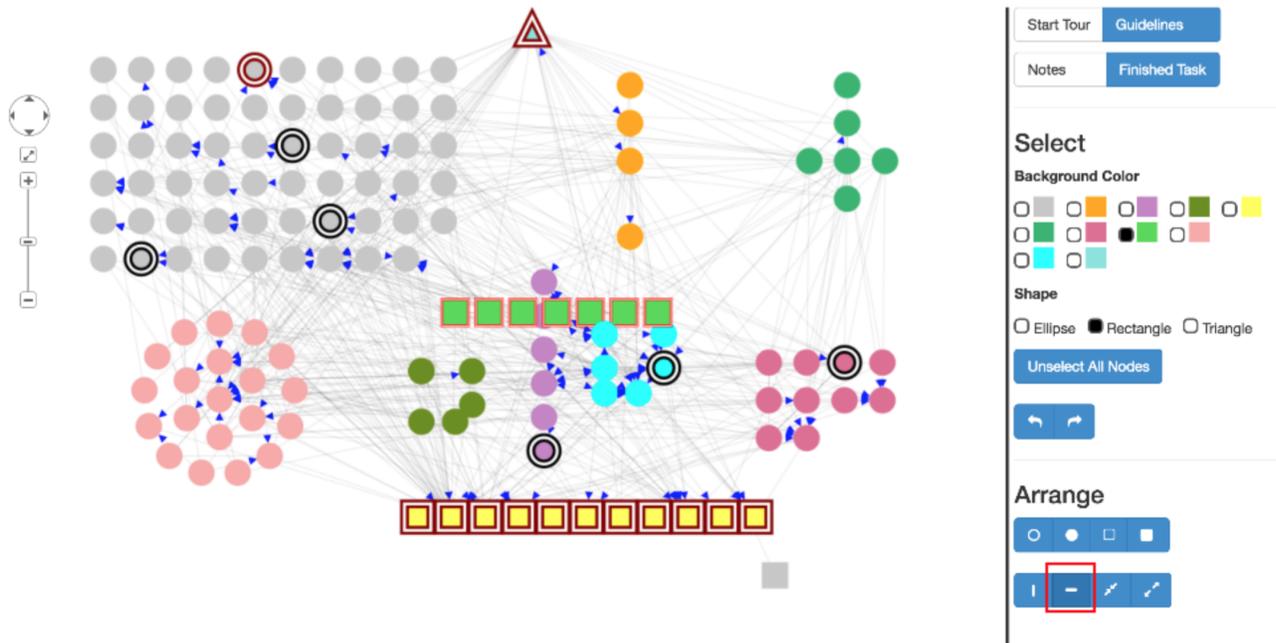
### Layout Editing Interface

Either the biologist or crowd workers can use CrowdLayout’s *layout editing interface* to design network layouts. The editing interface (Figure 2) consists of a workspace with the visualization on the left, and a sidebar of editing tools on the right.

The workspace displays the network visualization, along with controls for zooming and panning. By default, the workspace streamlines the visualization and removes unnecessary details to help the user focus on the layout task, including removing node and edge labels and deemphasizing edges by using a light gray color. Biologists in Manual mode can disable these simplifications. Node types are differentiated by shapes (circle, rectangle, triangle) and color (many possibilities). CrowdLayout extracts these properties automatically from the network’s JSON. Users can click on one or more nodes to select them, and drag to reposition the selected nodes elsewhere in the workspace.

The sidebar provides more powerful editing tools beyond click-and-drag. Users can select all nodes in the workspace by color and by shape. The selection criteria are presented as

<sup>1</sup><https://github.com/Murali-group/GraphSpace>



**Figure 2.** The CrowdLayout layout editing interface. In the workspace on the left, the user can click-and-drag to move nodes around, and pan and zoom around the workspace using the controls on the far left. In the sidebar on the right, the user can take a quick tour of the interface, view the design guidelines, and provide notes on this layout. The user can also use the selection tools to select groups of nodes by shape or color, and the arrangement tools to arrange the nodes in one of several patterns, such as circles, squares, and lines (indicated by the red box).

checkboxes, so combinations are supported. Users can also click a button to automatically arrange selected nodes in one of six patterns: circle, filled circle, rectangle, filled rectangle, horizontal line, and vertical line. Additional buttons allow the user to expand or contract the spacing in between nodes for any of these arrangements. Because a single button click can substantially impact the layout, Undo and Redo buttons are also available.

Users can click a Start Tour button to participate in a brief walkthrough of the editing interface (activated by default for crowd workers). A Guidelines button displays design guidelines for effective layouts. A Notes button allows users to provide text notes — for example, issues a crowd worker wants to communicate to the biologist — and a Finished Task button saves the new layout and exits the editing interface.

### Design guidelines

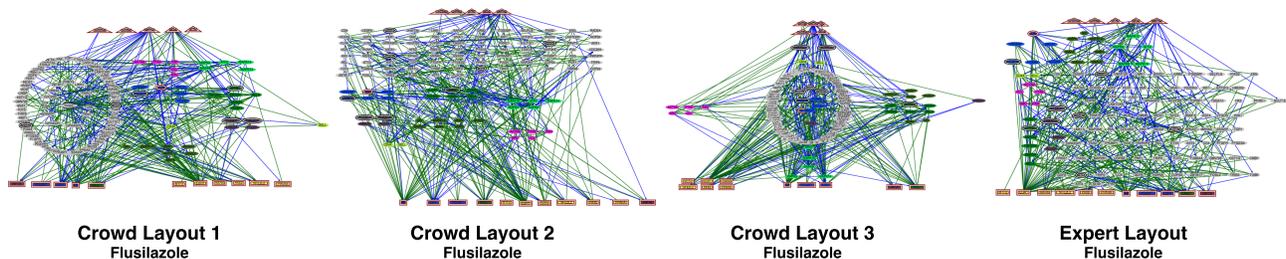
CrowdLayout provides a list of *design guidelines* to help users create effective biological network layouts. These guidelines are primarily intended for crowd workers, who are assumed to be novices with respect to biology and cell signaling pathways. However, they may also provide valuable reminders to experienced biologists. Expert 1, a tenured professor of computational biology with 15+ years of experience creating layouts of cell signaling pathways and an author of this paper, led the development of these guidelines. We iterated extensively in pilot studies with crowd workers to strike a balance between simplicity (avoid complex or numerous instructions)

and effectiveness (make the layout better). The final list has four prioritized guidelines:

1. Arrange triangles at the top of the graph
2. Arrange rectangles at the bottom of the graph
3. Arrange nodes of the same shape together
4. Arrange nodes of the same color together

These guidelines capture the intrinsic biological meaning of the networks. Triangles are receptor proteins that are located on the cell membrane, whose function is to sense the presence of the specific signals in the cell's environment. When a receptor senses such a signal, it communicates the information to the interior of the cell via interactions with other nodes in the network; their shapes and colors indicate their specific roles. Ultimately, the signal reaches proteins called transcription factors (rectangles), which move to the cell's nucleus to carry out its response to the signal. These guidelines seek to convey the flow of information from the cell membrane to the nucleus.

The guidelines become more interdependent from top to bottom, requiring users to leverage their creativity to optimize across the guidelines while making trade-offs within the design space. Therefore, multiple users may generate substantially different, yet equally effective, layouts for the same network data. CrowdLayout generates five crowd layouts to leverage the benefits of parallel prototyping, allowing the biologist to consider multiple alternatives (Figure 3).



**Figure 3.** Example layouts for the Flusilazole network illustrate the large design space of possible solutions. Different crowd workers created layouts 1–3 using CrowdLayout, with a judge’s overall quality rating of 4, 4, and 3, respectively. The expert computational biologist created the fourth layout, with a quality score of 5.

### Layout Reviewing Interface

CrowdLayout also provides a *layout reviewing interface* for users to review the quality of a network layout. The reviewing interface resembles the editing interface in that both include the workspace and simplified network visualization on the left. On the right, the sidebar in the reviewing interface contains reviewing criteria, rather than editing tools.

Users are asked to provide a discrete 1–5 (poor–perfect) rating for each criterion by moving a discrete slider. In the standard version of this interface, each criterion is presented as a question that directly maps onto the design four guidelines:

1. How well are the triangles arranged at the top of the graph?
2. How well are the rectangles arranged at the bottom of the graph?
3. How well are the nodes with the same shape arranged together?
4. How well are the nodes with the same color arranged together?

Once the user provides ratings for all criteria, he or she clicks the Submit Rating button to complete the rating task and exit the reviewing interface.

An alternative version of the reviewing interface, modified for an expert judge as detailed in the next section, first asks the judge to rate the overall quality of the layout on a 1–5 scale. After rating the overall quality of all layouts in the set, the judge then rates each layout again using the four guidelines. The purpose of separating these judging rounds is to first capture the judge’s unbiased perceptions of overall layout quality before presenting our guidelines.

### Usage Scenarios

CrowdLayout can benefit biologists in at least two real-world scenarios. First, many biologists want to efficiently examine hundreds or thousands of graphs, but it’s impractical for them to manually lay out each one. CrowdLayout can provide reasonably good layouts for all these graphs, in a few minutes, without biologist intervention. Second, CrowdLayout can substantially reduce the time biologists need to make a high-quality layout, e.g. for a publication. The biologist could use CrowdLayout to quickly create a good first draft of the layout, and then edit it to maximize quality. In the following sections, we present two evaluations of CrowdLayout demonstrating the viability of these scenarios.

### EXPERIMENT 1: CROWDSOURCING LAYOUT DESIGNS

For our first experiment, we investigated the following key research question with respect to crowdsourced layout designs:

- RQ1: Can crowds create layouts that are as effective as expert layouts?

RQ1 considers whether the crowdsourced design technique provided by CrowdLayout offers a viable solution to generating high-quality biological network layouts by comparing them to expert layouts. We hypothesize that the design guidelines, abstractions, and editing tools we provide will narrow the gap between novice crowds and experts, allowing them to produce similarly high-quality layouts.

#### Methods

##### *Generating layouts from algorithms, crowds, and an expert*

From our dataset of thousands of biological signaling networks, we selected six networks of varying complexity: Bisphenol, Etoazole, Fenbuconazole, Fludioxonil, Flusilazole, and Triclosan. These molecules are either industrial chemicals or drugs. Each network captures the proteins in the cell that are affected by the chemical. The networks were computed using a path-based algorithm. Our six networks had an average of 107 nodes (min: 36, max: 210) and 265 edges (min: 65, max: 543), typical of many of the 20,000 networks across 40+ research projects shared on GraphSpace. More complex networks could be divided and laid out as subgraphs using an approach similar to [60].

For each network, we generated automatic layouts to serve as a baseline using two complementary approaches. The first method was Graphviz, a popular, open-source graph drawing package developed by AT&T Labs, to generate an automatic layout to serve as a baseline. Specifically, we used the Graphviz *dot* algorithm [18] since it is designed to lay out a graph hierarchically in a level-wise manner. Thus, we expect nodes with no incoming edges (triangles, receptors) to appear at the top and the nodes with no outgoing edges (rectangles, transcription factors) to appear at the bottom of the layout (see Figure 1a).

Algorithms such as *dot* do not use any biological information. Therefore, we also used a second method, Cerebral [2], which considers information on the cellular locations of proteins to lay out networks. Specifically, Cerebral stratifies nodes by their locations. Thus, it is likely to place receptors in the

“plasma membrane” and transcription factors in the “nucleus” with other proteins in the network occupying these and other locations depending on their annotations. The ordering of these regions in the layout (Figure 1b) mimics the way biologists lay out signaling pathways in textbooks and publications.

Cerebral is widely used and endorsed by biologists. Gehlenborg et al. [21] reviewed 27 tools for visualizing protein interaction networks and recommended just two: Cerebral and Cytoscape. High-impact biology journals regularly publish Cerebral visualizations (e.g. [41, 28, 39, 50]). Our study experts also suggested Cerebral as an algorithmic baseline. For CrowdLayout, we used CerebralWeb [20], a JavaScript implementation of Cerebral.

We recruited crowd workers from MTurk to manually improve randomized network layouts using the CrowdLayout editing interface. We paid workers \$0.50 per layout task, a rate based on minimum hourly wage in our area adjusted for a 5-minute task. For a gold standard, Expert 1 also used CrowdLayout to create layouts.

To prepare the task for manual layouts, we generated six layouts with randomized node positions for each of the six networks. We then randomly assigned five layouts to crowd workers (one unique worker per layout) and the remaining one to the expert. We required crowd workers to spend at least 10 seconds on the layout task and move at least one node. We asked the expert to create a layout that they would be comfortable including in a publication or conference presentation, to simulate a reasonable gold standard.

In total, there were 48 layouts in the experiment. For each of the six networks, there were eight layouts: five crowd-improved layouts (one per worker), one expert-improved layout, and two algorithm-generated layouts (one Graphviz *dot*, one Cerebral).

#### *Evaluating the layouts*

We recruited Expert 2, an assistant professor of biology who does signaling pathways research and was not involved in CrowdLayout’s development, to evaluate all 48 layouts in two separate rounds. For the first round, this judge (Expert 2) used the modified CrowdLayout reviewing interface described above to provide an overall quality rating for each layout. For the second round, the judge rated each layout again, this time using our four design guidelines. For both rounds of judging, CrowdLayout presented the layouts in a random sequence and the judge was blind to condition; i.e. the reviewing interface did not reveal to the judge whether the layout was created by a crowd worker, an algorithm, or the expert.

We analyzed the judge’s ratings using Kruskal–Wallis tests. We used these non-parametric alternatives to ANOVAs because Shapiro–Wilk tests showed the ratings data failed a normality assumption ( $p < 0.05$  for all five rating types). Each test compared crowd, expert, and algorithm layouts across all six networks for one of the five rating types (triangle, rectangle, shape, color, and overall quality). We followed up significant Kruskal–Wallis tests with Mann–Whitney  $U$ -tests (a non-parametric alternative to  $t$ -tests) and Bonferroni corrections to adjust significance levels for multiple comparisons

( $\alpha = 0.05/3 = 0.01\bar{6}$ ). We also report effect sizes ( $r$ ) for these with the interpretation that  $r=0.1$  is a small effect size,  $r=0.3$  is medium, and  $r=0.5+$  is large. We used R to perform all statistical analyses.

To validate Expert 2’s layout reviews, Expert 1 followed the same review process described above and reviewed all layouts except the six he created himself (42 total). To measure inter-expert agreement, we calculated the weighted Cohen’s  $\kappa$  for the four guidelines and overall quality. The  $\kappa$  scores were: 0.47 (triangles), 0.75 (rectangles), 0.80 (shapes), 0.76 (colors), and 0.79 (overall). Thus, both experts showed substantial agreement in reviewing for overall quality and all guidelines except triangles. We conducted a follow-up analysis to understand why agreement for the triangles guideline was lower, and found that Expert 1 was more forgiving of common layout algorithm errors, such as flipping the relative positions of triangles and rectangles.

To simplify the presentation and discussion of results below, we combined the two layout algorithms, Graphviz *dot* and Cerebral, under the common heading of “algorithm.” We confirmed with Kruskal–Wallis tests that the two algorithms were not significantly different for any of the rating types.

## **Results**

### *G1: Crowd layouts as good as expert, better than algorithm*

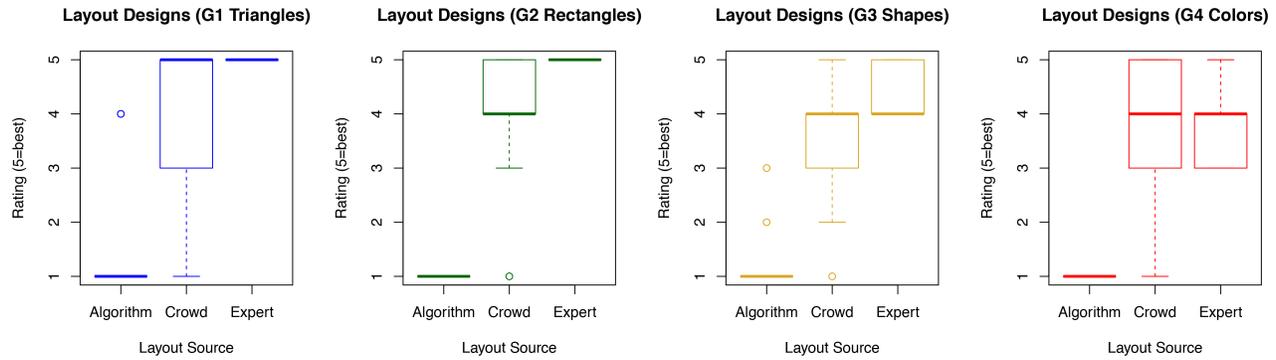
The box plot for Guideline 1 (triangles) is shown in Figure 4 (blue border). The expert had the highest average rating ( $\bar{x}=5$  out of 5,  $\sigma=0.0$ ), compared to the crowd ( $\bar{x}=4.27$ ,  $\sigma=1.08$ ), and the algorithm ( $\bar{x}=1.50$ ,  $\sigma=1.17$ ). A Kruskal–Wallis test showed a significant effect of layout creator on G1 rating ( $\chi^2(3)=30.98$ ,  $p < 0.01$ ). Post-hoc Mann–Whitney  $U$ -tests showed significant differences between crowd and algorithm ( $p < 0.001$ ,  $r=0.68$ ) and between expert and algorithm ( $p < 0.001$ ,  $r=0.84$ ), but not between crowd and expert ( $p=n.s.$ ).

### *G2: Crowd layouts as good as expert, better than algorithm*

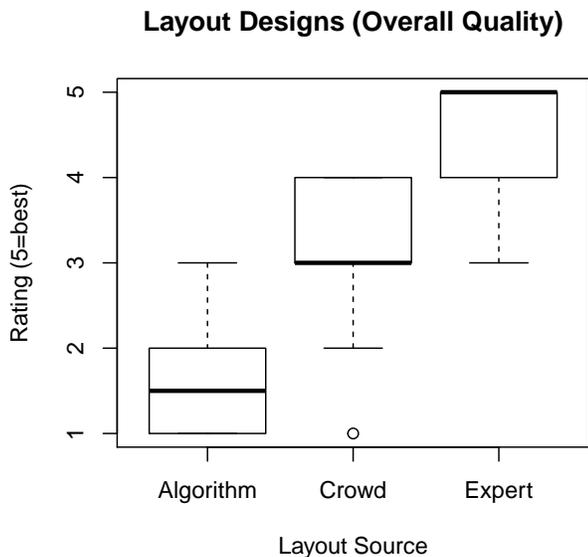
The box plot for Guideline 2 (rectangles) is shown in Figure 4 (green border). The judge gave the expert the highest ratings ( $\bar{x}=5$ ,  $\sigma=0.0$ ), followed by the crowd ( $\bar{x}=4.17$ ,  $\sigma=0.99$ ), and then the algorithm ( $\bar{x}=1.0$ ,  $\sigma=0.0$ ). A Kruskal–Wallis test showed a significant effect of layout creator on G2 rating ( $\chi^2(3)=34.79$ ,  $p < 0.001$ ). Post-hoc Mann–Whitney  $U$ -tests showed significant differences between crowd and algorithm ( $p < 0.001$ ,  $r=0.76$ ) and between expert and algorithm ( $p < 0.001$ ,  $r=0.92$ ), but not between crowd and expert ( $p=n.s.$ ).

### *G3: Crowd layouts as good as expert, better than algorithm*

The box plot for Guideline 3 (shapes) is shown in Figure 4 (yellow border). The expert had the highest average rating ( $\bar{x}=4.33$ ,  $\sigma=0.52$ ), compared to the crowd ( $\bar{x}=3.63$ ,  $\sigma=1.04$ ), and the algorithm ( $\bar{x}=1.25$ ,  $\sigma=0.62$ ). A Kruskal–Wallis test showed a significant effect of layout creator on G3 rating ( $\chi^2(4)=28.98$ ,  $p < 0.001$ ). Post-hoc Mann–Whitney  $U$ -tests showed significant differences between crowd and algorithm ( $p < 0.001$ ,  $r=0.70$ ) and between expert and algorithm ( $p < 0.001$ ,  $r=0.82$ ), but not between crowd and expert ( $p=n.s.$ ).



**Figure 4.** Box plots comparing judge's ratings for algorithm-, crowd-, and expert-generated layouts for the four guidelines in Experiment 1. Crowds performed as well as the expert ( $p=n.s.$  for all), and both performed significantly better than the algorithm ( $p<0.001$  and  $r>0.50$  for all).



**Figure 5.** Box plot comparing judge's ratings for algorithm-, crowd-, and expert-generated layouts for overall quality in Experiment 1. Crowds performed significantly better than algorithm ( $p<0.001$ ,  $r=0.62$ ), but not as well as the expert ( $p<0.01$ ,  $r=0.45$ ).

**G4: Crowd layouts as good as expert, better than algorithm**  
 The box plot for Guideline 4 (colors) is shown in Figure 4 (red border). The judge gave the expert the highest ratings ( $\bar{x}=3.83$ ,  $\sigma=0.75$ ), followed by the crowd ( $\bar{x}=3.37$ ,  $\sigma=1.47$ ), and then the algorithm ( $\bar{x}=1.0$ ,  $\sigma=0.0$ ). A Kruskal–Wallis test showed a significant effect of layout creator on G4 rating ( $\chi^2(3)=23.15$ ,  $p<0.001$ ). Post-hoc Mann–Whitney  $U$ -tests showed significant differences between crowd and algorithm ( $p<0.001$ ,  $r=0.60$ ) and between expert and algorithm ( $p<0.001$ ,  $r=0.90$ ), but not between crowd and expert ( $p=n.s.$ ).

**Quality: Crowd higher than algorithm, lower than expert**  
 The box plot for overall quality is shown in Figure 5. The expert had the highest average rating ( $\bar{x}=4.50$ ,  $\sigma=0.84$ ), com-

pared to the crowd ( $\bar{x}=3.17$ ,  $\sigma=0.92$ ), and the algorithm ( $\bar{x}=1.58$ ,  $\sigma=0.67$ ). A Kruskal–Wallis test showed a significant effect of layout creator on overall quality rating ( $\chi^2(4)=28.72$ ,  $p<0.001$ ). Post-hoc Mann–Whitney  $U$ -tests showed significant differences between crowd and algorithm ( $p<0.001$ ,  $r=0.62$ ) between expert and algorithm ( $p<0.001$ ,  $r=0.75$ ), and between crowd and expert ( $p<0.01$ ,  $r=0.45$ ).

**Time: Crowd faster than expert, slower than algorithm**  
 The expert spent more time on each layout ( $\bar{x}=14$  min) than crowd workers ( $\bar{x}=5$ ). The time required to generate a Graphviz *dot* or Cerebral layout was trivial.

## Discussion

Experiment 1 considered how effectively CrowdLayout could enable novice crowds to design biological network layouts. RQ1 asked whether crowds could generate layouts as effective as those of experts, and we hypothesized that crowd layouts would be similarly high-quality. Our results partially support this hypothesis. The results show that for our four layout guidelines, crowd workers performed as well as the expert's gold standard, and significantly better than the Graphviz *dot* and Cerebral baselines, with a large effect size. Furthermore, the average ratings for both crowds and experts across all four guidelines was high, ranging from 4.17–5 for G1 and G2, and 3.37–4.33 for G3 and G4. In contrast, the algorithm-generated layouts performed poorly across all four guidelines, averaging between 1.0 and 1.5.

We observed a slight downward trend in average ratings for both crowds and the expert across the four guidelines. This trend seems to reflect the challenge of optimizing for increasingly interdependent guidelines. As CrowdLayout users move through the list, making changes to the layout, they encounter more conflicts between guidelines that require them to make trade-offs. While G1 and G2 focus on node position relative to the workspace, G3 and G4 focus on node positions relative to one another — a more interdependent task.

We also measured the overall quality of layouts independent of our guidelines. With this complementary measure of effectiveness, we aimed to strengthen the ecological validity of the study, not only by providing an independent perspective

on the effectiveness of each layout, but also by enabling comparisons to our guidelines-based measure. We found that all three layout sources were significantly different with respect to overall quality. The layouts created by the expert scored consistently high (4.5 on average). Quality ratings for crowd layouts were lower than expert layouts, but still reasonably good — an average of 3.17, with 75% scoring 3 or higher. In contrast, both crowds and the expert performed significantly better than algorithms, with an average of only 1.58.

Taken together, the two measures of effectiveness suggest the following. First, the algorithm-generated layouts scored low on each layout guideline and overall quality, and are significantly worse than either crowds or the expert. Even Cerebral, which incorporates biological knowledge to visualize networks, did not provide effective layouts in this context. This supports our positioning of algorithmically generated layouts as a baseline.

Second, the expert performed highly across both measures. This provides a helpful reality check that experts can create as well as recognize high-quality layouts on their own terms, and are an appropriate gold standard here.

Finally, crowds performed as well as the expert on all the layout guidelines, but not as well as the expert in terms of overall quality. This suggests that the guidelines we developed may not fully capture the elements of a highly effective layout that an expert, with domain knowledge and years of experience, brings to the task. For example, experts could use their knowledge of specific protein combinations to create more meaningful sub-layouts. Additional guidelines may help close this gap, but some tacit knowledge, such as aesthetic considerations, may be impossible to distill.

Nevertheless, our results suggest that our guidelines provide a useful approximation, enabling even novices to create reasonably good layouts. When we factor in elapsed time for each source, we see that CrowdLayout’s novice crowds offer a scalable alternative to expert-generated layouts, creating layouts that are significantly better than current algorithms and satisfy key criteria as well as an expert, in a third of the time, on average.

This experiment demonstrated that crowd workers can generate high-quality layouts, but knowing *which* layouts are high quality is an equally important goal, and one that is currently difficult to achieve through automated techniques. We recruited an expert to judge these layouts, but this solution does not scale well, and with CrowdLayout we seek to ease the burden on biologists. Therefore, we conducted a second experiment to determine if crowds could review layouts as effectively as they create them.

## EXPERIMENT 2: CROWDSOURCING LAYOUT REVIEWS

For our second experiment, we investigated two key research questions with respect to crowdsourced layout reviews:

- RQ2: How do crowd reviews of layouts differ from expert reviews?
- RQ3: How much do crowd workers agree with one another on layout reviews?

RQ2 addresses the possibility of whether crowds provide reviews that are sufficiently consistent with experts that crowds could be used to both create and assess the quality of layouts. We hypothesize that the crowd’s reviews of layouts will be similar to experts.

RQ3 considers agreement among crowd workers to shed light on whether crowd reviews are most useful in aggregate, or individual workers perform reliably enough to provide useful reviews on their own. We hypothesize that crowd workers will generally agree with one another on layout reviews, because they are drawing on similar (novice) backgrounds and are exposed to the same evaluation criteria.

## Methods

For our dataset, we again used the six biological networks from Experiment 1. To ensure a diverse range of layouts, we collected 11 unique layouts for each network — five from paid crowd workers, five from computational biology students at our institution, and one from the Graphviz *dot* algorithm. In total, there were 66 layouts.

We sought to recruit five crowd workers from MTurk to review each of the 66 layouts. Due to automatic task balancing, we over-collected a sixth crowd review for six layouts and under-collected one review for a student-created Bisphenol layout. Therefore, the total number of crowd tasks (and unique workers) was 335, not 330 (66×5).

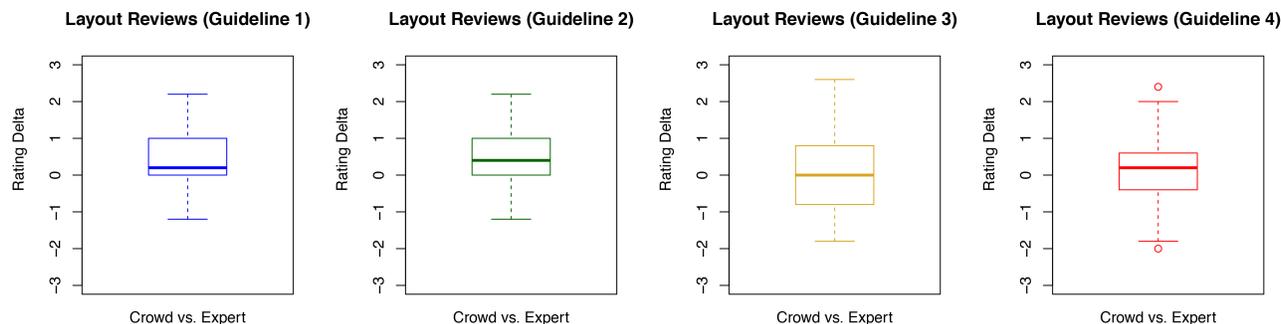
We paid each worker \$0.35 to review one of the randomly assigned 66 layouts using the CrowdLayout reviewing interface. Because crowd workers lack domain expertise, we focused on their reviews of our guidelines and did not collect overall quality ratings. Each worker reviewed a layout by moving a discrete 1–5 (perfect) ratings slider for each of the four guidelines. We required workers to spend at least three seconds on the review task and move a ratings slider in order for the work to be accepted. Most workers spent 2–3 minutes on their review tasks.

For a gold standard, Expert 2 used CrowdLayout to provide new reviews for all 66 layouts. We used the Wilcoxon signed-rank test to measure agreement between the crowd and the expert. This non-parametric test for matched pairs allows us to compare the average crowd reviews to the expert reviews, per guideline, while controlling for differences in quality across the 66 layouts.

We also considered the extent to which multiple crowd workers agreed with one another’s review for the same layout(s). To measure within-crowd agreement per layout, we used Fleiss’  $\kappa$ , a variant of Cohen’s  $\kappa$  used for more than two raters.

## Results

*Crowd and expert reviews average  $\geq 90\%$  similar per layout*  
We found that on average, the expert rated layouts slightly higher than the crowd across all guidelines. Wilcoxon signed-rank tests showed that for G1, the average  $\Delta$  (expert rating minus average crowd rating) per layout was 0.44 ( $\sigma=0.77$ ), and this 9% difference was significant,  $p<0.001$ . For G2, the average  $\Delta$  was 0.51 ( $\sigma=0.73$ ), and this 10% difference was also significant,  $p<0.001$ . For G3, the average  $\Delta$  was



**Figure 6. Boxplots comparing layout reviews provided by crowds vs. the expert for each of the four guidelines in Experiment 2. Overall average  $\Delta$  (expert minus crowd) ranged from 0.10 to 0.51 (2% to 10%) on the 5-point quality scale. The differences were significant for G1 and G2 ( $p < 0.001$ ), but not G3 or G4.**

0.10 ( $\sigma=1.09$ ), but the difference was not significant. For G4, the average  $\Delta$  was 0.18 ( $\sigma=0.86$ ), but this difference was not significant either. The boxplots are shown in Figure 6.

#### *Within-crowd agreement mostly fair*

This analysis considers how much the five workers who rated each layout agreed with one another. We used Fleiss'  $\kappa$  to measure the crowd's agreement reliability, with the following commonly accepted thresholds for  $\kappa$ :  $\leq 0.2$ =slight agreement; 0.21–0.4=fair agreement, 0.41–0.6=moderate agreement, 0.61–0.8=substantial agreement,  $\geq 0.8$ =almost perfect agreement.

For Guidelines 1, 2, and 4, the Fleiss'  $\kappa$  scores were 0.40 ( $z=16.1$ ,  $p < 0.001$ ), 0.31 ( $z=14.1$ ,  $p < 0.001$ ), and 0.25 ( $z=12.1$ ,  $p < 0.001$ ), respectively, indicating fair agreement for all three. Guideline 3 had the lowest  $\kappa$ , 0.13 ( $z=6.6$ ,  $p < 0.001$ ), indicating slight agreement. A category-wise breakdown of the  $\kappa$  scores showed that, within each guideline, workers had relatively high agreement about extreme ratings of 1 and 5, and much less agreement about intermediate ratings of 2–4.

## Discussion

Experiment 2 considered how effectively crowds could use CrowdLayout to review the quality of biological network layouts. RQ2 addressed how crowd reviews compared to experts, and we hypothesized they would be similar. Our results partially support this hypothesis. Our analysis of expert vs. crowd reviews per layout found relatively small differences. For Guidelines 1 and 2, the differences averaged 8–10% on the five-point quality scale, and these were statistically significant ( $p < 0.001$  for both). For Guidelines 3 and 4, the differences averaged 2–4% and were not significant ( $p = \text{n.s.}$  for both). Thus, for a given layout, average crowd reviews were indistinguishable from the expert (G3 and G4) or 90–92% similar to the expert (G1 and G2).

How and why were crowd reviews different from experts for the first two guidelines? We found that crowds review layouts more harshly than our expert did across all four guidelines; for G1 and G2, the difference was significant. A possible explanation is that our expert is simply more forgiving than the crowd, perhaps drawing on past experience with what constitutes an effective biological network layout. A different expert might be slightly more or less strict with their reviews.

These results are promising, because currently there is no effective automated technique for identifying high quality layouts, and expert reviews scale poorly and increase the burden on biologists. If crowds can review layouts as effectively as experts, then they can potentially replace experts, providing a scalable solution to the problem.

High-quality crowd reviews close the loop by providing a reliable mechanism for assessing the quality of crowd-generated layouts. Without biologist intervention, crowds can assess when enough high-quality layouts have been generated and provide a stopping rule for an iterative workflow (e.g., when a layout has received high reviews, stop generating crowd layouts). Crowds can also flag poorly reviewed layouts for improvement, feedback, or penalization.

RQ3 considered the agreement of crowd reviews for a given layout, and we hypothesized that agreement would be high. While the average of five crowd workers' reviews for a layout was similar to an expert's, the agreement among those crowd workers ranged from fair to slight. One explanation for this mismatch is methodological. Although the 1–5 ratings are ordinal, Fleiss'  $\kappa$  does not allow for weighted categories, so a disagreement between workers of one point (e.g., 3 vs. 4) is penalized as much as a disagreement of four points (e.g., 1 vs. 5). Therefore, workers might actually agree more (or less) than the  $\kappa$  scores suggest.

## BROADER IMPLICATIONS

### Bootstrapping from Design Guidelines to Review Criteria

In developing CrowdLayout, we first aimed to create a set of guidelines that would allow non-expert crowds to design biological network layouts that would be meaningful to scientists. In seeking to measure the quality of those layouts in early pilots, however, we soon realized that the same attributes of biological visualizations that make them challenging to automatically lay out (complexity, size, fidelity to real-world structures, aesthetics) also made them challenging to evaluate. This problem recalls the broader challenge of evaluating complex and creative crowd work that has been well-documented in other domains [32, 61].

Our solution was to attempt to bootstrap our design guidelines into review criteria with minimal changes, reasoning that

the collaborative, iterative process that yielded the guidelines would capture many of the important criteria for evaluation. This solution appealed to us not only because of its elegance — a single list of statements could fulfill both needs — but also its fundamental fairness in evaluating crowd workers' performance against the same goals they were provided. This approach extends earlier work showing that self-assessments help crowd workers learn to perform better [14], and canned instructions derived from earlier performance can be reused to provide feedback [34, 15].

While this bootstrapping approach offers advantages in efficiency and symmetry, it can also pose a threat to validity if not approached carefully. In some domains, instructions for performing a task may not be appropriate for assessing performance, and system designers must consider whether other, validated quality measures already exist. Such measures did not exist for biological network layouts, so we developed them in an effort led by an experienced computational biologist, Expert 1. We also recruited a second computational biologist, Expert 2, whose independent reviews of overall layout quality supported our guidelines. Expert 2 told us:

*I develop algorithms for pathway analysis that generate hundreds of biological networks. CrowdLayout enables me to sift through these networks by looking at layouts determined by non-experts using a set of straightforward rules. I can glean biological insight about pathway topology by looking at these layouts, saving me an immense amount of time.*

These efforts represent our attempts at reasonable and pragmatic measures of effectiveness, but other equally valid or superior measures are certainly possible.

### Exploring Trade-offs and Synthesizing Alternatives

Network layout is a constraint satisfaction problem and therefore also a design problem. In biological contexts like cell signaling pathways, an effective layout must balance clarity (drawing on principles of aesthetics and design) and accuracy (correctly reflecting real-world cell structures). By bringing crowd workers to this design space, we sought to leverage human intelligence and creativity to explore this design space and effectively balance these trade-offs. By default, CrowdLayout launches five design tasks for a network. Our original goal with this feature was primarily to increase the odds that at least one layout would be acceptable for the biologist to use or improve upon. However, as our initial evaluations showed that crowd-designed layouts were generally high-quality, we observed a second, potentially greater benefit: multiple workers produced different, but often equally effective, layouts for the same network (Figure 3). This result provided the biologist with a portfolio of reasonable layout alternatives, but CrowdLayout does not yet provide any targeted support for helping them select among them. Parallel prototyping has been shown to lead to more creative and effective designs, but this outcome relies on the designer's ability to consider alternatives and combine the best ideas from each into a new design that is stronger than the sum of its parts [15, 16].

Recent studies of crowdsourced creativity and innovation [7, 33, 30, 57, 58] offer insights into how a system like CrowdLayout could leverage the crowd's diversity of ideas by providing mechanisms for exploring design trade-offs in their layouts and supporting the synthesis of multiple good ideas into a coherent whole. Adapting these approaches, such as libraries of examples or expert facilitation, for biological network layout is a nontrivial but important goal for future work.

### Crowdsourcing Network Layouts Beyond Biology

This paper focuses on the challenge of creating effective visualizations in the biological domain of cell signaling pathways. However, this context represents only one type of network data within biology; Expert 2 suggested CrowdLayout could be helpful for other types such as gene regulatory networks, cancer networks, and protein co-complex networks. Outside of biology, the possibilities are wide-ranging. Networks are powerful and pervasive as a data structure partly because of their flexibility to represent such a broad range of phenomena, from social networks to citation networks [4, 9]. Likewise, we expect that the techniques CrowdLayout embodies can generalize beyond cell signaling pathways to other domains where the visualizing networks for human inspection is as important as generating them and analyzing them computationally. Because CrowdLayout's design and review process is organized around a short list of interchangeable guidelines, and its editing tools and representations (colors, shapes) are abstracted, the software and its underlying concepts could be adapted for other types of networks. The success of this translation depends on front-loaded effort to develop of a new set of guidelines that are usable for novice crowd workers. Once created, however, these guidelines can be widely deployed to support network visualization in the new domain at scale.

### CONCLUSION

Network data is widely used in biological research, but creating meaningful visualizations of these networks is challenging for algorithms and typically requires significant time and expertise for biologists to perform manually. We presented CrowdLayout, a system that leverages crowdsourced human intelligence and creativity to design layouts of biological network visualizations. CrowdLayout uses design guidelines, abstractions, and special editing tools to help crowd workers without biological expertise create layouts that mimic real-world biological structures. The guidelines also serve as evaluation criteria for crowds to assess the quality of their layouts. In two experiments, we showed that crowds could design layouts that satisfy these guidelines as effectively as expert biologists, and crowds could also provide reviews of layouts that were similar to an expert. Consequently, CrowdLayout enables biological network visualization at scale, giving scientists access to meaningful visualizations of many more networks than they would be able to create manually, and potentially supporting new scientific breakthroughs.

### ACKNOWLEDGEMENTS

We wish to thank Aditya Bharadwaj, Anna Ritz, the MTurk workers, and the anonymous reviewers. This research was supported by NIH grants 1UH2CA203768-01 and R01-GM095955.

## REFERENCES

1. Albert-Laszlo Barabasi, Natali Gulbahce, and Joseph Loscalzo. 2011. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics* 12, 1 (01 Jan. 2011), 56–68.
2. A. Barsky, T. Munzner, J. Gardy, and R. Kincaid. 2008. Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (Nov. 2008), 1253–1260. DOI: <http://dx.doi.org/10.1109/TVCG.2008.117>
3. Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks. *ICWSM* 8 (2009), 361–362.
4. Yochai Benkler. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.
5. Aditya Bharadwaj, Divit P. Singh, Anna Ritz, Allison N. Tegge, Christopher L. Poirer, Pavel Kraikivski, Neil Adames, Kurt Luther, Shiv D. Kale, Jean Peccoud, John J. Tyson, and T. M. Murali. 2017. GraphSpace: stimulating interdisciplinary collaborations in network biology. *Bioinformatics* 33, 19 (Oct. 2017), 3134–3136. DOI: <http://dx.doi.org/10.1093/bioinformatics/btx382>
6. Rick Bonney, Jennifer L. Shirk, Tina B. Phillips, Andrea Wiggins, Heidi L. Ballard, Abraham J. Miller-Rushing, and Julia K. Parrish. 2014. Next Steps for Citizen Science. *Science* 343, 6178 (March 2014), 1436–1437. DOI: <http://dx.doi.org/10.1126/science.1251554>
7. Joel Chan, Steven Dang, and Steven P. Dow. 2016. Comparing Different Sensemaking Approaches for Large-Scale Ideation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2717–2728. DOI: <http://dx.doi.org/10.1145/2858036.2858178>
8. Duen Horng Chau, Aniket Kittur, Jason I. Hong, and Christos Faloutsos. 2011. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 167–176. DOI: <http://dx.doi.org/10.1145/1978942.1978967>
9. Nicholas A. Christakis and James H. Fowler. 2011. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives – How Your Friends' Friends' Friends Affect Everything You Feel, Think, and Do* (reprint edition ed.). Back Bay Books.
10. Jeffrey P. Cohn. 2008. Citizen science: can volunteers do real research? *BioScience* 58, 3 (March 2008), 192–197.
11. Catalina M. Danis, Fernanda B. Viegas, Martin Wattenberg, and Jesse Kriss. 2008. Your place or mine?: visualization as a community component. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. ACM, Florence, Italy, 275–284. <http://portal.acm.org.www.library.gatech.edu:2048/citation.cfm?id=1357054.1357102&coll=portal&dl=ACM&CFID=874565&CFTOKEN=84180943>
12. Wouter De Nooy, Andrej Mrvar, and Vladimir Batagelj. 2011. *Exploratory social network analysis with Pajek*. Vol. 27. Cambridge University Press.
13. Ugur Dogrusoz, Erhan Giral, Ahmet Cetintas, Ali Civril, and Emek Demir. 2009. A layout algorithm for undirected compound graphs. *Information Sciences* 179, 7 (2009), 980 – 994.
14. Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. 2012. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 1013–1022. DOI: <http://dx.doi.org/10.1145/2145204.2145355>
15. Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2010. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-efficacy. *ACM Trans. Comput.-Hum. Interact.* 17, 4 (Dec. 2010), 18:1–18:24. DOI: <http://dx.doi.org/10.1145/1879831.1879836>
16. Steven P. Dow, Kate Heddleston, and Scott R. Klemmer. 2009. The Efficacy of Prototyping Under Time Constraints. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition (C&C'09)*. ACM, New York, NY, USA, 165–174. DOI: <http://dx.doi.org/10.1145/1640233.1640260>
17. T. Dwyer, Bongshin Lee, D. Fisher, K.I. Quinn, P. Isenberg, G. Robertson, and C. North. 2009. A Comparison of User-Generated and Automatic Graph Layouts. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (Nov. 2009), 961–968. DOI: <http://dx.doi.org/10.1109/TVCG.2009.109>
18. John Ellson, Emden Gansner, Lefteris Koutsofios, Stephen C North, and Gordon Woodhull. 2002. Graphviz—open source graph drawing tools. In *Graph Drawing*. Springer, 483–484.
19. M. Franz, C. T. Lopes, G. Huck, Y. Dong, O. Sumer, and G. D. Bader. 2015. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* 32 (Sep 2015), 309–311.
20. S. Frias, K. Bryan, F. S. Brinkman, and D. J. Lynn. 2015. CerebralWeb: a Cytoscape.js plug-in to visualize networks stratified by subcellular localization. *Database (Oxford)* 2015 (2015), bav041.
21. Nils Gehlenborg, Seán I. O'Donoghue, Nitin S. Baliga, Alexander Goesmann, Matthew A. Hibbs, Hiroaki Kitano, Oliver Kohlbacher, Heiko Neuweger, Reinhard Schneider, Dan Tenenbaum, and Anne-Claude Gavin. 2010. Visualization of omics data for systems biology. *Nature Methods* 7, 3s (March 2010), nmeth.1436. DOI: <http://dx.doi.org/10.1038/nmeth.1436>

22. A. Gerasch, D. Faber, J. Kuntzer, P. Niermann, O. Kohlbacher, H. P. Lenhof, and M. Kaufmann. 2014. BiNA: a visual analytics tool for biological network data. *PLoS ONE* 9, 2 (2014), e87397.
23. Michael D. Greenberg, Matthew W. Easterday, and Elizabeth M. Gerber. 2015. Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers. In *Proceedings of ACM Creativity & Cognition 2015*. ACM, Glasgow, Scotland.
24. Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. 2007. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, San Jose, California, USA, 1029–1038. <http://portal.acm.org/www.library.gatech.edu:2048/citation.cfm?id=1240624.1240781&coll=portal&dl=ACM&CFID=874565&CFTOKEN=84180943>
25. Zhenjun Hu, Yi-Chien Chang, Yan Wang, Chia-Ling Huang, Yang Liu, Feng Tian, Brian Granger, and Charles DeLisi. 2013. VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Research* 41, W1 (01 July 2013), W225–W231.
26. Weidong Huang. 2007. Using eye tracking to investigate graph layout effects. In *2007 6th International Asia-Pacific Symposium on Visualization, 2007. APVIS '07*. 97–100. DOI: <http://dx.doi.org/10.1109/APVIS.2007.329282>
27. Weidong Huang, Peter Eades, Seok-Hee Hong, and Chun-Cheng Lin. 2013. Improving multiple aesthetics produces better graph drawings. *Journal of Visual Languages & Computing* 24, 4 (Aug. 2013), 262–272. DOI: <http://dx.doi.org/10.1016/j.jvlc.2011.12.002>
28. Francesco S. Ielasi, Mitchel Alioscha-Perez, Dagmara Donohue, Sandra Claes, Hichem Sahli, Dominique Schols, and Ronnie G. Willaert. 2016. Lectin-Glycan Interaction Network-Based Identification of Host Receptors of Microbial Pathogenic Adhesins. *mBio* 7, 4 (Sept. 2016), e00584–16. DOI: <http://dx.doi.org/10.1128/mBio.00584-16>
29. Firas Khatib, Seth Cooper, Michael D. Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, David Baker, and Foldit Players. 2011. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108, 47 (Nov. 2011), 18949–18953. DOI: <http://dx.doi.org/10.1073/pnas.1115898108>
30. Joy Kim, Justin Cheng, and Michael S. Bernstein. 2014. Ensemble: Exploring Complementary Strengths of Leaders and Crowds in Creative Collaboration. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 745–755. DOI: <http://dx.doi.org/10.1145/2531602.2531638>
31. Sunyoung Kim, Christine Robson, Thomas Zimmerman, Jeffrey Pierce, and Eben M. Haber. 2011. Creek Watch: Pairing Usefulness and Usability for Successful Citizen Science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2125–2134. DOI: <http://dx.doi.org/10.1145/1978942.1979251>
32. Aniket Kittur, Susheel Khamkar, Paul André, and Robert Kraut. 2012. CrowdWeaver: visually managing complex crowd work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 1033–1036. DOI: <http://dx.doi.org/10.1145/2145204.2145357>
33. Chinmay Kulkarni, Steven P. Dow, and Scott R. Klemmer. 2014. Early and Repeated Exposure to Examples Improves Creative Work. In *Design Thinking Research*, Larry Leifer, Hasso Plattner, and Christoph Meinel (Eds.). Springer International Publishing, 49–62. [http://link.springer.com/chapter/10.1007/978-3-319-01303-9\\_4](http://link.springer.com/chapter/10.1007/978-3-319-01303-9_4)
34. Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulou, Justin Cheng, Daphne Koller, and Scott R. Klemmer. 2013. Peer and Self Assessment in Massive Online Classes. *ACM Trans. Comput.-Hum. Interact.* 20, 6 (Dec. 2013), 33:1–33:31. DOI: <http://dx.doi.org/10.1145/2505057>
35. Edith Law, Conner Dalton, Nick Merrill, Albert Young, and Krzysztof Z. Gajos. 2013. Curio: A Platform for Supporting Mixed-Expertise Crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*. <http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7534>
36. Chris Lintott and Jason Reed. 2013. Human Computation in Citizen Science. In *Handbook of Human Computation*, Pietro Michelucci (Ed.). Springer New York, 153–162. [http://link.springer.com/chapter/10.1007/978-1-4614-8806-4\\_14](http://link.springer.com/chapter/10.1007/978-1-4614-8806-4_14)
37. Kurt Luther, Scott Counts, Kristin B. Stecher, Aaron Hoff, and Paul Johns. 2009. Pathfinder: an online collaboration environment for citizen scientists. In *Proceedings of the 27th international conference on Human factors in computing systems*. ACM, Boston, MA, USA, 239–248. DOI: <http://dx.doi.org/10.1145/1518701.1518741>
38. Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 473–485. DOI: <http://dx.doi.org/10.1145/2675133.2675283>
39. Valeriya Malysheva, Marco Antonio Mendoza-Parra, Mohamed-Ashick M. Saleem, and Hinrich Gronemeyer. 2016. Reconstruction of gene regulatory networks reveals chromatin remodelers and key transcription factors in tumorigenesis. *Genome Medicine* 8 (May 2016), 57. DOI: <http://dx.doi.org/10.1186/s13073-016-0310-3>

40. Cathleen McGrath, Jim Blythe, and David Krackhardt. 1997. The effect of spatial arrangement on judgments and errors in interpreting graphs. *Social Networks* 19, 3 (Aug. 1997), 223–242. DOI: [http://dx.doi.org/10.1016/S0378-8733\(96\)00299-7](http://dx.doi.org/10.1016/S0378-8733(96)00299-7)
41. Marco-Antonio Mendoza-Parra, Valeriya Malysheva, Mohamed Ashick Mohamed Saleem, Michele Lieb, Aurelie Godel, and Hinrich Gronemeyer. 2016. Reconstructed cell fate regulatory programs in stem cells reveal hierarchies and key factors of neurogenesis. *Genome Research* (Sept. 2016). DOI: <http://dx.doi.org/10.1101/gr.208926.116>
42. H.C. Purchase, C. Pilcher, and B. Plimmer. 2012. Graph Drawing Aesthetics - Created by Users, Not Algorithms. *IEEE Transactions on Visualization and Computer Graphics* 18, 1 (Jan. 2012), 81–92. DOI: <http://dx.doi.org/10.1109/TVCG.2010.269>
43. Helen C. Purchase, Robert F. Cohen, and Murray James. 1996. Validating graph drawing aesthetics. In *Graph Drawing*, Franz J. Brandenburg (Ed.). Number 1027 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 435–446. <http://link.springer.com/chapter/10.1007/BFb0021827>
44. P Shannon, A Markiel, O Ozier, N S Baliga, J T Wang, D Ramage, N Amin, B Schwikowski, and T Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13, 11 (Nov. 2003), 2498–2504.
45. S. Tripathi, M. Dehmer, and F. Emmert-Streib. 2014. NetBioV: an R package for visualizing large network data in biology and medicine. *Bioinformatics* 30, 19 (Oct 2014), 2834–2836.
46. Frank van Ham and Adam Perer. 2009. "Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 953–960. DOI: <http://dx.doi.org/10.1109/TVCG.2009.108>
47. F. van Ham and B. Rogowitz. 2008. Perceptual Organization in User-Generated Graph Layouts. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (Nov. 2008), 1333–1339. DOI: <http://dx.doi.org/10.1109/TVCG.2008.155>
48. M. Vidal, M.E. Cusick, and A.L. Barabási. 2011. Interactome networks and human disease. *Cell* 144, 6 (2011), 986–998.
49. Fernanda B. Viegas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. 2007. Many Eyes: A Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1121–1128.
50. Pia Wentker, Martin Eberhardt, Florian S. Dreyer, Wilhelm Bertrams, Martina Cantone, Kathrin Griss, Bernd Schmeck, and Julio Vera. 2017. An Interactive Macrophage Signal Transduction Map Facilitates Comparative Analyses of High-Throughput Data. *The Journal of Immunology* 198, 5 (March 2017), 2191–2201. DOI: <http://dx.doi.org/10.4049/jimmunol.1502513>
51. Andrea Wiggins and Kevin Crowston. 2014. Surveying the citizen science landscape. *First Monday* 20, 1 (Dec. 2014). <http://firstmonday.org/ojs/index.php/fm/article/view/5520>
52. Andrea Wiggins and Yurong He. 2016. Community-based Data Validation Practices in Citizen Science. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1548–1559. DOI: <http://dx.doi.org/10.1145/2818048.2820063>
53. Wesley Willett, Shiry Ginosar, Avital Steinitz, Björn Hartmann, and Maneesh Agrawala. 2013. Identifying Redundancy and Exposing Provenance in Crowdsourced Data Analysis. *IEEE VIS 2013* (2013).
54. Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. 2012. Strategies for crowdsourcing social data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 227–236. DOI: <http://dx.doi.org/10.1145/2207676.2207709>
55. Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. 2011. CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3131–3140. DOI: <http://dx.doi.org/10.1145/1978942.1979407>
56. Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-experts. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1433–1444. DOI: <http://dx.doi.org/10.1145/2531602.2531604>
57. Lixiu Yu, Robert E. Kraut, and Aniket Kittur. 2016. Distributed Analogical Idea Generation with Multiple Constraints. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1236–1245. DOI: <http://dx.doi.org/10.1145/2818048.2835201>
58. Lixiu Yu and Jeffrey V. Nickerson. 2011. Cooks or cobblers?: crowd creativity through combination. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*. ACM, New York, NY, USA, 1393–1402. DOI: <http://dx.doi.org/10.1145/1978942.1979147>
59. Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Bjorn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *Proceedings of the 19th ACM Conference on*

*Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1005–1017. DOI :

<http://dx.doi.org/10.1145/2818048.2819953>

60. Xiaoru Yuan, Limei Che, Yifan Hu, and Xin Zhang. 2012. Intelligent Graph Layout Using Many Users' Input. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2699–2708. DOI :  
<http://dx.doi.org/10.1109/TVCG.2012.236>
61. Haiyi Zhu, Steven P. Dow, Robert E. Kraut, and Aniket Kittur. 2014. Reviewing Versus Doing: Learning and

Performance in Crowd Assessment. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1445–1455. DOI :

<http://dx.doi.org/10.1145/2531602.2531718>

62. L. Zhu, A. Malatras, M. Thorley, I. Aghoghogbe, A. Mer, S. Duguez, G. Butler-Browne, T. Voit, and W. Duddy. 2015. CellWhere: graphical display of interaction networks organized on subcellular localizations. *Nucleic Acids Res.* (Apr 2015).