Flud: a hybrid crowd-algorithm approach for visualizing biological networks

Aditya Bharadwaj
Virginia Tech
Blacksburg, Virginia, USA
adb@vt.edu

David Gwizdala
Bridgewater Associates
Stamford, Connecticut, USA
gwizdala@vt.edu

Yoonjin Kim
Virginia Tech
Blacksburg, Virginia, USA
ykim05@vt.edu

Kurt Luther
Virginia Tech
Arlington, Virginia, USA
kluther@vt.edu

T. M. Murali
Virginia Tech
Blacksburg, Virginia, USA
murali@vt.edu

ABSTRACT

Many fields of science require meaningful and visually appealing layouts of graphs. However, the problem remains challenging due to multiple conflicting criteria and complex domain-specific constraints. In this workshop paper, we present a gamified graph layout task where the goal of the players is to create a layout that optimises a score based on user-defined priorities. We propose a novel hybrid approach wherein non-experts and simulated annealing algorithm build on each other's progress. To facilitate this collaborative process, we have developed Flud, an online game with a purpose that leverages the combination of cognitive abilities of humans to observe patterns, and the computational accuracy of simulated annealing to draw graph layouts that can help scientists visualize and understand complex networks.

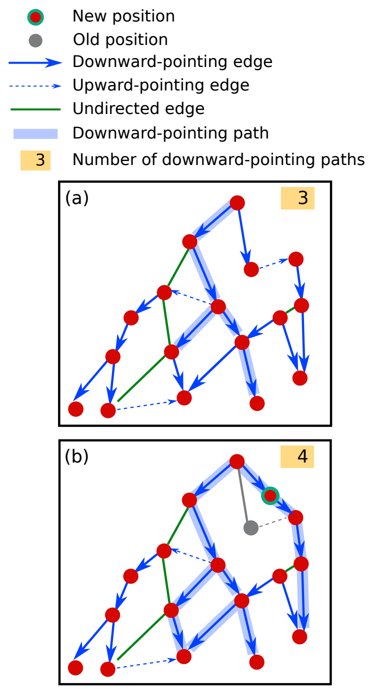


Figure 1: Downward-pointing paths. (a) Three paths. (b) Moving the node with a green border increases the number of paths to four.

KEYWORDS

crowdsourcing; graph drawing; computational biology, protein networks, graphs, citizen science, serious games, games with a purpose, optimization

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

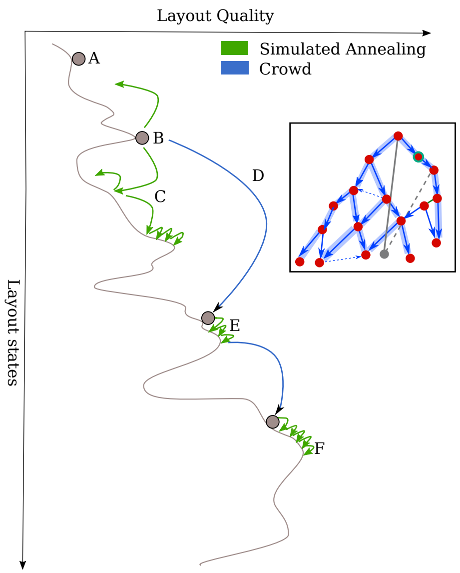


Figure 2: Illustration of hybrid crowd-algorithm collaboration facilitated by Flud. The y-axis represents the value representing layout quality as a function of layout state. The x-axis represents the layout state in 1-dimensional space for illustration purpose. (A) Initial graph layout position (B) Local Optima. Greedy algorithms can get stuck here! (C) Simulated annealing explores random non-local moves and accepts bad layouts with a small probability (D) Crowd workers creates a new downward pointing path leading to non-local move to a better layout (E) Simulated annealing fine-tunes the layout by exploring layouts in local neighborhood while preserving the overall layout structure (F) A high-quality layout after many iterations

INTRODUCTION

Modern experiments generate large quantities of data. Researchers need to be able to create visual representations of these data that are aesthetically pleasing while also conveying domain knowledge and meaning. A prominent example is the discipline of network biology where scientists use graphs to understand the chemical reactions and protein interactions that underlie processes that take place in the cell [1]. In order to present and analyze these networks, researchers require aesthetic layouts of these networks that clearly convey the relevant biological information.

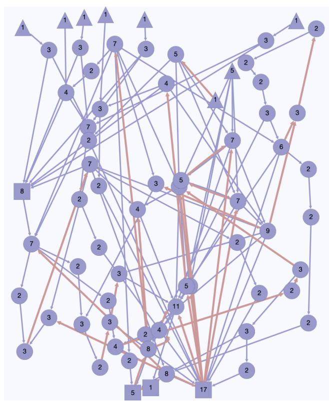
There are two major approaches for creating layouts. The first approach views humans as the deciding agent and primary creator with computers merely seen as support tools. An example is a graph drawing interface [4] that provides the layout tools for designers to use while drawing graphs. While this approach offers creative freedom, it lacks consistency and scalability. The second approach uses fully-automated methods that can generate data visualizations at scale [5]. However, these methods lack the ingenuity to capture complex visualization constraints and domain-specific needs. As a result, it is a common practice for biologists to manually improve automatically generated visualizations Interestingly, prior research on user-generated graph layouts [7, 20] suggests that non-experts can create graph layouts as well as or better than algorithms. However, due to the lack of expertise in machine learning and computer science in general, non-experts refrain from using sophisticated tools they do not fully understand. Some systems even limit their users to simple 'click and drag' interactions [6, 7].

In this work, we argue that by gamifying the visualization task and using HCI design principles we can make algorithmic approaches more readily accessible to humans with no expertise in machine learning or computer science. However, such a collaboration between non-experts and algorithms poses several open research questions. First, it remains unclear how can we elucidate a contribution made by a complex algorithm to a non-expert. Second, how can we design a human-algorithm collaborative experience that is enjoyable to the human? Third, where do we draw a line between a support tool augmenting the user's work and it removing the user's ability to make novel contributions to the system's goals?

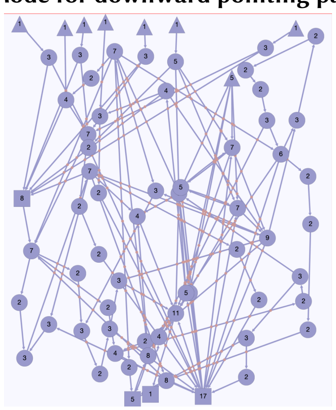
In an attempt to answer these questions in the context of graph data visualization task, we developed Flud, an online game with purpose (GWAP) that allows humans with no expertise to design graph layouts with the help of algorithmic graph layout techniques. The goal of the game is to move nodes in a given graph so as to create a layout that optimizes a score based on certain design criteria (see Table 1). Flud uses a novel sequential and collaborative process wherein humans and a simulated annealing-based layout algorithm [2] build upon each other's progress. This collaborative process leverages the combination of 1) cognitive abilities of humans to observe flow, crossings, and distances with 2) the fine-tuning capabilities of simulated annealing in order to draw graph layouts that meet specified domain-specific and aesthetic criteria.

Increase the number of downward pointing paths from sources to targets
Increase the number of non-crossing edge pairs
Decrease the distance between connected pair of nodes
Increase the distance between disconnected pair of nodes
Increase the separation between pair of nodes and edges

Table 1: Five layout criteria used in Flud



(a) Mode for downward pointing paths



(b) Mode for edge crossings

Figure 3: Screenshots of Flud when the player selects a given mode.

BIOLOGICAL GRAPH DATA VISUALIZATION

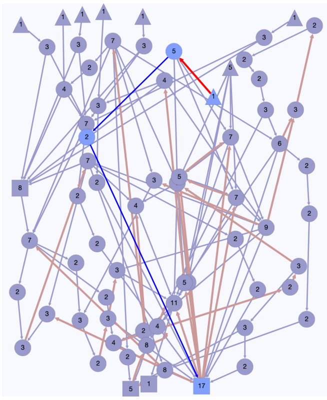
In this work, we are interested in a graph layout model where the user defines and prioritizes a set of design criteria (see Table 1) with the goal of creating a layout that optimizes a score based on these priorities. The model includes widely-used aesthetic criteria such as minimizing the number of edge crossings, keeping nodes connected by an edge close to each other, dispersing disconnected node pairs, and increasing the separation between nodes and edges. We also introduce a new criterion inspired by the biological application to cellular signaling: maximize the number of downward pointing paths in the layout (see Figure 1). These types of paths visually draw attention to sequences of edges that lead from proteins in the cell membrane through internal nodes to effector molecules in the nucleus.

HYBRID HUMAN-ALGORITHM APPROACH

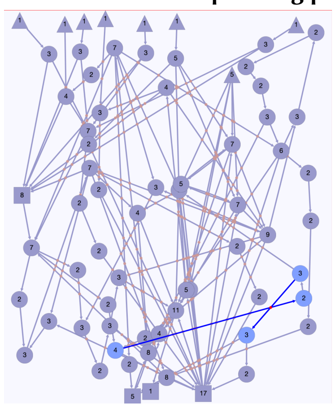
The hybrid approach is designed with the aim of utilizing the complementary strengths of humans and algorithms at the graph drawing task. In Flud, players and algorithms collaborate at different stages.

Human players. One of the challenges of graph layout task is that the aesthetic criteria may conflict with each other. Therefore, automated methods are limited to heuristic solutions and can get stuck in local optima. For instance, the correct orientation of several edges in a path may be required to make it downward-pointing. In Flud, we leverage the cognitive ability of humans to observe patterns and identify solutions to escape the local optima. However, some graphs can be very complex, and it can be challenging for players to figure out what the next move should be. To relieve the players of the cognitive load and make the interaction with the system enjoyable, we implemented two important game features—criterion-specific modes and clues (see Figures 3 and 4). In a criterion-specific mode, the visual representation of the graph highlights the elements that are relevant to the criterion-specific task. Moreover, if the player is still stuck, the player can click a "Clue" button to algorithmically highlight a small subset of nodes and edges in the graph; changing the positions of these elements is likely to improve the score for the corresponding criterion. With the help of these guidelines, clues, and layout tools, we expect game players to make modifications that may be global in nature and out of the scope of an automated method.

Algorithmic engine. We selected Simulated annealing [2] as our algorithmic approach to assist humans with the layout task due to its flexibility to accommodate all of our layout criteria. Simulated annealing uses random non-local moves to escape local optima. However, the algorithm requires a large number of iterations and random moves to escape local optima and find a better solution. In our work, we use the method developed by Harel and Sardas [3] for incremental improvement of layouts of planar graphs. This work serves as inspiration for us to use fine tuning as a way to make the random moves are "local," i.e., a node can move only to a nearby position. Such local moves further optimize the design criteria while preserving the overall structure of the layout created by the players.



(a) Clue for downward pointing paths



(b) Clue for edge crossings.

Figure 4: Screenshots of Flud when the player selects a given clue.

ACKNOWLEDGEMENTS

We thank Lee Lisle, Parker Irving, and Jeffrey Law. This research was supported by NIH grant 1UH2CA203768-01.

CONCLUSION

In this work, we are exploring the potential of human-algorithm collaboration for a design task such as graph data visualization. We describe how we gamified the task and make algorithmic approaches more accessible to humans, even if they have no biological or computer science expertise. We expect that such a collaboration between humans and algorithms will lead to higher scoring layouts than either from humans or algorithms alone.

Overall, we believe that lessons from our work can be generalized to other design tasks such as creating websites and posters or even art. These tasks are atypical of problems traditionally solved by automated intelligent systems. These systems are typically used to solve decision-making and classification problems. In contrast, design tasks where we are trying to optimize a non-linear value function, algorithms have to play a more active role. An exploration of mixed initiative systems for such tasks opens up the opportunity to understand and address open-ended problems without a definitive correct answer. This paper presents a step in this direction by leveraging the complementary strengths of both humans and algorithms to find the best possible solution for such problems.

REFERENCES

[1] A.L. Barabasi and Z.N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5, 2 (2004), 101–13.

[2] Ron Davidson and David Harel. 1996. Drawing graphs nicely using simulated annealing. *ACM Transactions on Graphics (TOG)* 15, 4 (1996), 301–331.

[3] David Harel and Meir Sardas. 1995. *An incremental drawing algorithm for planar graphs.* Technical Report. Cornell University.

[4] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 11 (2003), 2498–2504.

[5] Roberto Tamassia, Giuseppe Di Battista, and Carlo Batini. 1988. Automatic graph drawing and readability of diagrams. *IEEE Transactions on Systems, Man, and Cybernetics* 18, 1 (1988), 61–79.

[6] Frank van Ham and Bernice Rogowitz. 2008. Perceptual organization in user-generated graph layouts. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1333–1339.

[7] 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 (2012), 2699–2708.