ISSN 2186-7437

NII Shonan Meeting Report

No. 2015-1

Big Graph Drawing: Metrics and Methods

Takayuki Itoh
Karsten Klein
Giuseppe Liotta

January 12–15, 2015

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Big Graph Drawing: Metrics and Methods

Organizers:
Takayuki Itoh (Ochanomizu University, Tokyo, Japan)
Karsten Klein (Monash University, Melbourne, Australia)
Giuseppe Liotta (University of Perugia, Perugia, Italy)

January 12–15, 2015

1 Introduction

Graphs provide a versatile model for data from a large variety of application domains, including for example biology, finance, information security, telecommunication, software engineering, and social sciences. Graph visualization helps scientists and engineers to understand critical issues in these domains. However, the depth of understanding depends on the quality of the graph representation.

Since the 1990s, the size of relevant data sets has grown exponentially. In biotechnology, sequencing machines produce one terabase of data in a single sequencing run. In the finance arena, algorithmic trading has reached hundreds of thousands of trades per second. Data from social networks, wireless networks, and the web itself continue to grow at a rate that outstrips current methodologies. Human understanding of this data has clear social and economic benefits; this has led to a strong demand for visualization of huge graphs.

Good visualization can facilitate efficient visual analysis of the data to detect patterns and trends, and can also be used for the presentation of the underlying information. Important aspects for the development of graph drawing methods are the efficiency and accuracy of the algorithms, and the quality, i.e. readability, of the resulting pictures.

For a successful application of visualization approaches in practice these approaches need to meet the particular requirements for the intended use-cases. Quality metrics need to be developed to measure and optimize the graph drawing quality for the task at hand. These metrics need to cover both how accurate the drawing represents the data as well as how well the drawing can be perceived by the human user.

Classical graph drawing quality metrics such as the number of crossings do not seem to be well suited to capture the readability, and thus the usefulness, of big graph drawings. And while the research in the graph drawing field focuses mainly on the geometry of the drawing based on the graphs structure, successful approaches in practice usually need to cover also additional semantics and data annotations. Network analysis that includes such information can greatly support the understanding and knowledge discovery, allowing for example the search for important actors, identification of their role, or detection of interaction patterns in biological and social networks. A useful graph representation should reflect this information and enable the user to interactively explore the graph. As the graphs might be too large for representation, navigation

methods that allow to only view focus areas within surrounding context, and aggregated views that employ for example clustering methods to simplify the representation have to be considered.

This seminar brought together researchers working on graph visualization and people from application areas to discuss these topics and to develop new methods and metrics for big graph visualization.

2 Program and Schedule

The seminar's schedule reflected the idea that most time should be devoted to collaborative work on open research questions, instead of having a large number of talks on recent research results. Thus the main part of the seminar time was used for working groups and general discussion sessions. Only a few selected talks were given that gave an overview on the state of the art of large graph visualization, and presented challenges from application areas.

- Stephen Kobourov gave an overview from a Graph Drawing perspective,
- Kwan-Liu Ma presented an overview from a Information Visualization angle,
- Tom Freeman, Michel Westenberg, and Sean O'Donoghue presented challenges from application areas and concepts to cope with the corresponding problems.

After the introductory and talks sessions on the first morning, a general discussion lead to the formulation of challenges and research questions for working groups. The rest of the seminar was spent in working groups, except for a few feedback and reporting sessions.

2.1 Schedule

Day 1:

09:00:	Welcome and self-introduction
10:30:	Break
11:00:	Talks session
12:30:	Lunch
14:00:	Discussion - formulation of challenges and research questions
15:30:	Break
16:00:	Breakout groups - tackling selected topics

Day2:

09:00:	Talks session
10:30:	Break
11:00:	Working groups
12:00:	Group photo
12:30:	Lunch
14:00:	Working groups
15:30:	Break
16:00:	Working groups and group reports

Day3:

09:00:	Working groups
10:30:	Break
11:00:	Working groups
12:30:	Lunch
13:30:	Excursion
19:00:	Banquet

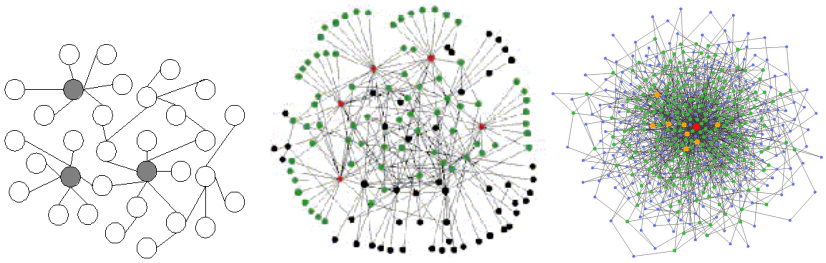
Day4:

09:00:	Talks session
09:30:	Working groups
11:00:	Final group presentations and wrap up
12:30:	Lunch – end of the seminar

3 Working Group Reports

3.1 Cognitive Scalability of Graph Visualization

Participants: Daniel Archambault, Stephan Diehl, Tim Dwyer, Chun-Cheng Lin, Kazuo Misue, Tamara Munzner, Hsiang Yun Wu



Introduction: There has been much work done on designing algorithms that can efficiently scale to create pictures of very large graphs. However, what remains a more open question, is whether pictures of very large and complex networks are "useful" in supporting tasks that have real-world application. In cognitive psychology Millerś "seven plus or minus two" is commonly accepted as a rule-of-thumb for the limitation on peoples' working memory. Working memory is an example of one of the "cognitive ceilings" that might affect peoples' ability to reason about large networks. Does working memory and other cognitive limitations have implications for the size and complexity of graphs that we should be trying to visualize? The goal of this working group is to better understand these cognitive limitations and hence "rules of thumb" for the scalability of graph vis. This knowledge would guide builders of graph visualisation systems in choosing when to filter, aggregate or otherwise "chunk" parts of large graphs in order to produce a less overwhelming display. We have noticed that controlled experiments tend to focus on graph datasets of size within a fairly limited window. For example, tens to a few hundred nodes and low density.

Scope:

- The answer is surely task dependent; we need to restrict our study to a small set of very concrete tasks.
- We need to restrict the graph data sets considered to particular structures (e.g. small world/scale free) so that the density is high enough to be the difficult rather than trivial case.
- The type of tasks that are most likely to be adversely affected by graph complexity are those where a subgraph of interest is difficult to visually isolate from the rest.
- Finding the threshold.

Goals: Literature survey: What are the scalability limits of node-link graph drawings?

- A complete list of all graph drawing evaluation studies using node-link representations. Out of scope: other visual representations (matrix, hybrid). In scope: both static and interactive, both flat and multi-level representations.
- Compile data for graph sizes/densities, graph types, and tasks reported on in these studies. (14 coded so far)
- Glean evidence from those studies concerning scalability of node-link representations: when do we hit the wall?
  - Hypothesis/Complaint: most papers do not report ceiling/floor effects explicitly, many papers do not explicitly explain/justify why they chose the sizes used.
  - A comprehensive list of tasks that have been tested in this context, categorized by:
    - Preattentive vs slower, higher-level cognitive analysis
    - Local vs global scope
- Timeline
  - Maybe some work in next few months
  - Begin in earnest starting in April
  - Study design as possible next step
    - Test for transition points between A and B
      - Visual representation change (cf treemap/barchart from Heer&Agrawala, small multiples vs superimposed line charts from Javed/Elmqvist)
      - Visual representation parameters (horizon graph Heer et al)
      - Flat vs multi-level layout (Archambault et al)

3.2 Dynamic Graphs

Participants: Daniel Archambault, Tom Freeman, Leszek Gasieniec, Seokhee Hong, Hiroshi Hosobe, Takayuki Itoh, Kwan-Liu Ma, Kazuo Misue, Mark Ortmann, Huamin Qu, Ken Wakita

Introduction: Most of the real-world applications generate time-varying graph data. Some examples include biological networks, social networks, collaboration networks, and the Internet. While many solutions have been introduced for visualizing static graphs, the problem of visualizing dynamic graphs has not been comprehensively studied. The main objective of this discussion group is thus to identify all the essential issues concerning the problem of visualizing and understanding dynamic graphs. We also wish to distinguish different types of dynamic graphs and corresponding analysis tasks. This writing summarizing our discussion attempts to provide a guide for those who intend to pursue research in dynamic graph visualization.

Classification of Dynamic Graphs: There are different types of dynamic graphs. A classification may be done as follows:

1. By the fashion the data is obtained:
    - Pre-collected data
    - Real-time, streaming data
2. By the given network properties:
    - Structure-varying networks
    - Attribute-varying networks
3. By the derived/observed attributes/phenomena:
    - Periodical
    - Chaotic/random
    - Converging

According to the type of the data, appropriate methods are then designed/employed for visualization.

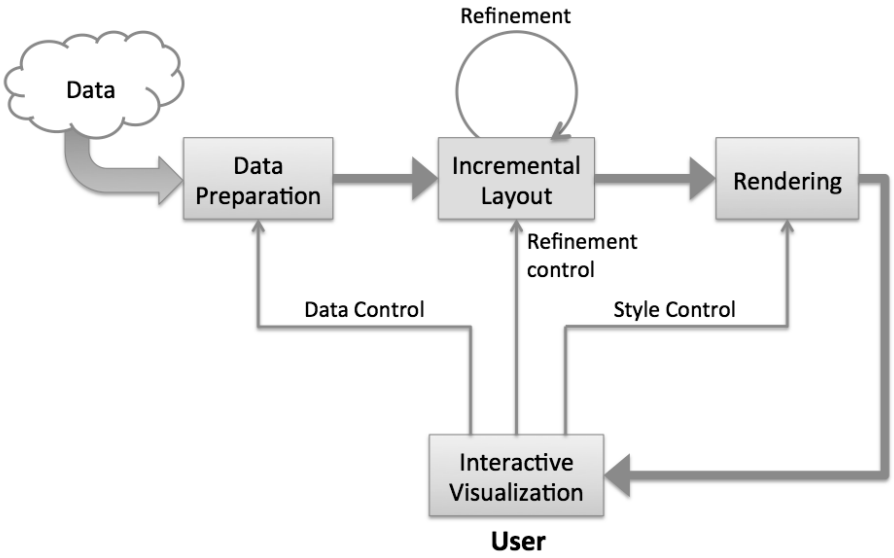
Scientific/Business Objectives of Dynamic Graph Visualization We consider the following objectives unique to visualizing dynamic graphs: 1. Depicting and understanding process, evolution, and temporal trend 2. Detecting abnormalities 3. Finding dependencies 4. Making prediction based on the above to support decision making

Research Topics We have collectively identified research topics that will help us enhance our ability to handle dynamic graphs.

1. What are the visual paradigms to use?
2. Visualization tasks to consider:
    - Real-time monitoring
    - Sporadic checking
    - Storytelling and offline analysis
3. How to deal with varying rate input, multi-source data, different resolutions, etc.?

4. How to achieve coherent motion to preserve mental map while highlighting outliers?
5. How to make history-preserving visualization while highlighting new entities/events?
6. Precomputed clustering versus incremental clustering
7. User input and user directed visualization

Process/Framework: A basic process with essential steps to handle and visualize streaming network data was proposed by Kwan-Liu Ma. It was extensively discussed and approved by the group.



The research topics correspond to each step in this process are given below:
Data preparation: 2, 3, 5, 6
Layout (& mapping): 1, 2, 3, 4, 5
Rendering: 4, 5
Interactive visualization: 1, 2, 3, 4, 5, 6, 7
We have also considered the possibility of adding an intelligent system for quality control and making recommendations to the users.

References J. Abello, D. Archambault, J. Kennedy, S. Kobourov, K.-L. Ma, S. Miksch, C. W. Muelder, and A. Telea, Temporal Multivariate Networks, Ch. 8 in Multivariate Network Visualization. Lecture Notes in Computer Science, 8380. Springer-Verlag, 2014

D. Archambault, H. Purchase, and B. Pinaud. Animation, Small Multiples, and the Effect of Mental Map Preservation in Dynamic Graphs. IEEE Transactions on Visualization and Computer Graphics, 17(4):539-552, 2011.

F. Back, M. Burch, S. Diehl, D. Weiskopf. The State of the Art in Visualizing Dynamic Graphs, EuroVis 2014 STAR Report.

B. Bach, E. Pietriga, J.-D. Fekete. GraphDiaries: Animated Transitions and Temporal Navigation for Dynamic Networks. IEEE Transactions on Visualization and Computer Graphics, 20 (5):740 754, 2014.

U. Brandes, N. Indlekofer, M. Mader. Visualization Methods for Longitudinal Social Networks and Stochastic Actor-Oriented Modeling, Social Networks, 3(24):291-208, 2012.

A. Sallaberry, C. W. Muelder, and K.-L. Ma. Clustering, Visualizing, and Navigating for Large Dynamic Graphs, in proceedings of the 20th International Conference on Graph Drawing, 2012, pages 487-498.

3.3 Network + Set-System Topology (Hypergraphs)

Participants: Sean O'Donoghue, Karsten Klein, Shigeo Takahashi, Michel Westenberg, Michael Wybrow.

Introduction: Biological networks have a growing importance for the interpretation of high-throughput omics data. Integrative network analysis makes use of statistical and combinatorial methods to extract smaller subnetwork modules, and performs enrichment analysis to annotate the modules with ontology terms or other available knowledge. This process results in an annotated module, which retains the original network structure and includes enrichment information as a set system. Visualizing an annotated module amounts to visualizing a hypergraph consisting of binary edges (interactions) between nodes and n-ary edges (annotation sets). Previous work consists of an approach that displays annotations as contours on top of a node-link diagram (called eXamine, see http://www.biomedcentral.com/1471-2105/15/201). The nodes, links, and contours are laid out in a unified way by an extension of a self-organizing map algorithm. The approach was designed to accurately convey small and annotated modules, consisting of up to about thirty nodes and categories of up to about twenty annotations. While this covers common analysis cases in the application domain, scalability is a concern as the approach focuses on small modules to enable accurate depiction of set contours. The main research question is how to address this scalability issue (computational and perceptual) so that a comprehensive layout can be made if the module consists of hundreds of nodes and/or if there are dozens of annotation sets to visualize at the same time.

Progress: In the breakout group, we developed several ideas in terms of both network layout as well as set topology representation to deal with the scalability issue. We plan to extend these ideas further, and implement these in a prototype. Some first steps have already been made after the return from the Shonan seminar, and an initial version of the prototype has been made. We aim at publishing the results at a biological data visualization venue, such as BioVis or the Eurographics workshop on Visual Computing for Biology and Medicine. We also intend to make the resulting tool available publically for the research community.

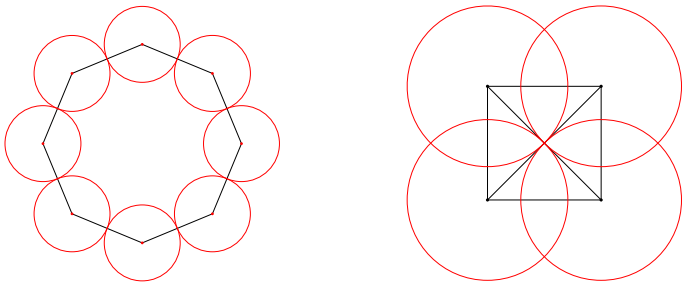
3.4 Low Ply Drawings

Participants: Emilio Di Giacomo, Walter Didimo, Seok-hee Hong, Michael Kaufmann, Stephen Kobourov, Giuseppe Liotta, Kazuo Misue, Antonios Symvonis, Hsu-Chun Yen.

Introduction: Let D be a straight-line drawing of a graph. For each vertex $v \in D$, let C_v be the open disk centered at v and whose radius r_v is half the length of the longest edge incident to v. Denote by S_q the set of disks sharing a point $q \in \mathbb{R}^2$, i.e., $S_q = \{C_v \mid \|v - q\| < r_v\}$. The *ply number* of D, denoted by $\mathsf{pn}(D)$, is defined as

$$\mathsf{pn}(D) = \max_{q \in \mathbb{R}^2} |S_q|.$$

In other words, the ply number of D is the maximum number of disks C_v mutually intersecting in D (see Figure 2 for an illustration). We studied the problem of computing a drawing of a given graph G with minimum ply number over all possible straight-line drawings of G. The motivation behind the study of this problem is that drawings with small ply number seem to have several good quality measures: they tend to have uniform edge length, good angular resolution, and uniform vertex distribution. Thus, the ply number seems to be a reasonable measure to optimize when drawing graphs. The inspiration for the definition of the ply number comes from a paper by Eppstein and Goodrich where it is experimentally shown that real road networks have small ply number [2].



(a) (b)

Figure 2: (a) A drawing with ply number 1. (b) A drawing with ply number 2.

We first investigated drawings with ply number 1. It is easy to see that if a drawing has ply number 1, then all its edges must have the same length. Namely, let $e_1 = (u, v)$ and $e_2 = (v, w)$ be two edges of a drawing D that share the vertex v and suppose that e_1 is longer than e_2 in D. The disk C_v centered at v has a radius that is at least half of the length of e_1 and therefore more than half of the length of e_2. On the other hand, the disk C_w centered at w has radius that is at least half of the length of e_2. It follows that C_v and C_w have a non-empty intersection. The following result can be easily proven.

Theorem 1 *A graph has a drawing with ply number 1 if and only if it has a contact representation with unit disks*[1]

Proof: If we connect the centers of the circles of a unit disk representation of a graph G, then we get a drawing of G with ply number 1. Conversely, suppose that D is a drawing of G such that $\mathsf{pn}(D) = 1$. As observed above all edges of D have the same length. It follows that all disks C_v have the same radius and they touch in the middle of each edge. Thus, the drawing consisting of all disks C_v is a unit disk representation of G. □

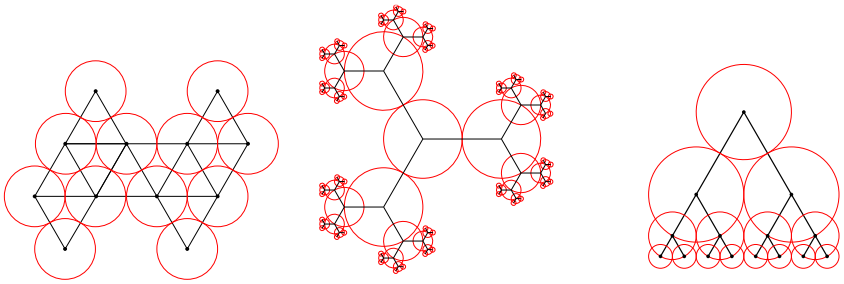
[1] A contact representation with unit disks is a representation where each vertex is represented by a disk of radius one and two vertices are adjacent if and only if the two corresponding disks touch each other.

Since it is NP-hard to decide whether a graph admits a contact representation with unit disks [1], we immediately obtain the following result.

Corollary 1 *It is NP-hard to decide whether a graph G admits a drawing with ply number* 1.

On the positive side we can prove the following result.

Theorem 2 *Every maximal outerplanar graph with maximum vertex degree* 5 *admits a drawing with ply number* 1 *(see Figure 3a for an example).*



(a) (b) (c)

Figure 3: (a) A drawing of a maximal outerplanar graph of maximum degree 5 with ply number 1. (b) A drawing of a binary tree with ply number 2. (c) A drawing of a binary tree with ply number 2.

We also investigated the properties of the drawings of cycles that have ply number 1. The following necessary condition can be easily proven.

Theorem 3 *If a drawing D of a cycle has $\mathsf{pn}(D) = 1$, then all the edges have the same length and all internal angles are at least* $60°$.

One can create a drawing of a k-cycle that satisfies the necessary condition of Theorem 3 by placing the vertices of the cycle at the corners of a regular k-gon. This type of drawing has indeed ply number 1 (see Figure 2a). On the other hand, there exist drawings of a cycle with ply number 1 where the cycle is not drawn as a regular k-gon.

We investigated drawings with ply number 2 for binary trees and proved the following.

Theorem 4 *Every binary tree has a drawing D such that $\mathsf{pn}(D) = 2$, which is worst case optimal.*

We developed two different drawing techniques for binary trees, which are illustrated in Figures 3b and 3c. Notice that both techniques produce drawing with polynomial area for balanced binary trees, but exponential area is obtained if the tree is not balanced. This naturally motivates the following problem.

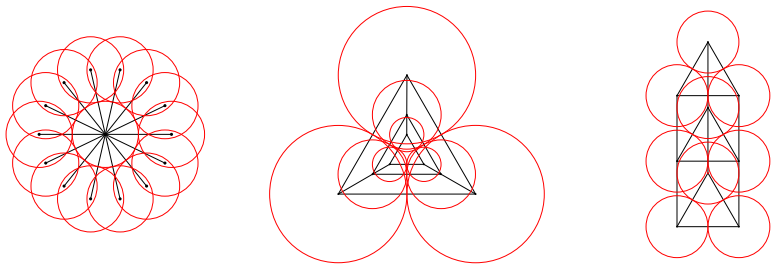
Problem 1 *Is it possible to draw a general binary tree in polynomial area with ply number* 2*?*

Other problems for trees that are worth investigating are the following.

Problem 2 *Is it possible to draw ternary trees with ply number* 2*?*

Problem 3 *What is the ply number for general trees?*

About Problem 3, it is easy to see that not all trees admit a drawing with constant ply number. Namely, any drawing D of a star of degree d has $\mathsf{pn}(D) >= \lceil \frac{d}{2\pi} \rceil$, and in fact this lower bound is worst case optimal (see Figure 4a).



(a) (b) (c)

Figure 4: (a) A drawing of a star with degree $d = 14$; the ply number is 3. (b) A planar drawing of a graph G with ply number $O(n)$. (c) A non-planar drawing of G with ply number 2.

An interesting observation about drawings with small ply number, is that non-planar drawings can have smaller ply number than planar ones. In Figure 4b and 4c two drawings of the same graph are shown. The first one is planar but its ply number is $O(n)$, the second one is non-planar but its ply number is 2.

We finally considered some generalizations of the drawings with small ply number. An *empty-ply drawing* is a drawing of a graph such that the disk C_v of each vertex v contains no vertex other than v. In other words, in an empty-ply drawing, intersections of disks are accepted provided that no vertex is inside the disk of another vertex. Our intuition is that empty-ply drawings have small ply number. The following problems can be investigated.

Problem 4 *Is the ply number of empty-ply drawings constant?*

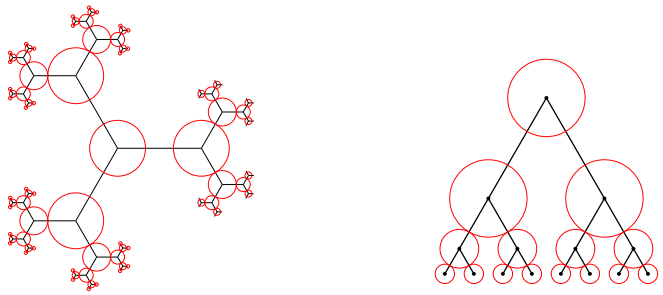
Problem 5 *Characterize those graphs that admit an empty-ply drawing.*

Another generalization is the following. For each vertex $v \in D$, let C_v^α be the open disk centered at v and whose radius r_v^α is α times the length of the longest edge incident to v. Let $S_q^\alpha = \{C_v^\alpha \mid \|v - q\| < r_v^\alpha\}$. The *$\alpha$-ply number* of D, denoted by α-$\mathsf{pn}(D)$, is defined as

$$\alpha\text{-}\mathsf{pn}(D) = \max_{q \in \mathbb{R}^2} |S_q^\alpha|.$$

With this definition we consider smaller or larger disks around each vertex depending on the value of α (the "standard" definition of ply number is obtained for $\alpha = 0.5$), thus reducing or increasing the ply number of a drawing. For example, the drawings produced by the techniques behind Theorem 4 have α-ply number 1 if we choose $\alpha \leq \frac{1}{3}$ (see Figures **??** and **??** for an illustration).

Theorem 5 *Every binary tree has a drawing D such that α-$\mathsf{pn}(D) = 1$, for $\alpha \leq \frac{1}{3}$.*



(a) (b)

Figure 5: Two drawings of a binary tree with α-ply number 1 for $\alpha = \frac{1}{3}$.

The definition of α-ply number opens up many research directions. Drawings with different values of the α-ply number for different values of α can be investigated. One problem suggested by Theorem 5 is the following.

Problem 6 *For which value of α every d-ary tree admits a drawing D such that α-$\mathsf{pn}(D) = 1$?*

References

[1] H. Breu and D. G. Kirkpatrick. Unit disk graph recognition is NP-hard. *Computational Geometry*, 9(12):3 – 24, 1998. Special Issue on Geometric Representations of Graphs. doi:10.1016/S0925-7721(97)00014-X.

[2] D. Eppstein and M. T. Goodrich. Studying (non-planar) road networks through an algorithmic lens. In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '08, pages 16:1–16:10, New York, NY, USA, 2008. ACM. doi:10.1145/1463434.1463455.

4 Participants

- Daniel Archambault, Swansea University
- Emilio Di Giacomo, University of Perugia
- Walter Didimo, University of Perugia
- Stephan Diehl, University of Trier
- Tim Dwyer, Monash University
- Tom Freeman, Roslin Institute, The University of Edinburgh
- Leszek Gasieniec, University of Liverpool
- Seokhee Hong, University of Sydney
- Hiroshi Hosobe, Hosei University

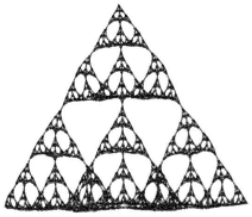
- Takayuki Itoh, Ochanomizu University
- Masahiko Itoh, The University of Tokyo / NICT
- Michael Kaufmann, Eberhard-Karls-University Tübingen & yworks GmbH
- Karsten Klein, Monash University
- Stephen Kobourov, University of Arizona
- Chun-Cheng Lin, National Chiao Tung University
- Giuseppe Liotta, University of Perugia
- Kwan-Liu Ma, University of California at Davis
- Kazuo Misue, University of Tsukuba
- Tamara Munzner, University of British Columbia
- Seán O'Donoghue, CSIRO & Garvan
- Mark Ortmann, University of Konstanz
- Huamin Qu, Hong Kong University of Science and Technology
- Antonio Symvonis, National Technical University of Athens
- Shigeo Takahashi, The University of Tokyo
- Ken Wakita, Tokyo Institute of Technology
- Michel Westenberg, Eindhoven University of Technology
- Hsiang-Yun Wu, The University of Tokyo
- Michael Wybrow, Monash University
- Hsu-Chun Yen, National Taiwan University

Overview of Talks

Force-Directed Layout of Node-Link Diagrams

Stephen Kobourov, University of Arizona

Some of the most flexible algorithms for calculating layouts of simple undirected graphs belong to a class known as force-directed algorithms. Also known as spring embedders, such algorithms calculate the layout of a graph using only information contained within the structure of the graph itself, rather than relying on domain-specific knowledge. Graphs drawn with these algorithms tend to be aesthetically pleasing, exhibit symmetries, and tend to produce crossing-free layouts for planar graphs. Force directed algorithms for drawing graphs have a long history and new variants are still introduced every year. Their intuitive simplicity appeals to researchers from many different fields, and this accounts for dozens of available implementations. As new relational data sets continue to be generated in many applications, force directed algorithms will likely continue to be the method of choice.



Three bioinformatics applications using big graphs: Aquaria, Minardo and Rondo

Seán O'Donoghue, CSIRO & Garvan Inst. of Medical Research, Sydney

In this talk I outline three recently developed bioinformatics applications that use big graphs to facilitate exploration, sense-making, and discovery.

(1) Minardo (http://dx.doi.org/10.1063/1.4825010 and http://minardo.org) is designed for time-series mass-spectroscopy data on post-translational modifications; these experiments are relatively new, with each experiment producing time-series data for up to 10,000 nodes (phosophosites), thus creating an analysis and visualization challenge that currently few other resources can address. Minardo uses a novel layout that incorporates time, cell topology, and a 'storyline' path-based representation for key proteins that have roles at multiple time points.

(2) Aquaria (http://nature.com/articles/nmeth.3258 and http://aquaria.ws) is a new web resource that, for each of ¿500,000 protein, uses graphs to create a visual summary of all related atomic-scale 3D models (typically 50 - 2,000 models per protein). The models are organized into groups, first by region of sequence match, and secondly by the number and identity of molecules present in each model. Feedback from users indicates Aquaria is effective in helping users quickly gain an overview of all available models and select which are most relevant for their work.

(3) Rondo is an application still in development for managing 'HiC' data (`http://doi.org/10.1126/science.1181369`), a very new kind of experiment that

determines millions - even billions - of physical connections occurring between parts of an entire genome. Rondo enables interactive exploration of all currently available HiC datasets, and has potential to give insight into fundamental mechanisms of genome organization, dynamics, and how these are perturbed in diseases such as cancer.

Biological Network Visualization

Michel Westenberg, TU Eindhoven

Biological networks pose many challenges for visualization of both topological structure and node/edge attributes. There are different types of network, each with specific structural characteristics. For example, gene regulatory networks (describing gene-to-gene interactions) exhibit out-degrees with a scale-free distribution, in-degrees bound by a low maximum and few and small cycles. Besides these global structural properties, subnetworks of a specific structure, called motifs, provide important knowledge about gene regulatory networks to domain experts. Standard layout methods typically do not exploit these structural characteristics. The first part of the talk is about visualization techniques that were designed specifically for a certain type of biological network, namely the Compressed Adjacency Matrix for gene regulatory networks and MetaViz for metabolic networks.

The second part of the talk is devoted to integrative network analysis, in which node attributes play an important role. Initial approaches aimed at visualizing whole networks using color coding of nodes to reflect node attribute values. However, after working several years in this direction ourselves, it became clear that a view on the whole network was not so relevant for interpretation by the domain experts. Recent analysis approaches use statistical and combinatorial methods to extract smaller sub-network modules (with tens of nodes rather than thousands). Further enrichment analyses can then rank annotation items (e.g. Gene Ontology terms or gene-to-KEGG pathway relationships) that relate to a significant number of module nodes. These annotated modules provide the input to our visualization tool eXamine, which extends upon Self Organizing Maps to simultaneously lay out nodes, links, and contours depicting annotation sets. eXamine provides a set-oriented visual analysis approach that allows the domain expert to focus on the relevant details only, and to quickly formulate hypotheses about the underlying biology.

Big Graphs and Biology: Visualisation and Analysis of Data Matrix

Tom Freeman, The Roslin Institute, University of Edinburgh

Emerging Topics in Big Network Visualization

Kwan-Liu Ma, University of California at Davis

Acknowledgments The organizers would like to thank all participants, NII, and the Shonan staff for all their contributions. We hope that the research during the seminar, as well as the exchange between the participants, will lead to new successful collaborations and research findings in the future.