COMP5048 Information Visualisation

2011 Semester 2

- Tuesday 3-5pm
- SIT Lecture Theater
- Lecturer: Seokhee Hong (shhong@it.usyd.edu.au)
- Consultation: Tuesday 5-6pm

COMP5048 Course Outline

- Unit Specification
- Assumed Knowledge
- Learning Outcomes
- Assessment
- USYD Policies
- Lecture Time Table
- References

Unit Specification

- Information Visualisation aim to make good pictures of abstract information, such as stock prices, computer networks, social networks and software diagrams.
- The main challenge is to design and implement effective and efficient algorithms that produce good pictures of abstract data.
- The unit will review basic concepts, techniques and fundamental algorithms to achieve good visualisation of abstract.

Assumed Knowledge

- Basic Knowledge in Algorithms/Data Structures
- Basic Skills in Programming

Learning Outcomes

- understanding of basic concepts, techniques and algorithms for good visualisation of abstract data
- using efficient algorithms to produce good visualisation of abstract data
- applying effective visualisation methods for specific application area

Assessment

- Week 7: Assignment 1 (10%)
- Week 11-13: Presentation (10%)
- Programming Assignment (40%) - group work
  - Week 9: Report 1 (10%)
  - Week 13: Report 2/demo/presentation (30%)
- Exam: 40%
USYD/SIT Policies

- You are required to carefully read the policies on
  - Academic Honesty and Plagiarism
  - Special consideration due to illness and misadventure
  - No late submissions allowed

Topics Covered: Week 1-7

Approximate schedule: topics are subject to change

- Week 1 (SH): Introduction
- Week 2 (SH): Tree Drawing
- Week 3 (PE): Spring Algorithm
- Week 4 (PE): Drawing Clustered Graphs
- Week 5 (SH): Network Analysis
- Week 6 (SH): Visual Analytics
- Week 7 (SH): 3D and Interaction

Topics Covered: Week 8-13

Approximate schedule: topics are subject to change

- Week 8 (FF): Sugiyama Method
- Week 9 (TH): Evaluation
- Week 10 (FF): Drawing Planar Graphs
- Week 11: Student Presentation
- Week 12: Student Presentation
- Week 13: Student Presentation

References

- Information Visualization by Robert Spence, Pearson Addison Wesley, 2000
- Conference Proceedings: GD, IEEE InfoVis, EuroVis, VAST, PacificVis
- Journals:
  - Information visualization, www.palgrave-journals.com/ivs/

Lecture 1

Information Visualisation and Graph Drawing

Visualisation

- Visualisation: the use of computer-supported, interactive, visual representations of data to amplify cognition.
- Scientific visualisation: the use of computer-supported, interactive, visual representations of scientific data to amplify cognition.
- Information visualisation: the use of computer-supported, interactive, visual representations of abstract data to amplify cognition.
Information Visualisation

the loss of Napoleon’s army

• Edward R. Tufte, The Visual Display of Quantitative Information
  • by Charles Joseph Minard (1781-1870)
  • Russian-Polish border 422,000 men / Moscow 100,000 men.

Information Visualisation research aims to make pictures of abstract data so that humans can understand, navigate, and manipulate the data.

Good visualisation

• H. Beck

Bad visualisation

• ALP
Visualisation of Abstract Information

There are two steps:
1. **Analysis**: extracting a graph from the information
2. **Visualisation**: Graph drawing

Visualisation of football transfers

- Drew moved from the Panthers to the Eels
- Miles moved from the Roosters to the Eagles
- Green moved from the Cowboys to the Roosters
- O’Hara moved from the Bulldogs to the Raiders

Visualisation of Social Networks

- **email friends**

Email log files reflect relationships between people

Definition: friends network
  - X and Y are *email friends* if
  - X sends more than 5 messages per day to Y, and
  - Y sends more than 5 messages per day to X.

The *email_friends* graph can be derived from email log files.
There are two steps to visualising graphs:
1. **Analysis**: extracting a graph from the information
2. **Graph drawing**

**Reference Model for Visualisation**
- Data
- Data Table
- Visual Structures
- Views
- Data Transformations
- Visual Mappings
- View Transformations
- Human Interaction

**Bad Visualisation**

**Good Visualisation**

**Tree Maps**
[Johnson, Shneiderman 91] Treemaps: A Space-filling Approach to the Visualization of Hierarchical Information

**Cone Tree**
[Robertson, Mackinlay, Card, CHI 91] Cone Trees: Animated 3D visualizations of hierarchical information.
**2D Hyperbolic Tree Browser**

- [Lamping, Rao, Pirolli, CHI'95] A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies.
- Distortion & Hierarchy

**M.C. Escher, Circle Limit IV (Heaven and Hell)**

**H3: 3D Hyperbolic**

- data: web hyperlinks
- quasi-hierarchical graphs: can find reasonable spanning tree using domain-specific information
- goal: scalability

**Walrus - Graph Visualization Tool**

**Network Data Visualisation**

- [Becker, Eick, Wilks 95] SeeNet
- [Cox, Eick 95] 3D Displays of Network Traffic
- SeeNet3D

**Focus+Context (FishEye View)**

- the perspective wall (Mackinlay, Robertson and Card, 1991)
Visualizing the Topology of the MBone

- **time:** 1996
- **data:** MBone tunnels
- **task:** find badly placed tunnels
- **goal:** simple baseline
- **method:** 3D geographic

Tamara Munzner and Eric Hoffman and K. Claffy and Bill Fenner

Database Visualisation

VISDB
Five-dimensional artificially generated data set (100,000 points) in simple configuration.

NicheWorks: Exploring Large Networks

- **NicheWorks - Interactive Visualization of Very Large Graphs,** by Graham J Wills.
- Typical analyses performed using NicheWorks have between 20,000 and 1,000,000 nodes.

International Calling Fraud

Overview of calling patterns: 40,000 calls involving 35,000 callers

High users’ calling patterns
International Calling Fraud

- Possible fraud pattern
- The Israel-Jordan-UAE generated subset
- Zooming in to those callers calling more than one country

SPIRE

- SPIRE: Spatial Paradigm for Information Retrieval and Exploration
- SPIRE provides a wealth of tools for exploring the information, including query, subset, and trend analysis tools.

http://www.pnl.gov/infoviz/spire/spire.html
Pacific Northwest National Laboratory, USA

Galaxies

- The Galaxies visualization uses the image of stars in the night sky to represent a set of documents.
- Each document is represented by a single "docustar."
- Closely related documents cluster together while unrelated documents are separated by large distances.

Several analytical tools are provided with Galaxies to allow users to investigate the document groupings, query the document contents, and investigate time-based trends.

ThemeView

- The topics or themes within a set of documents are shown as a relief map of natural terrain.
- The mountains in the ThemeView indicate dominant themes.
- The height of the peaks indicates the relative strengths of the topics in the document set.
- Similar themes appear close together, while unrelated themes are separated by larger distances.

ThemeView provides a visual overview of the major topics contained in a set of documents. Combined with its exploration tools, ThemeView permits the analyst to identify unanticipated relationships and examine changes in topics over time.
CAIDA

- CAIDA, the Cooperative Association for Internet Data Analysis, provides tools and analyses promoting the engineering and maintenance of a robust, scalable global Internet infrastructure.
- http://www.caida.org

The AS Internet graph

- One of CAIDA’s skitter project goals is to develop techniques to illustrate relationships and depict critical components of the Internet infrastructure.
- The graph reflects 626,773 IP addresses and 1,007,723 IP links of skitter data from 16 monitors probing approximately 400,000 destinations spread across over 48,302 (52%) of globally routable network prefixes.
- Then aggregate this view of the network into a topology of Autonomous Systems (ASes), each of which approximately maps to an Internet Service Provider (ISP).
- The abstracted graph consists of 7,624 Autonomous System (AS) nodes and 25,126 peering sessions.

The AS Internet graph

A Macroscopic Visualisation of the Internet During October, 2000

- Bill Cheswick, Bell Labs and Hal Burch, CMU
- A long-term project to collect routing data on the Internet.
- This mapping consists of frequent traceroute-style path probes, one to each registered Internet entity.
- They build a tree showing the paths to most of the nets on the Internet.
- These paths change over time, as routes reconfigure and the Internet grows.
- They are preserving this data to show how the Internet grows.

Internet Mapping Project

- Bill Cheswick, Bell Labs and Hal Burch, CMU

Graph Drawing
Graph Drawing

- Graphs are abstract structure that are used to model relational information.
- Graph $G=(V,E)$
  - $V$: set of vertices (objects)
  - $E$: set of edges connecting vertices (relationship)
- Graph Drawing: automatic construction of geometric representations of graphs in 2D or 3D.

Graph Drawing

The classical graph drawing problem is to develop algorithms to draw graphs.

The input is a graph with no geometry.

The output is a drawing of the graph; the drawing should be easy to understand, easy to remember, beautiful.

Applications

- Software engineering
- Database
- Information system
- Realtime system
- Computer Network
- VLSI
- AI
- Data Mining
- Bioinformatics
- Decision support system
- Biology
- Chemistry
- ...
Graphs

- tree
  - free tree
  - binary tree
  - rooted tree
  - ordered tree
- planar graphs
- general graphs
- directed graphs
- extended graph model
  - hierarchical graphs
  - clustered graphs
  - hyper graphs
  - higraphs

Drawing conventions

- polyline drawing
- straight-line drawing
- orthogonal drawing
- grid drawing
- planar drawing
- upward drawing
- convexity
- …

Aesthetics

- readability: the drawing should be easy to read, easy to understand, easy to remember, beautiful.

- crossings
- area
- symmetry
- edge length
  - total edge length, maximum edge length, uniform edge length
- bends
  - total bends, maximum bends, uniform bends
- angular resolution
- aspect ratio
Avoid edge crossings
Avoid edge bends
Avoid long edges

One should spread the nodes evenly over the page. This can be measured:
• minimise area (for fixed size nodes)
  or equivalently
• maximise resolution
  (for a fixed size screen).

Avoid edge crossings
Avoid edge bends
Avoid long edges

Aesthetics
There are many aesthetic criteria for good diagrams:
• minimum edge crossings,
• minimum bends,
• minimum edge lengths,
• maximum resolution,
and many more.

Graph Drawing Algorithms
• Tree Drawing
  • Tidy drawing
  • Free tree drawing
• Planar Graphs
  • Straight-line drawing
  • Orthogonal (grid) Drawing
• Undirected graphs: Spring algorithm (force directed methods)
• Directed graphs: Sugiyama method
  (Layered/Hierarchical drawing)
• Clustered graphs ...

NP-hardness
• minimize crossings
• minimize area
• maximize symmetry
• minimize total edge length
• minimize number of bends
• maximize angular resolution
• ...

Conflicts
Minimize edge crossings
Maximize symmetry

Latour tree visualization system
• I. Herman, G. Melançon, M.S. Marshall, VisSym99

Radial layout of 29773 nodes
Reingold-Tilford layout of 3255 nodes
On-line Graph Navigation System

Maolin Huang

Graphviz, AT&T, USA

GDToolkit, Italy

• Hermes: internet topology

OGDF, Germany

Tom Sawyer Software, USA

Hierarchical layout  Symmetric layout  Circular layout

Social network
Actor Collaboration Network

A Few Good Man  Days of Thunder (1990)
Far and Away (1992)
Eyes Wide Shut (1999)

Kevin Bacon Number 1  Kevin Bacon Number 2

Far and Away

IEEE InfoVis Research

Important researchers (research groups) with their research area:
Ahmed et al.: EuroVis 05 (IEEE InfoVis 2004 competition winner)

Evolution of research area

Visualisation of FIFA 2002 World Cup

Ahmed, Fu, Hong, Quan, Xu: GD 2006 competition winner

Visualisation of Stock Market Data

Visualisation of Fund Manager Movement Graph
Tim Dwyer
Visualisation of Patterns – motif
- "Motifs": small pattern in the network which occurs with significantly high frequency.
- Data: transcriptional regulation network of Escherichia coli.

Visualisation and Analysis of Network Motifs (IV 2005)

Map of protein-protein interactions. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-lethal; orange, slow growth; yellow, unknown). By Hawoong Jeong

Homework
Examples of Good Visualisation
Bad Visualisation
send me a ppt slide with pictures
+ source: data, paper, author, website
by next monday