Evaluation

Learning Objectives

- Evaluating visualizations
- Case studies on Graph Visualization
  - Controlled experiment
  - Non-controlled experiment (regression test)
  - Beyond time and error
    - Mental effort and overall quality measurement
    - Eye tracking
    - Questionnaire and interviews

What is visualisation?

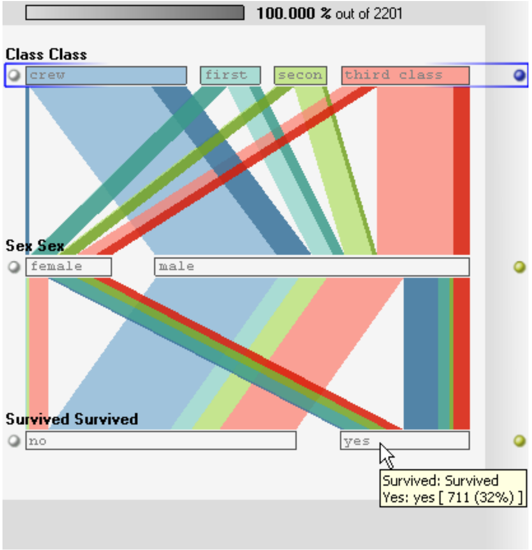
- A readable (visually inspectable) representation of non-visual data.

- A minimal set of requirements for any visualization (R. Kosara, 2007)
  - Is based on data
  - Produces an image
  - Is readable and recognizable

Kosara, R. (2007, 4-6 July 2007). Visualization Criticism - The Missing Link Between Information Visualization and Art. 2007 11th International Conference Information Visualization (IV '07),

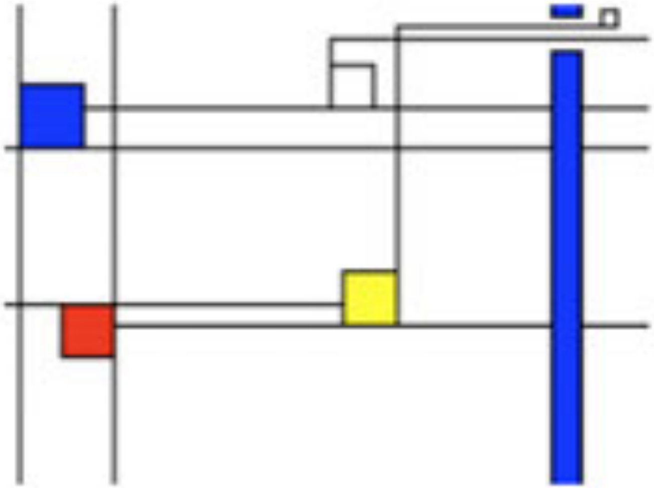
Type of Visualisation – Part I

- Readable/Recognizable



Type of Visualisation – Part II

- Not readable but recognizable



Type of Visualisation – Part III

- Neither readable nor recognizable

Type of Visualisation – Part IV

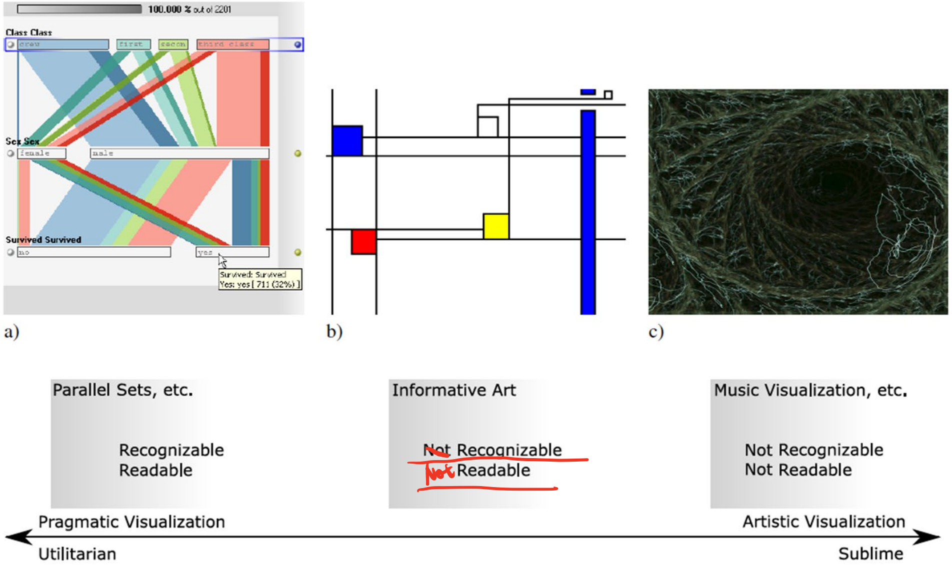
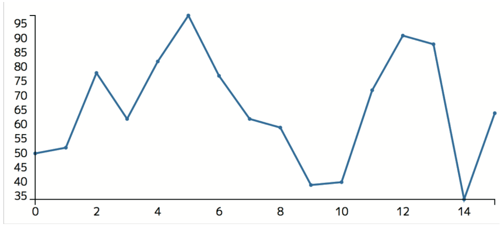
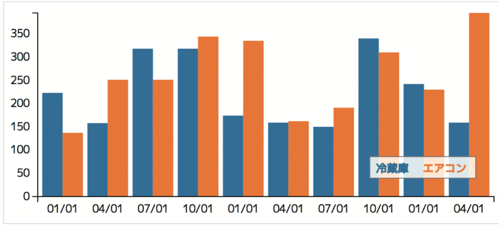
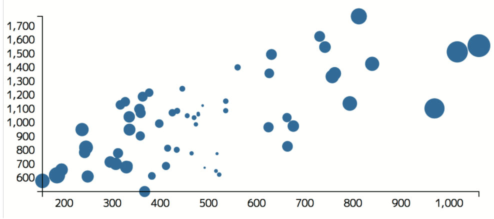


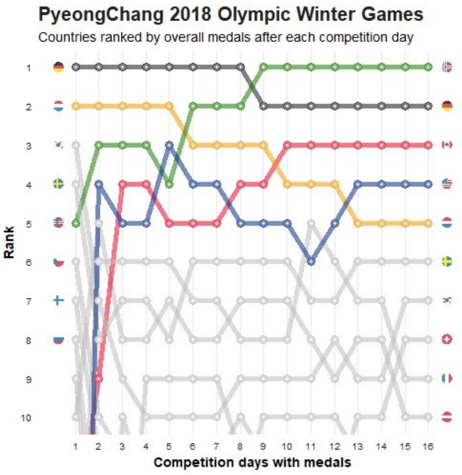
Figure 1. The gamut of data-based visualization. a) Parallel Sets [12] show data about the people on the Titanic, and are readable and recognizable as a visualization; b) Ambient visualization [18] visualizing a bus schedule are readable but require more effort and are not readily recognizable as a visualization; c) Music visualization like MilkDrop [23] is also based on data, but not readable.

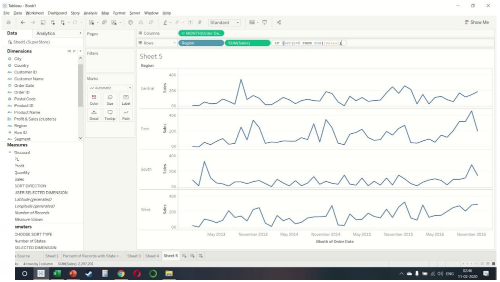
Examples of visualisation (I)

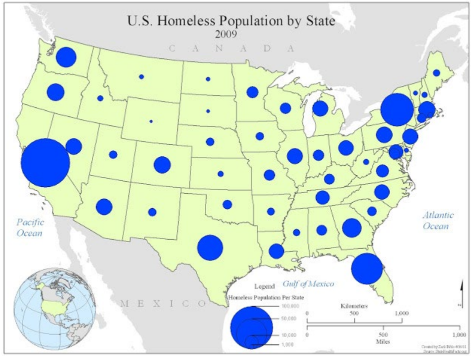




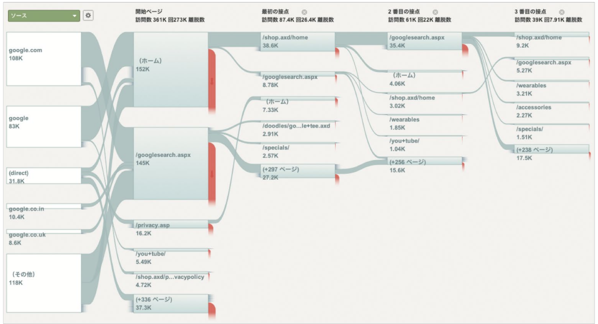


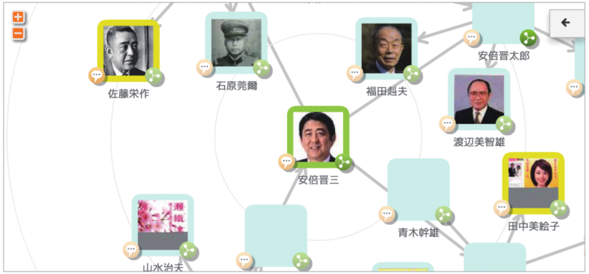




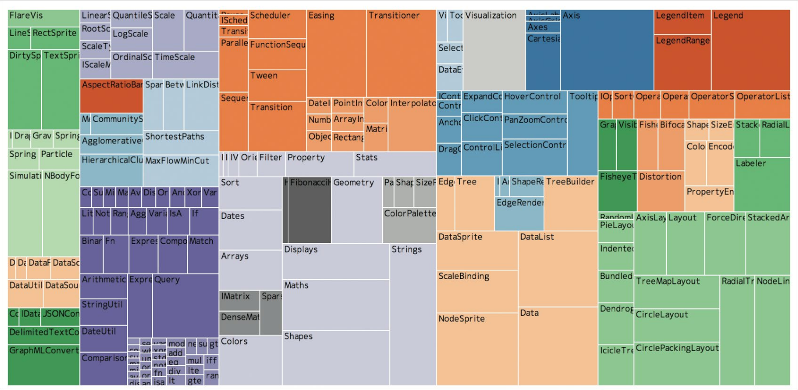


Examples of visualisation (II)





※『スパイシー』
(http://spysee.jp/%E5%AE%89%E5%80%8D%E6%99%8B%E4%B8%89/71814/profile?ref1=from_edge_list)



※『d3.js garally』(http://bl.ocks.org/mbostock/4063582)

Benefits of Visualisation

- Increased resources
  - Offload work to the perceptual system
  - External memory
- Reduced search space
  - Grouping
  - High data density
- Enhanced recognition
  - Recognition instead of recall
  - Abstraction and aggregation
- Manipulable medium

Stuart Card, Jock Mackinlay, and Ben Shneiderman, Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, 1999.

Evaluation Methods

Evaluation

- An assessment of visualisation to see if benefits have been provided.

Is good visualisation == pretty pictures?

To make better visualisation, we need to

- Know how people make sense of data

- Evaluate the visualisations

How does people perceive data? – Part I

Recall…

Gestalt Principles:

- Similarity
- Proximity
- Common Region
- Closure
- Continuity
- Connection

How does people perceive data? – Part II

Visual system (including cognitive process) may yield untruthful results.

Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoatnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the human mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

Evaluation Methods

- Interview
- Questionnaire
- Analytic inspection
- Empirical evaluation

Interview

Qualitative technique

- Gathering information about users by talking directly to them
- A method for discovering facts and opinions of the users

Format:

- It is usually done by one interviewer speaking to one user at a time
- Structured interviews : a pre-defined set of questions and users
- Open-ended interviews : allows for an exploratory approach to uncover unexpected information

Problems:

- The unstructured nature of the resulting data can be easily misinterpreted.

Questionnaire

Qualitative technique

- But results can be quantified. (refer to the previous lecture)

Preparation:

- Keep questions simple, be clear and concise
- Group questions appropriately / give explanation

Pilot questionnaire before distributing it

- It is still unreasonable to think that any one person can anticipate all the potential problems

Problems:

- It is only as good as the questions it contains. you can't get more than what you asked

Question Types – Part I

- General
  - On average, how much time per week do you spend on this system?

    1) Less than 1 hour
    2) 1 to less than 4 hours
    3) 4 to less than 10 hours
    4) over 10 hours

- Open-ended
  - What are the features you think helpful, if any?
  - What are the features you think can be improved, if any?

Question Types – Part II

- Closed
  - Which of the following have you used? (tick all that apply)

    1) Word processor 2) database 3) spreadsheet

  - How easy was it to understand the drawing?

    1) Very Easy 2) Easy 3) Average 4) Difficult 5) Very Difficult

- Scale
  - Indicate how much effort you devoted for this task based on a scale from 0 – 6?

    0 (extremely easy) 1 2 3 4 5 6 (extremely difficult)

comparision is difficult

Questionnaire examples

- Established questionnaires will give more reliable and repeatable results than ad-hoc questionnaires.

- Three questionnaires for assessing the perceived usability of an interactive system:
  - Questionnaire for User Interface Satisfaction (QUIS)
    (Chin, J. P., Diehl, V. A. and Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. Proceedings of SIGCHI '88, (pp. 213-218), New York: ACM/SIGCHI)
  - Computer System Usability Questionnaire (CSUQ)
    (Lewis, J. R. (1995) *IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use.* International Journal of Human-Computer Interaction, 7:1, 57-78)
  - System Usability Scale (SUS)
    (Brooke, J. (1986). "SUS: a "quick and dirty" usability scale". In P. W. Jordan; B. Thomas; B. A. Weerdmeester; A. L. McClelland (eds.). *Usability Evaluation in Industry*. London: Taylor and Francis)

Analytic Inspection

Benefits

- Generate results quickly with low cost
- Can be used early in the design phases

Heuristic evaluation

- Experts review the systems against a list of principles

Cognitive walkthrough

- Starts with a task analysis that specifies the sequence of steps or actions required by a user to accomplish a task
- Then work through the steps

Empirical evaluation

Usability test
- Think aloud, eye tracking
- Formative: helps/guides design
- Against the single UI (visualization)
- Identify usability problems
- Qualitative feedback from users

Controlled experiment
- Summative: measure the final result of the task
- Compare multiple UIs (visualizations)
- Quantitative results, statistical significance

Case Study on Graph Visualisation

Aspects of a high quality graph

- Efficiency:
  - The running time of algorithms should be reasonably fast
- Elegance:
  - Algorithms should be easy to understand and easy to code;
  - Final drawings should be beautiful
- Effectiveness:
  - Graph viewers should understand the underlying data quickly and correctly

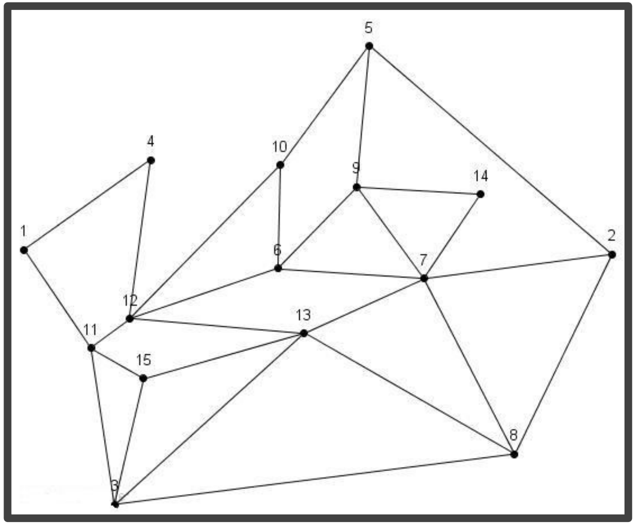
Effectiveness of a graph

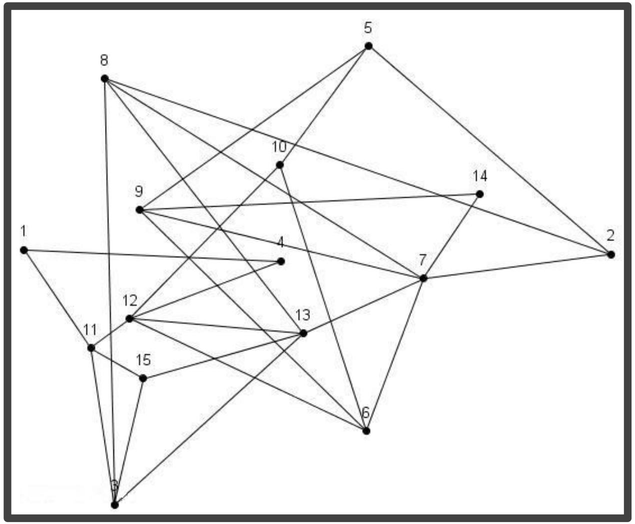
- Visualisation designers are often satisfied with the “coolness” of the technologies they introduced.
  - technology that looks “cool” to the designer might be too complex or unneeded for real users

- It is assumed that graphs should be effective when drawn conforming to some predefined criteria.
  - Maximise symmetry
  - Minimise edge crosses
  - Maximise angular resolution

- However, common senses and intuitions are not reliable.
  - Users, data sets, tasks are different

Which one is better?





We do not know until we actually evaluate them

Case Study I

Case study I: Controlled experiment on performance

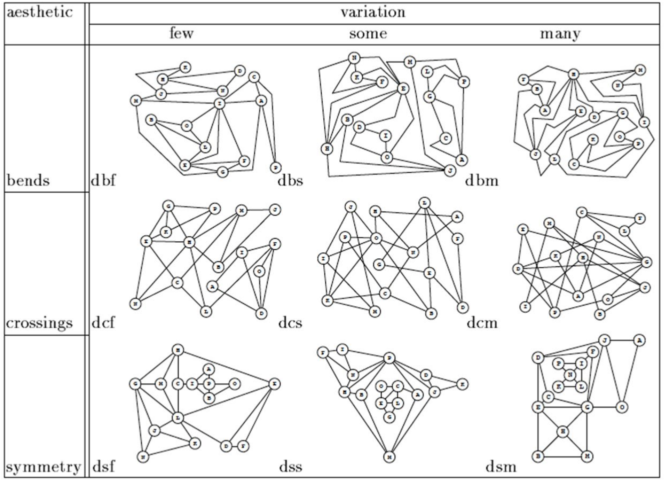


Fig. 1. The Experimental Graph Drawings: dense graph

Purchase, H. C., Cohen, R. F., & James, M. (1996). Validating graph drawing aesthetics. Graph Drawing, Berlin, Heidelberg.

Case study I
Design and analysis

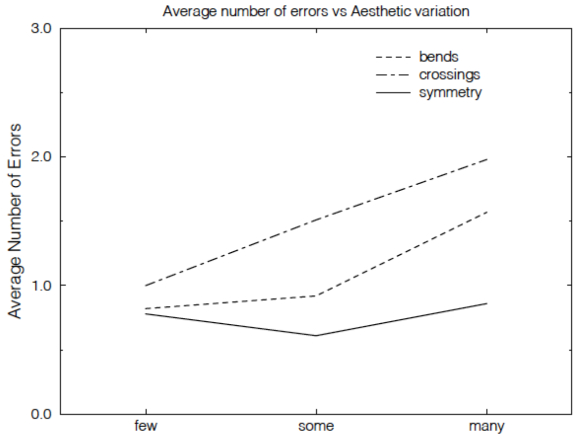
- On dense/sparse graph
- bends, crosses, symmetry
- high / medium / low presence
- accuracy in fixed time (45 seconds)
- Task: find the shortest path between two nodes

- Paper based, within-subjects, random order a “filler” task

Case study I
Results and discussion

- Bends and crossing are important
- Symmetry needs further examination

- A pioneering work that provides empirical evidence for intuition-based aesthetic criteria



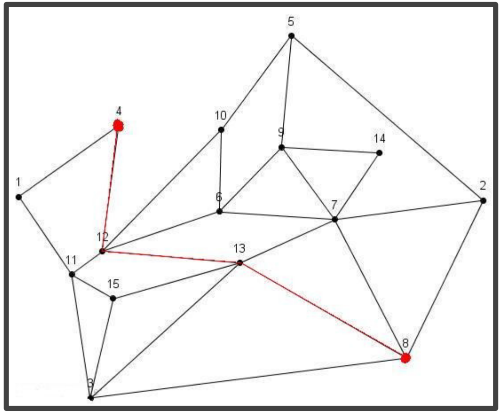
bends	6	18	30
crossings	6	24	42
symmetry	4.57	25.7	51.0

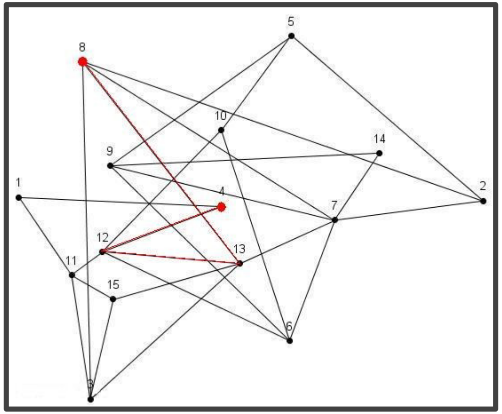
Aesthetic Variation

Fig. 3. Results for the dense graph

Case study I
Limitations

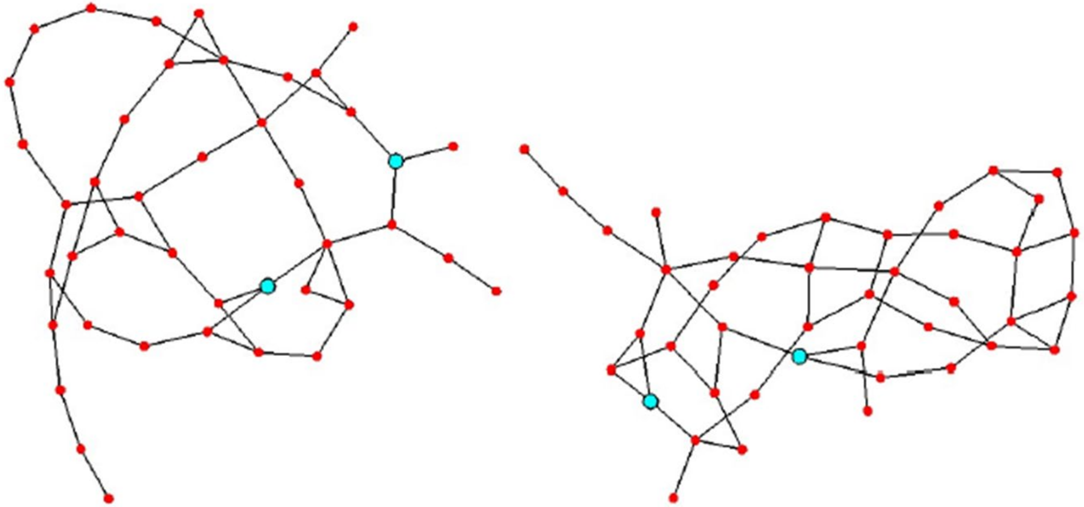
- Graphs are difficult to control
  - Elements are interconnected to each other
- Change in one criterion can lead to change in another
- Ex. Making more crossings can also make more sharp turns on the path
  - Gestalt principle of continuity





Case Study II

Case study II: An experiment without manipulation/interaction



Ware et al. (2003) Cognitive measurements of graph aesthetics. Information visualization. 1(2), 103-110.

Case study II
Design and analysis

- Create a set of random graphs, drawn with a spring algorithm
- Highlight two nodes, record the measurements of the predictor variables
  - Continuity, number of crossings, crossing angles, number of branches, shortest path length….
- Record the response variable
  - Response time
- Regress response variable on predictor variables to detect their relationships.
- Task: shortest path between two highlighted nodes
- Analysis: Correlation and multiple regression

- Finding: Path continuity and number of crossing on the shortest path are important factors.

Case study II

Design and analysis

- the correct value for the shortest path length (**spl**): 3, 4 or 5
- the continuity ('bendiness') of the shortest path (**con**): measured in degrees.
- the number of crossing edges on the shortest path (**cr**)
- the cosine of the angle at which each crossing edge crossed the shortest path. The purpose of this was to allow us to weight shallow angle crossings higher than orthogonal crossings. From this we computed the average cosine crossing angle (**aca**)
- the total number of crossed edges in the graph drawing (**tcr**)
- the number of edges branching from nodes along the shortest path (**br**)
- average geometric line length along the shortest path, per edge (**all**). This was computed in arbitrary units. To get centimeters it is necessary to multiply by 1.4.
- total geometric line length of the shortest path (**tll**): same units as **all**

Case Study III

Case study III: Beyond time and error

- Where the time is spent and how the performance is impacted?
- What is the mechanism of crossings affecting performance?

- Time and error performance logging
  - Treat the human as a “black box”, which tells us what but not how and why

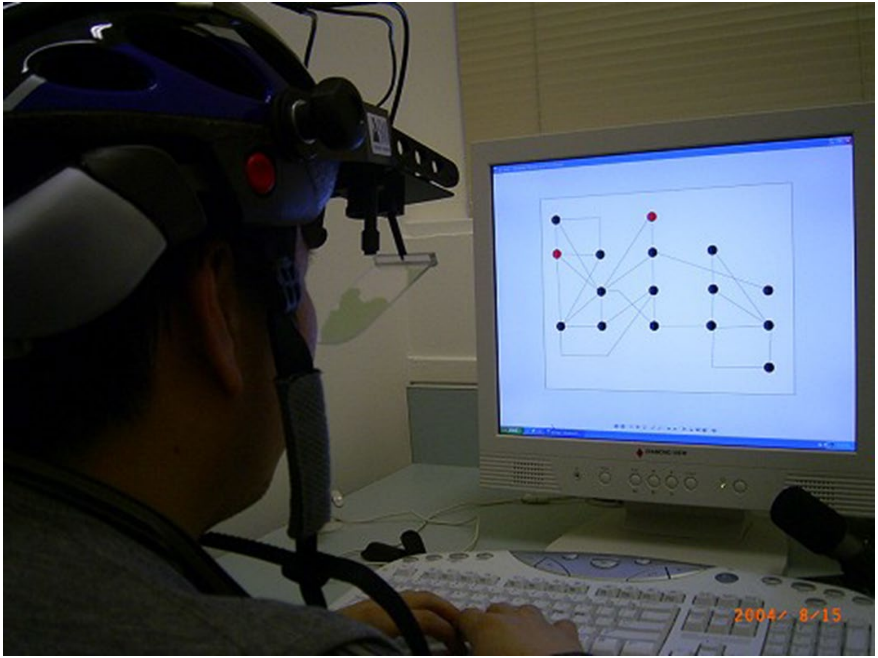
- Eye tracking may give insight as to how
- Post-interview and questionnaire tell us why

Case study III: Beyond time and error (cont.)

- Tow exploratory eye tracking experiments
  - Ex1 : small and sparse graphs
  - Ex2 : large and dense graphs

- Three confirmatory controlled experiments
  - Ex3a : existence of geodesic-path tendency
  - Ex3b : effects of geodesic-path tendency

Eye tracker



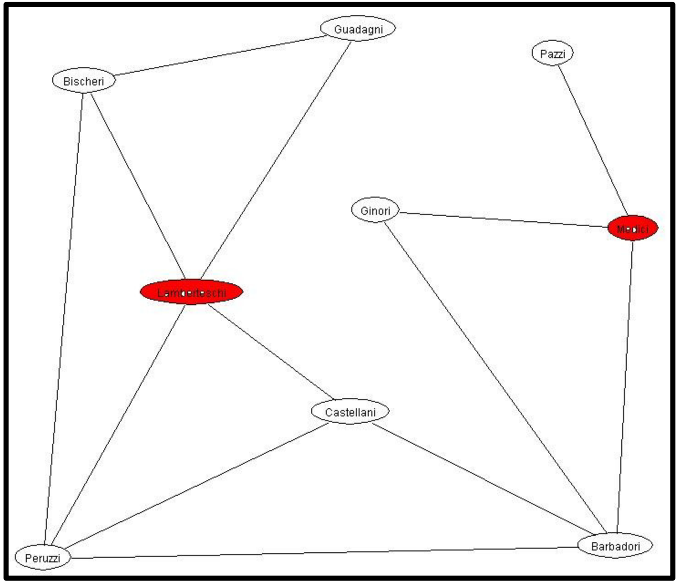
Case study III

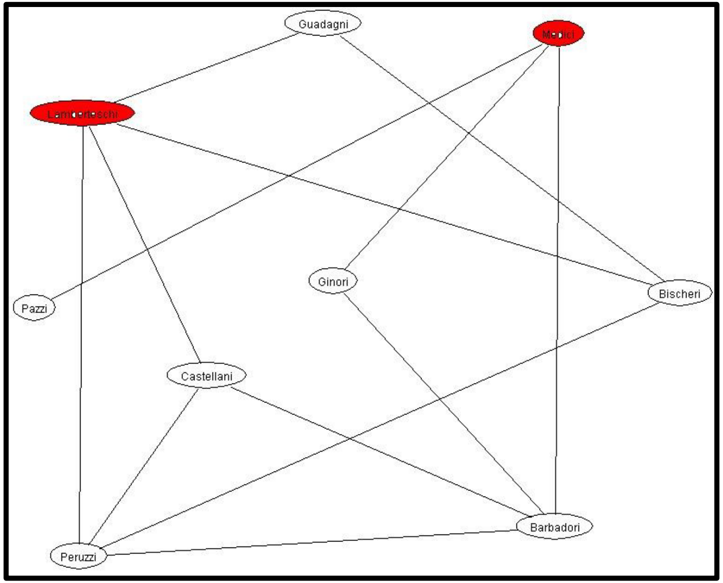
Experiment I – Description

- Task: find the shortest path between two highlighted nodes.
- Time, error and eye movements were recorded.
- Questionnaires and interviews.

Case study III

Experiment I – Example of Stimuli





Case study III

Experiment I – Results of Time and Error

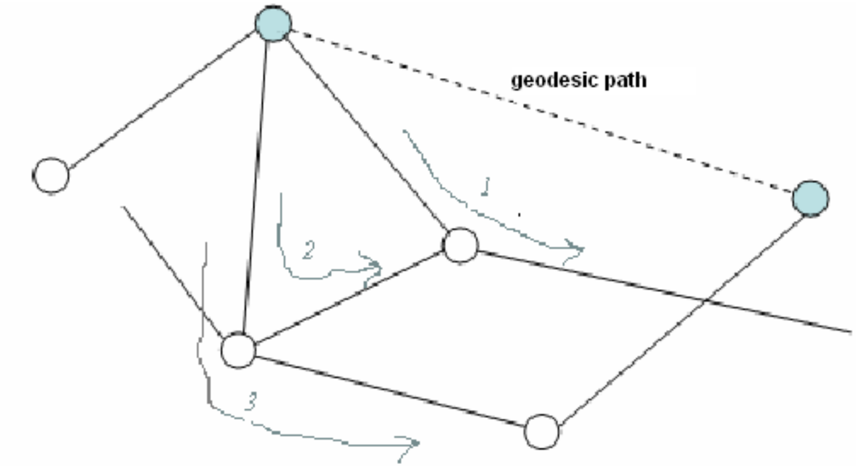
- Overall, subjects spent significantly more time with crossing drawings than with non-crossings

- However, on some specific instances, this was not the case

Case study III

Experiment I – Results of Eye Tracking

- Crossings had little impact on eye movements.
- Geodesic-path tendency: subjects seemed to follow the geodesic path between the current node and target node



Case study III

Possible reasons for the lack of crossing effect

- Crossing angles may inhibit readability (Ware et al., 2003).

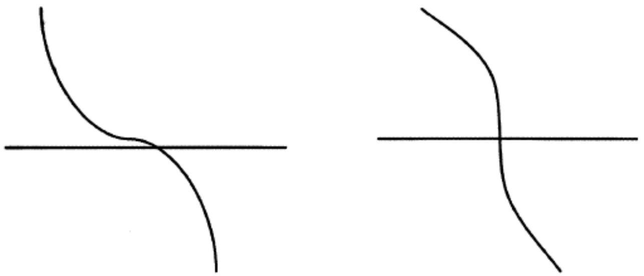


Figure 1 The pattern on the left (a) is perceived as a curved line overlapping a rectangle (b) rather than as shown in (c).

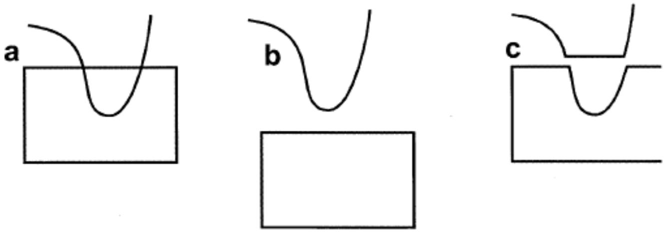


Figure 2 The coarse orientation tuning of edge detectors in the brain suggests that lines that cross at an acute angle as shown on the left are more likely to be confusion than lines that cross nearly at 90° as shown on the right.

Case study III

Experiment II – Description

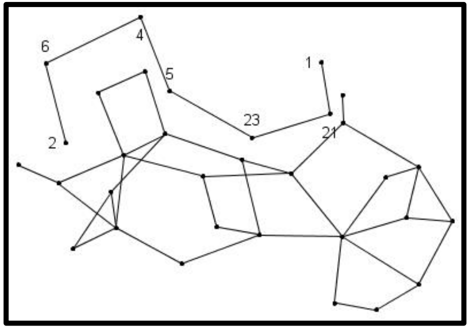
Crossing angle

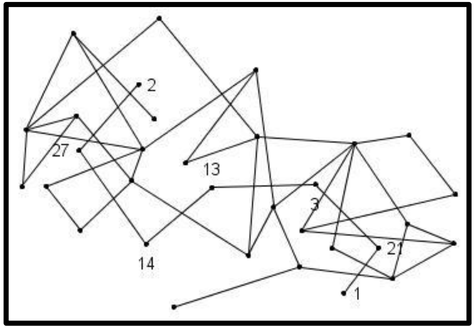
Graphs were drawn with three conditions:

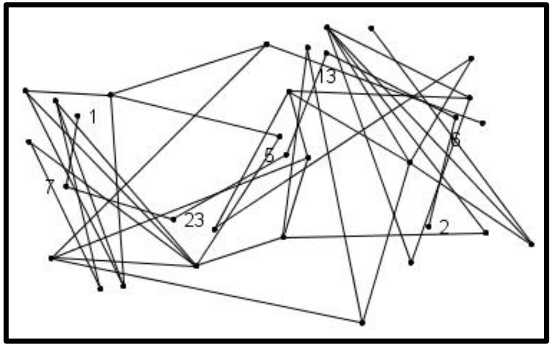
- No crossings on the path
- Small-angle crossings
- Large-angle crossings

Case study III

Experiment II – Example of Stimuli

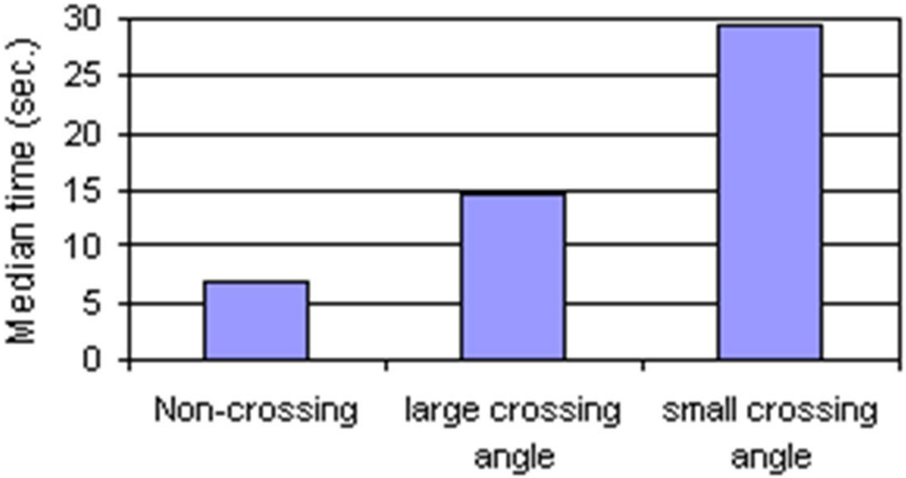






Case study III

Experiment II – Results



- Effects of crossing angles were significant

Case study III

Experiment II – Results of Eye Tracking

- No crossings: eye movements were smooth and fast.
- Large crossing angle: eye movements were still smooth, but slower.
- Small crossing angle: eye movements were very slow and no longer smooth (back-forth moves at crossing points).

Summary

To make sure that visualization is effective, we need

- Principles and theories to guide the visualization design
- Once visualizations are produced, we need proper methods and measurements to evaluate them

THE UNIVERSITY OF
SYDNEY