Visualisation of Abstract Information

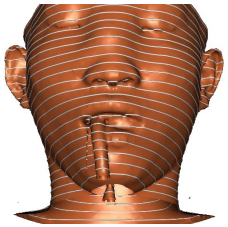
Visualisation – Lecture 17

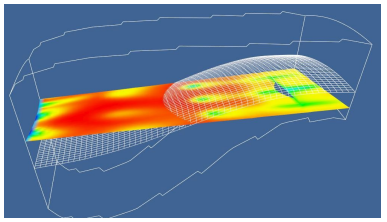
Taku Komura

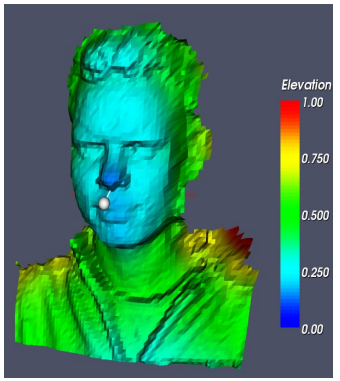
Institute for Perception, Action & Behaviour
School of Informatics

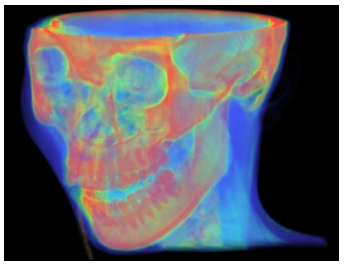
Information Visualisation

- Previously – **data with inherent topology**
  - 2D / 3D datasets
  - multi-dimensional information









- Today : **Abstract Information**
  - constructing topology in otherwise disjoint data measurements for visualisation
  - **information visualisation**

What is Information Visualisation ?

- visualising ***discrete*** **data with no topology**

- Information with no obvious visual representation
  - events in an O/S, financial records, documents, medical records
  - **'Visualising the non-visual'**
  - **'Visual data mining'**

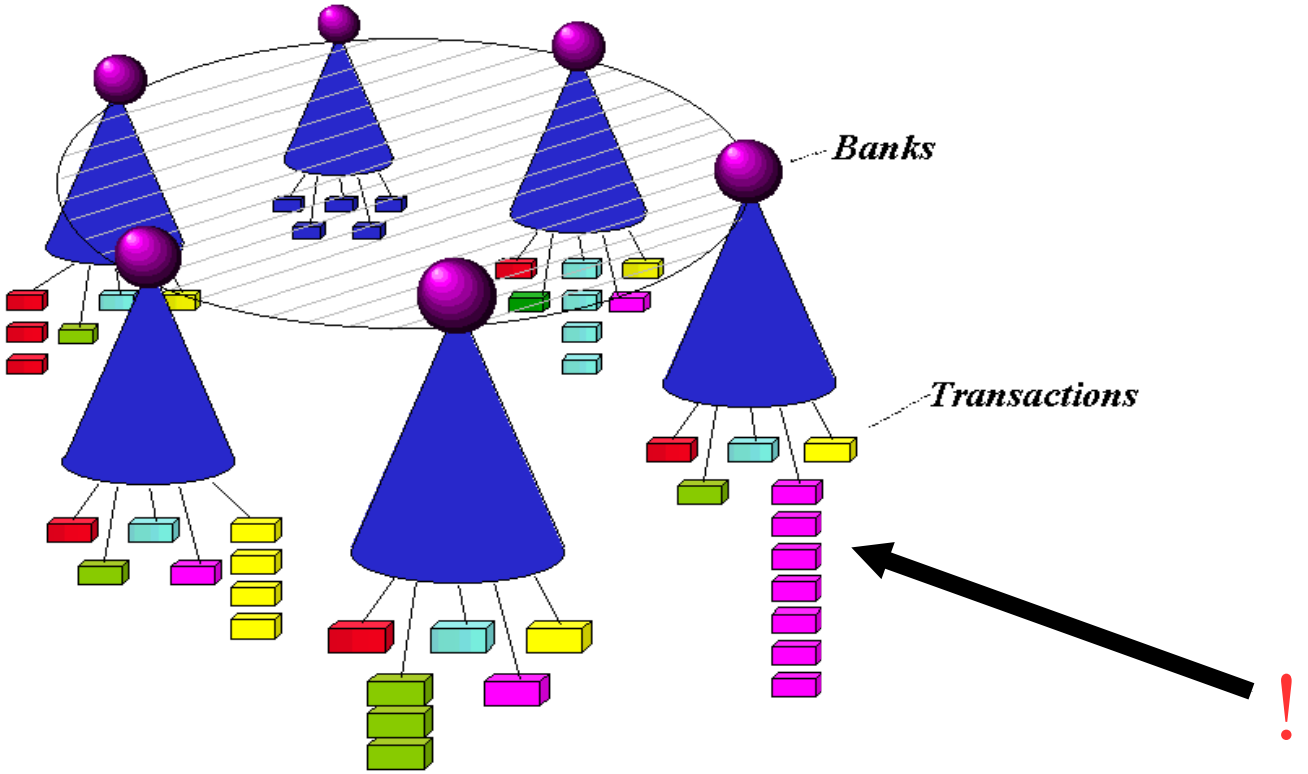
- Interest in data routinely collected and archived:
  - Consumer 'store loyalty cards' *(who buys what, when ?)*
  - Mobile communication *(who call who, when ?)*
  - Internet crime *(firewall logs – who attacks who, when?)*

Visualisation of Connectivity

- Essentially identifying for **patterns in connections between multiple objects**

- **Example : system failures**
  - Usually multiple causes, or a combination of events
  - Identify for patterns
- **Example : Fraud**
  - Identify for anomalous or unusual behaviour
  - Connections between fraudsters
  - *humans poor at numerical analysis, use visual analysis*

Example : fraud detection



- Raw information on transaction visualised in a graphical format
  - anomalies easily apparent

Connectivity : netmap

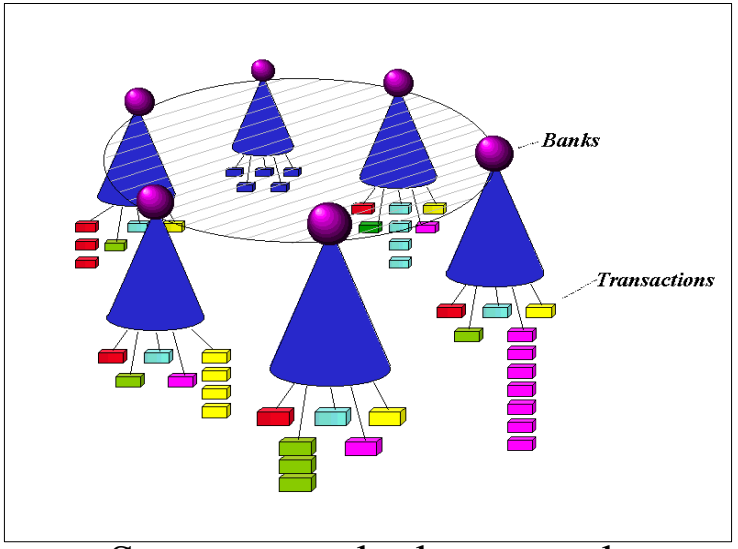
- Popular tool for **information visualisation**
- **Concept:** display data items around a 2D circle with *n* attributes displayed next to each other
  - draw dissecting lines to indicate connections between attributes
  - lines form **graph indicating strength and correlations of connections**
    - represent **relationships in the attributes of the information**

Example : insurance data

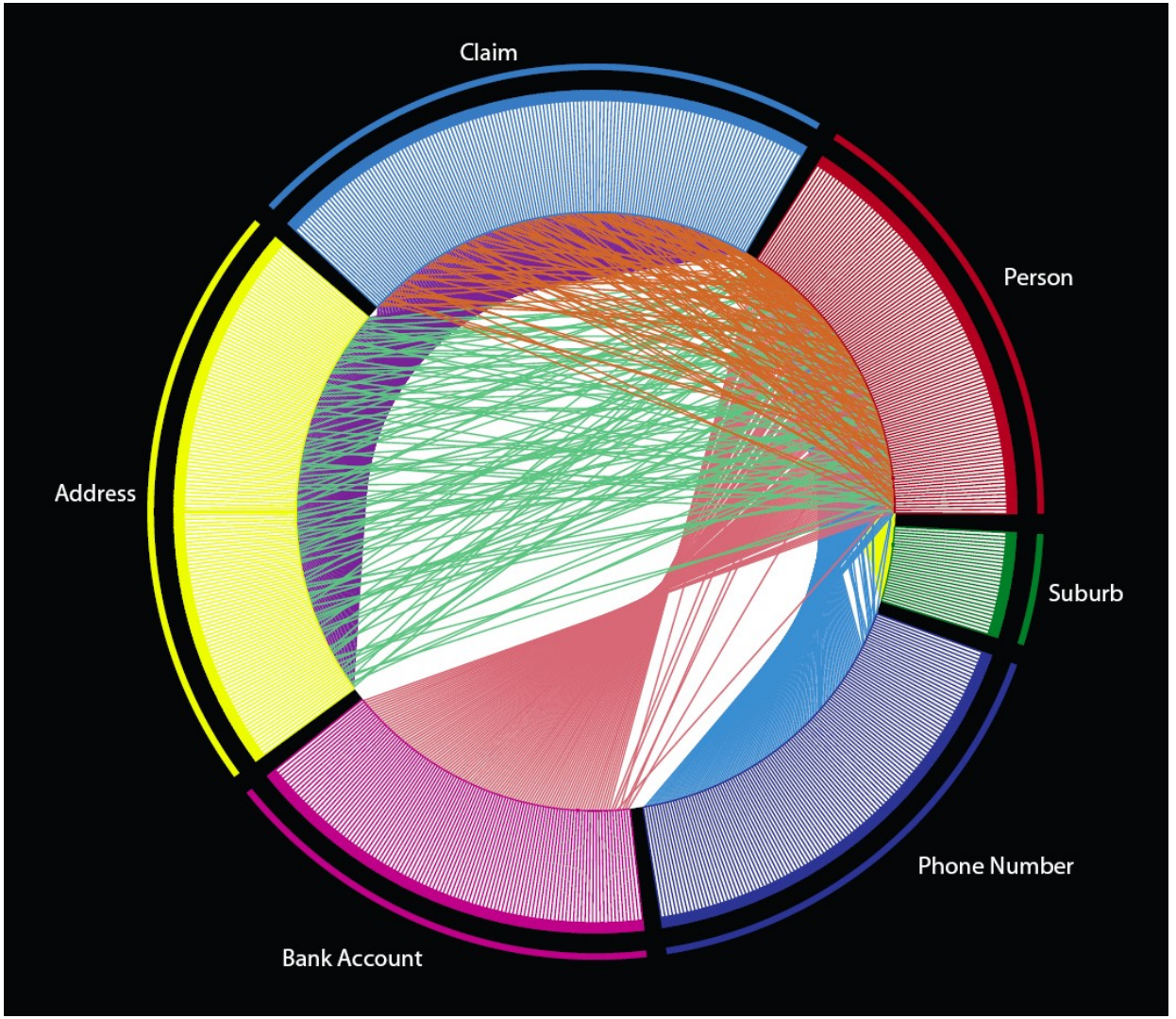
Detecting insurance fraud

12 months insurance claim data

http://www.netmapanalytics.com/technical/Fraud_Crime_Demo.pdf.



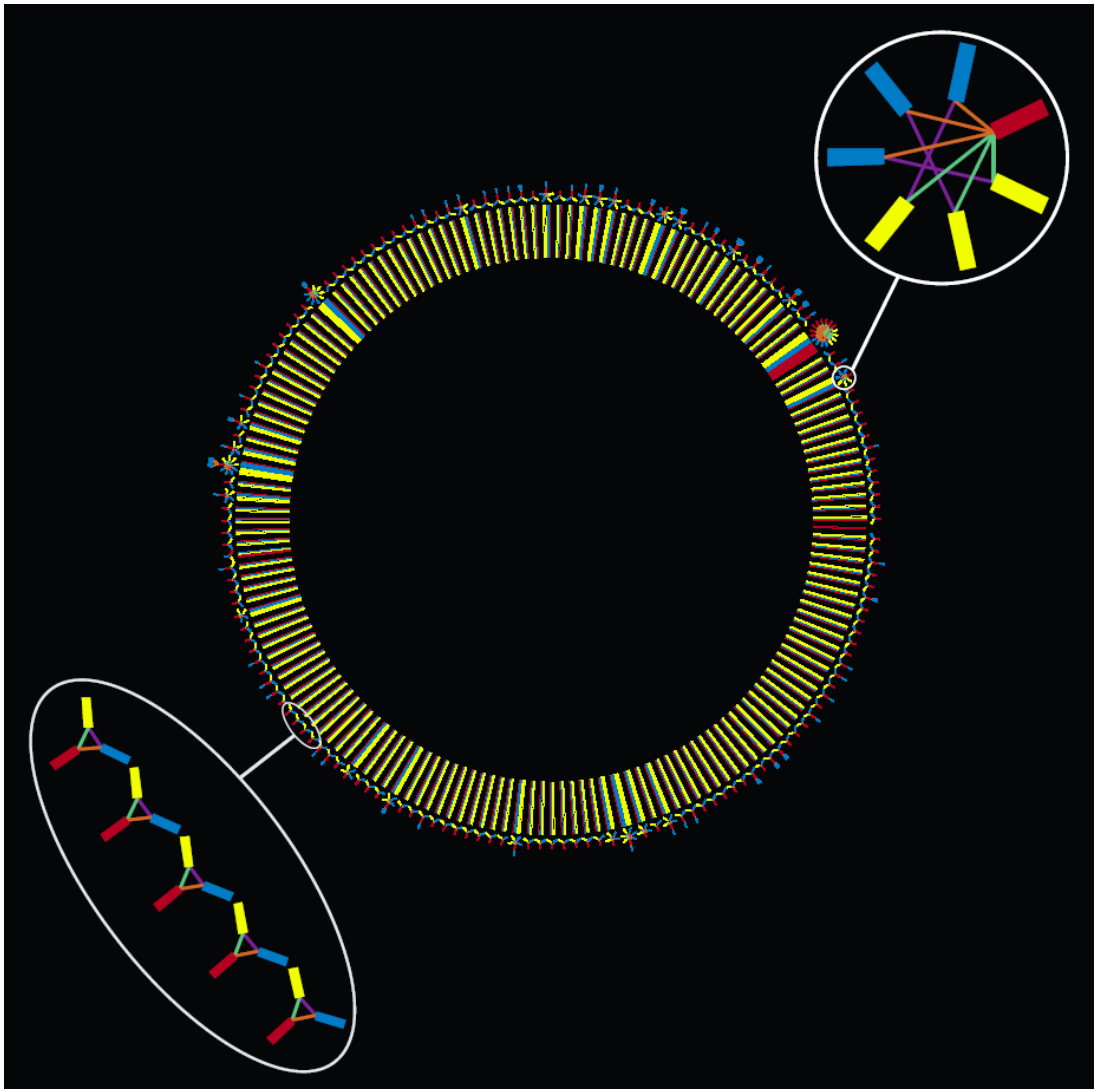
Same example, larger scale.



Example : insurance data

The results of a clustering algorithm.

A normal cluster links 1 person, address and claim.

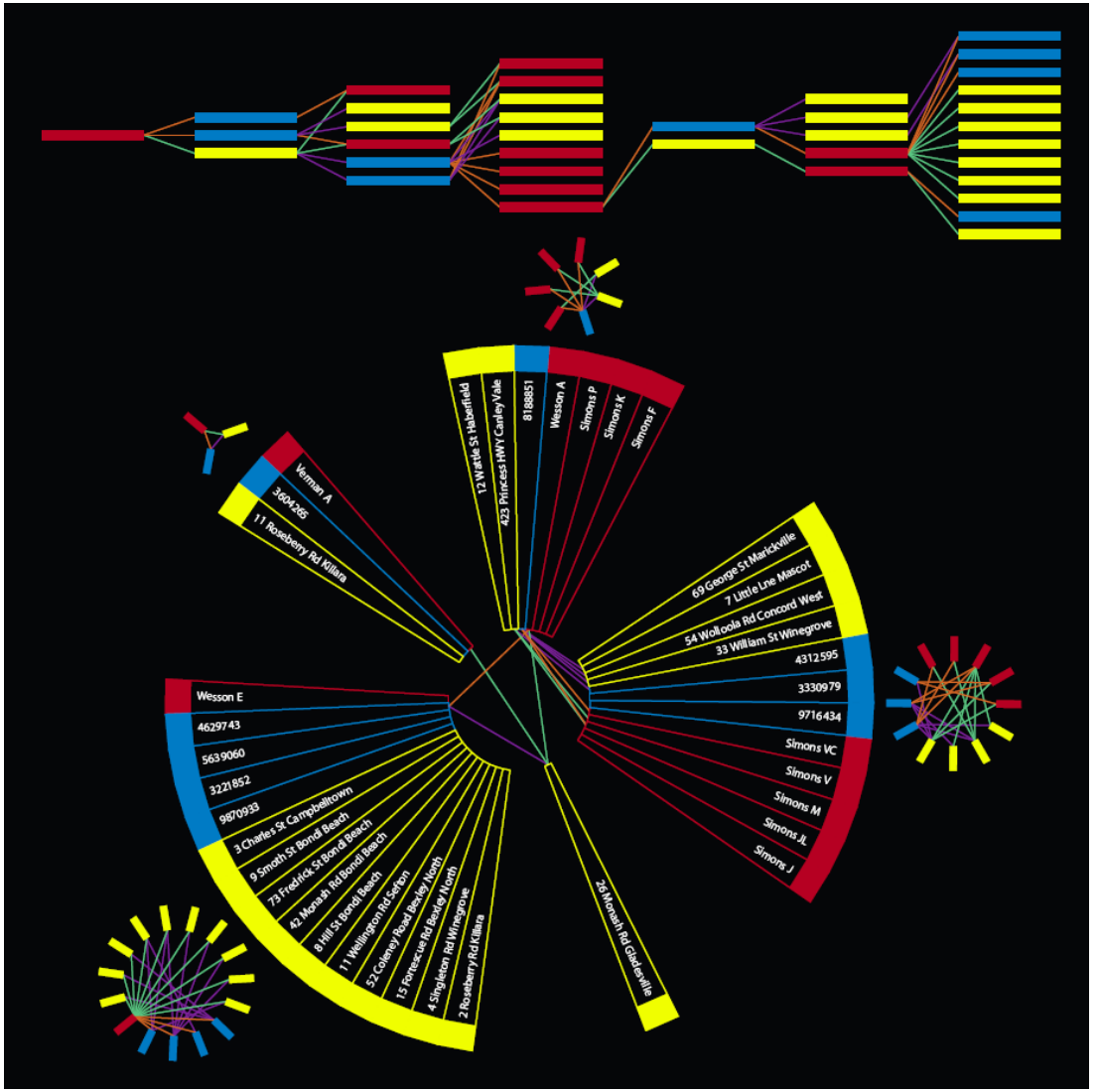


Example : insurance data

Visualising a ‘cluster in linear form (top).

6 steps of connectivity, 10 people linked (no third parties like lawyers).

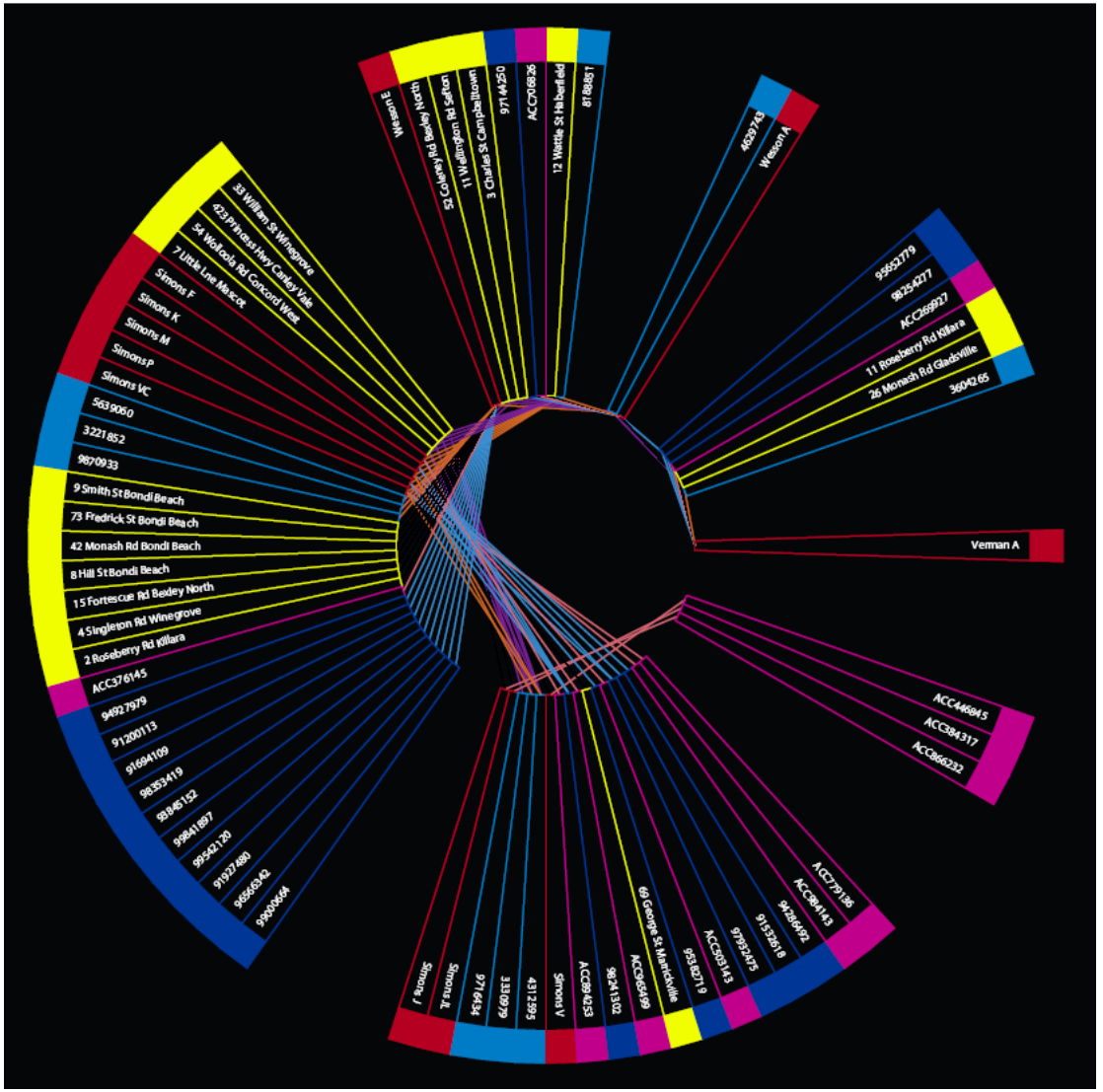
Links between clusters at bottom.



Example : insurance data

All the information relating to ‘Mr Verman’ is displayed anti-clockwise and in clusters – **shows one person linked to multiple claims!**

The real ‘Mr Verman’ is in jail…



Success Stories

Tracking the Backpacker Murderer

- In the early 1990s seven young backpackers were murdered in what was to become Australia's most notorious serial murder case.
- The police had developed a profile of the killer. However, to track their suspect down, they faced an enormous volume of data from numerous different sources.
- Investigators therefore applied NetMapping technology to RTA vehicle records, gym memberships, gun licensing and internal police records.
- As a result, the list of suspects was progressively narrowed from extensive list of individuals to a short list of 230, and then a still shorter list of 32, which included the killer.
- Thanks to NetMap, thousands of precious police hours were saved and police were able to focus their investigations on a more manageable list of potential suspects, leading to the eventual successful conviction of the Backpacker Murderer.

Success Stories

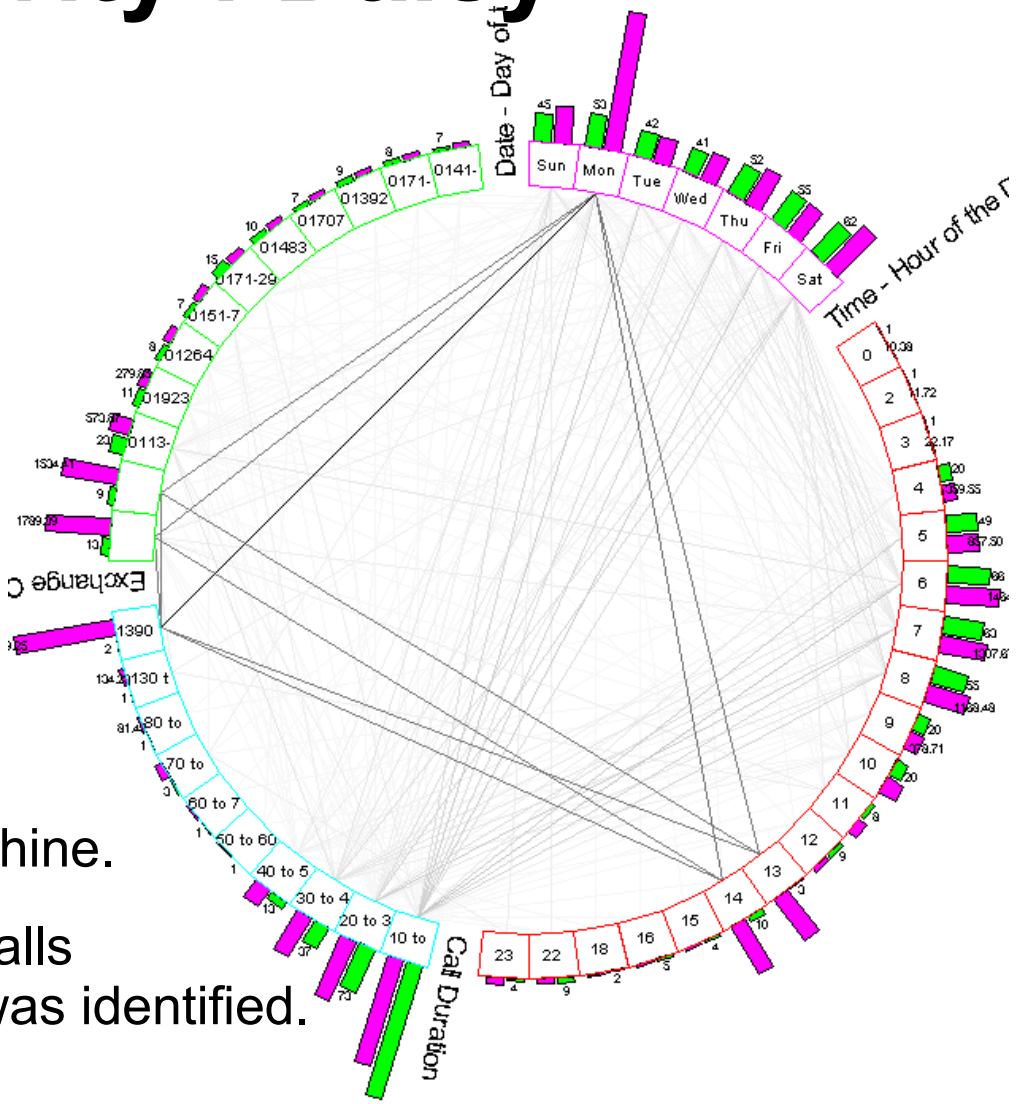
The Mystery TNT Options Trader

- The Australian Securities and Investment Commission (ASIC) has the unenviable responsibility of regulating the many millions of transactions that flow across the Australian Stock Exchange.
- For some years now, ASIC has been a regular user of NetMap technology to successfully detect irregularities.
- One of the most high profile examples of this took place in August 1996 when investigators enlisted NetMapping technology to help them track down a mystery TNT options trader who had become a millionaire, just two weeks after purchasing options.
- The criminal had cleverly hidden behind multiple layers of transactions, false identities and third party bank accounts. However, with the help of NetMap, ASIC was able to unravel this complex trail, leading to the discovery and arrest of a Macquarie Bank Director, who was successfully convicted following his 1999 appeal

Connectivity : Daisy

- Variant on netmap concept
  - **Displays quantities as well in a bargraph form.**

www.daisy2000.com



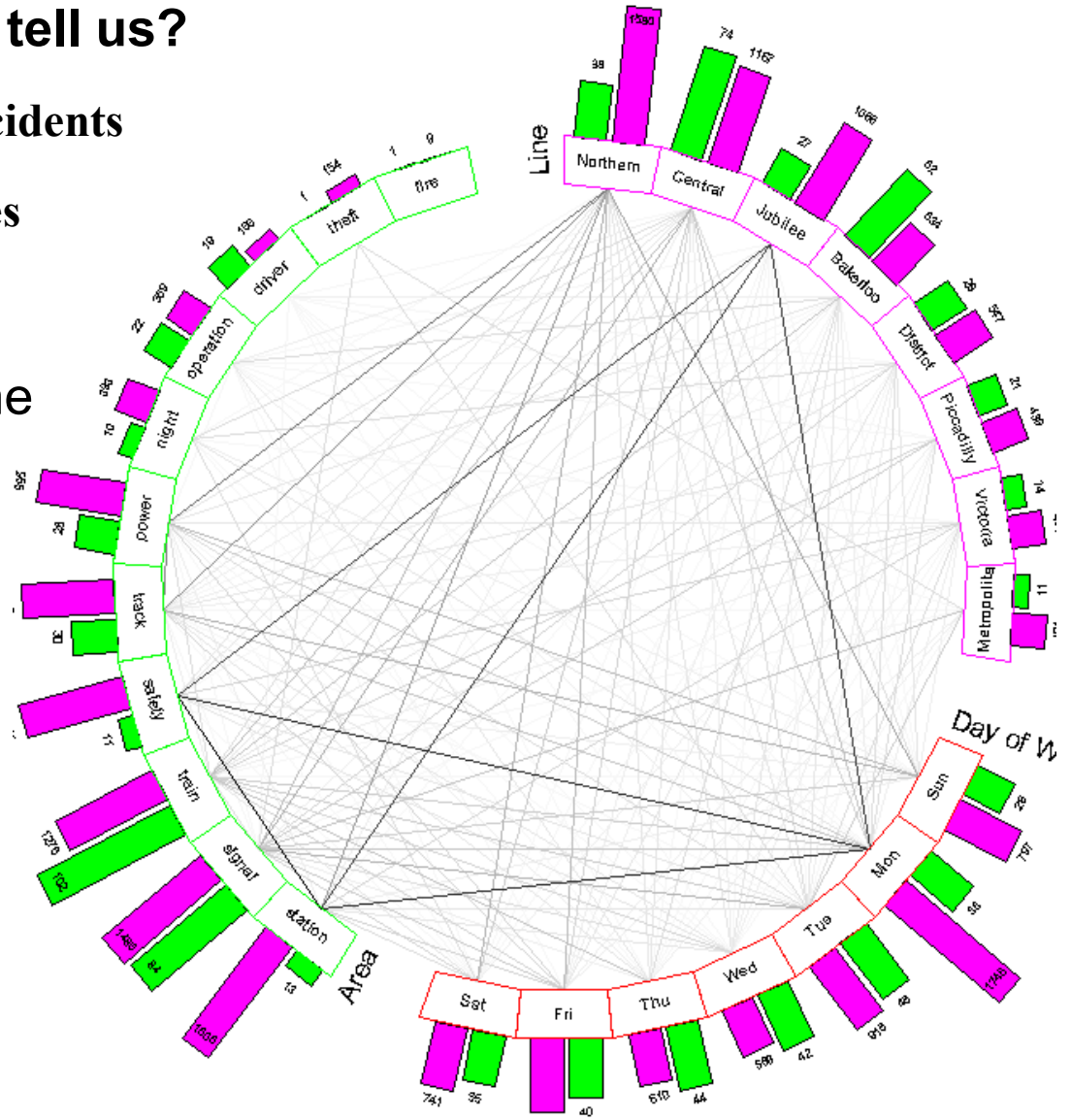
- Visualisation of calls made by a fax machine.
  - Triangle indicates large number of calls which timed out. As a result a fault was identified.

Connectivity : Daisy

- **What does the visualisation tell us?**
- **the green shows the Number of Incidents**
- **the cyan the Total Delay in Minutes**

- Many delays on the Jubilee line on Mondays – this was due to a station being upgraded.

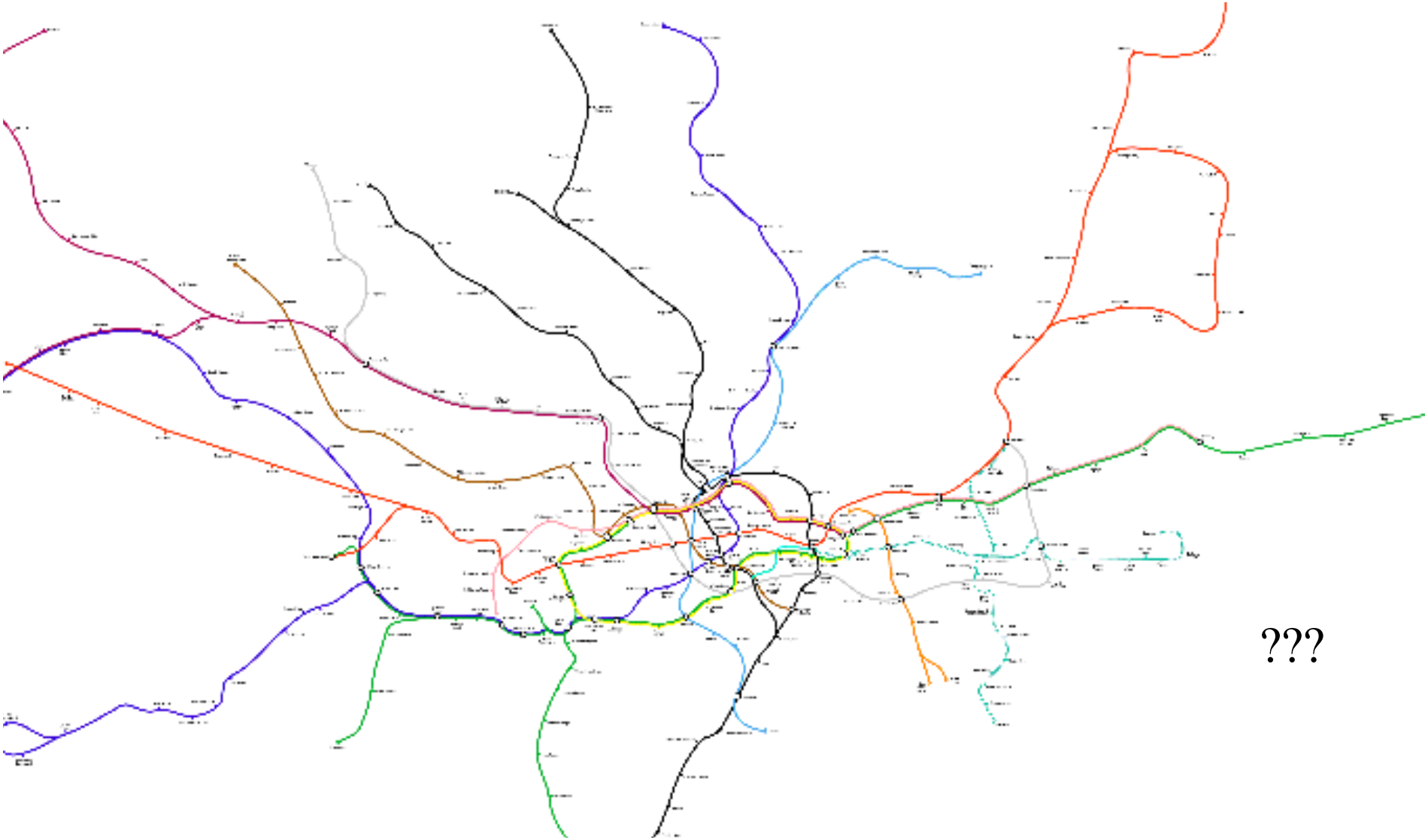
- Most delays are essentially random
- Northern line is particularly bad



Visualising Connectivity

???

Visualising Connectivity



???

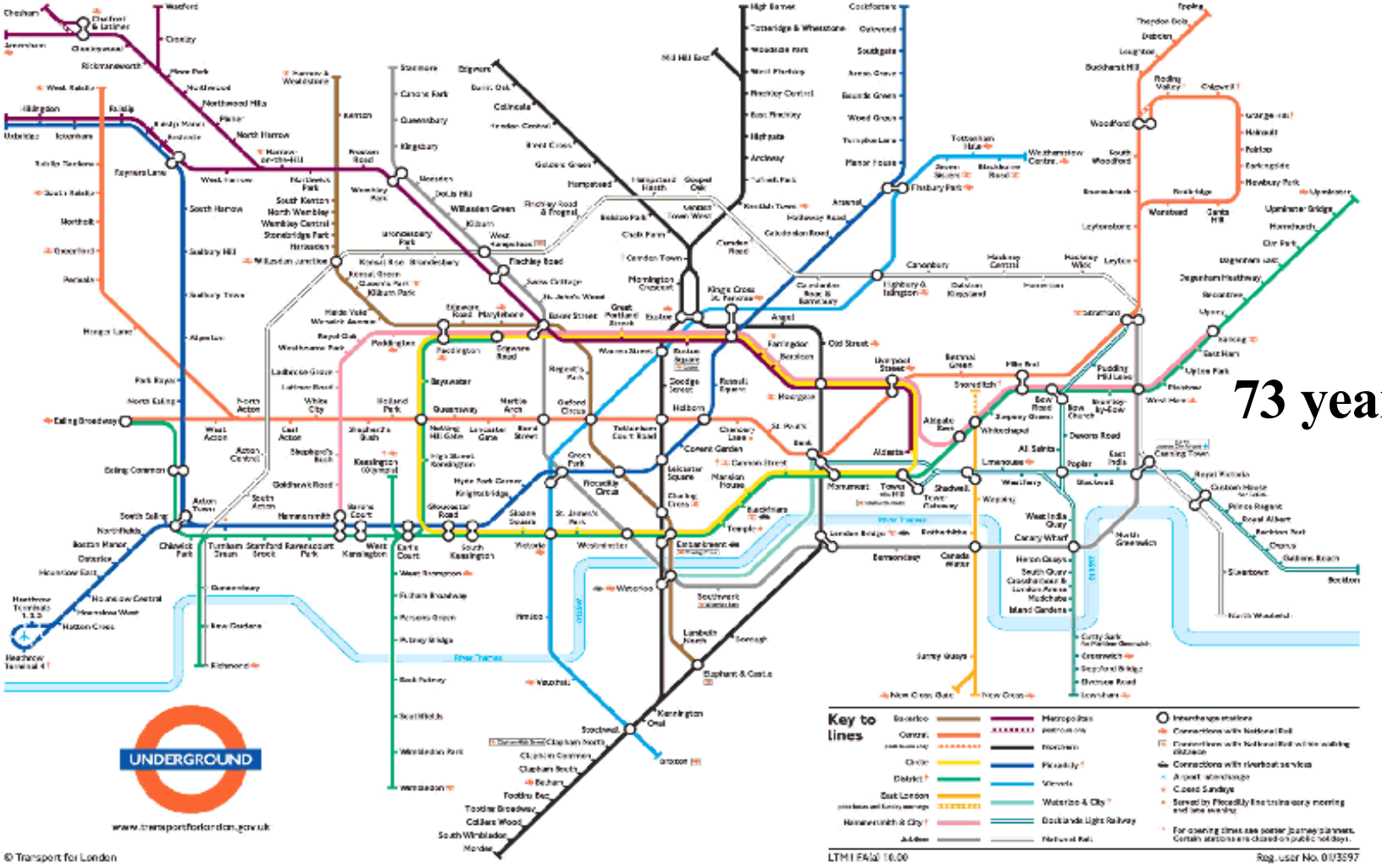
Visualising Connectivity

1933 London Underground Map

N.B. Beck used colour mapping in 1933!

- Here the **connectivity** is important to the user
  - design : H. Beck 1933 (electrical draftsman)

Visualising Connectivity



- **Beck's design** : topology of the data preserved but geographically inaccurate
  - **topology represents the connectivity** (thus it is maintained in the design)

Document Visualisation

- Motivation:

Action	Units of Information transfer
Typing at 10 bytes per second	1
Mouse Operations.	2
Reading	3-40
Hearing	60
Visualisation and Pattern Recognition	12,500

  - visualisation is considerably faster than hearing / reading

Source :
Silicon Graphics Inc.

Visualisation of Documents

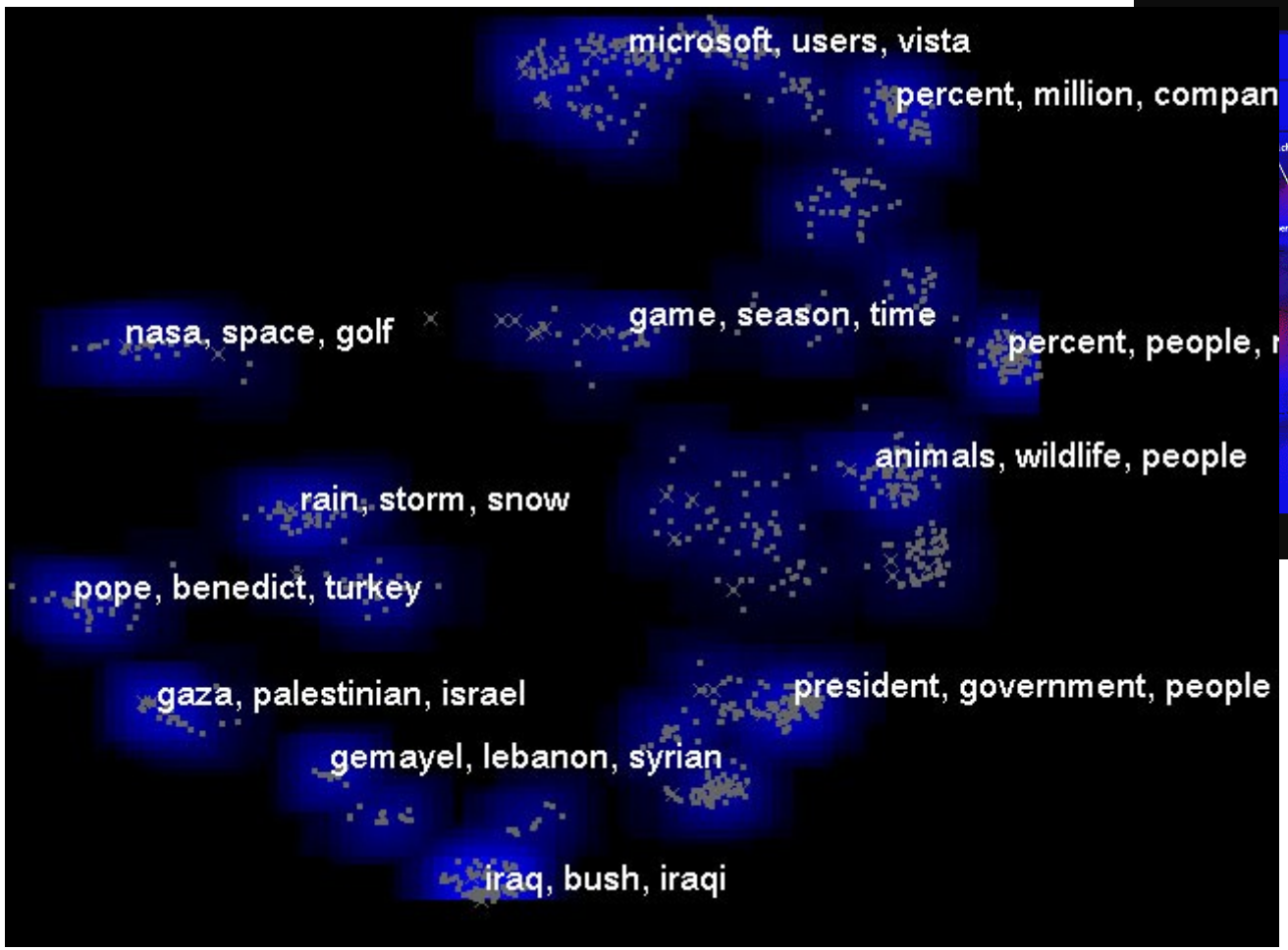
- **Motivation : large bandwidth of human visual system**
  - 100s millions of documents available on-line
  - information only in textual form
- **'Visualising the non-visual'**
  - **searching for scientific papers**
  - analysing witness statements
  - awareness of events in news bulletins
- **Concerned with searching and visualising occurrences of query words**

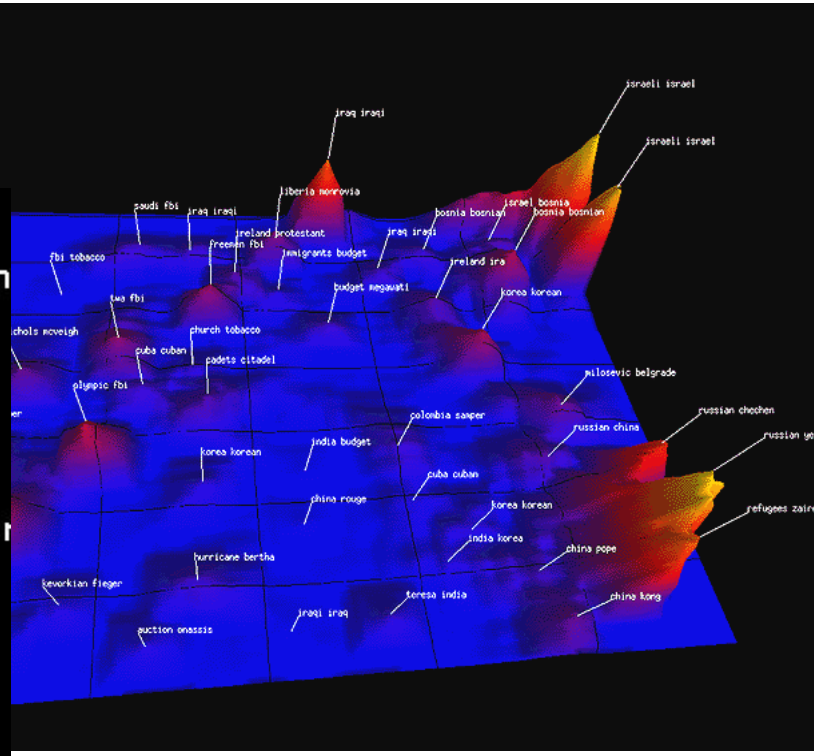
Document Visualisation - Stages

- **Query**
  - "keywords" from user specification
  - comparison to sample "reference" document
- **Representation** of results
  - form high-dimensional vector (one for each word, ~10000+)
  - cluster documents based on vector similarity (e.g. Nearest-Neighbour)
- **Visualisation** of clustered results
  - projection to lower dimensional space
  - 2D "galaxy" / "theme-scape" / 1D "theme-river"

2D and 3D projections of documents

Closely related documents cluster together while unrelated documents are further apart





Articles in a collection of news items (2D).

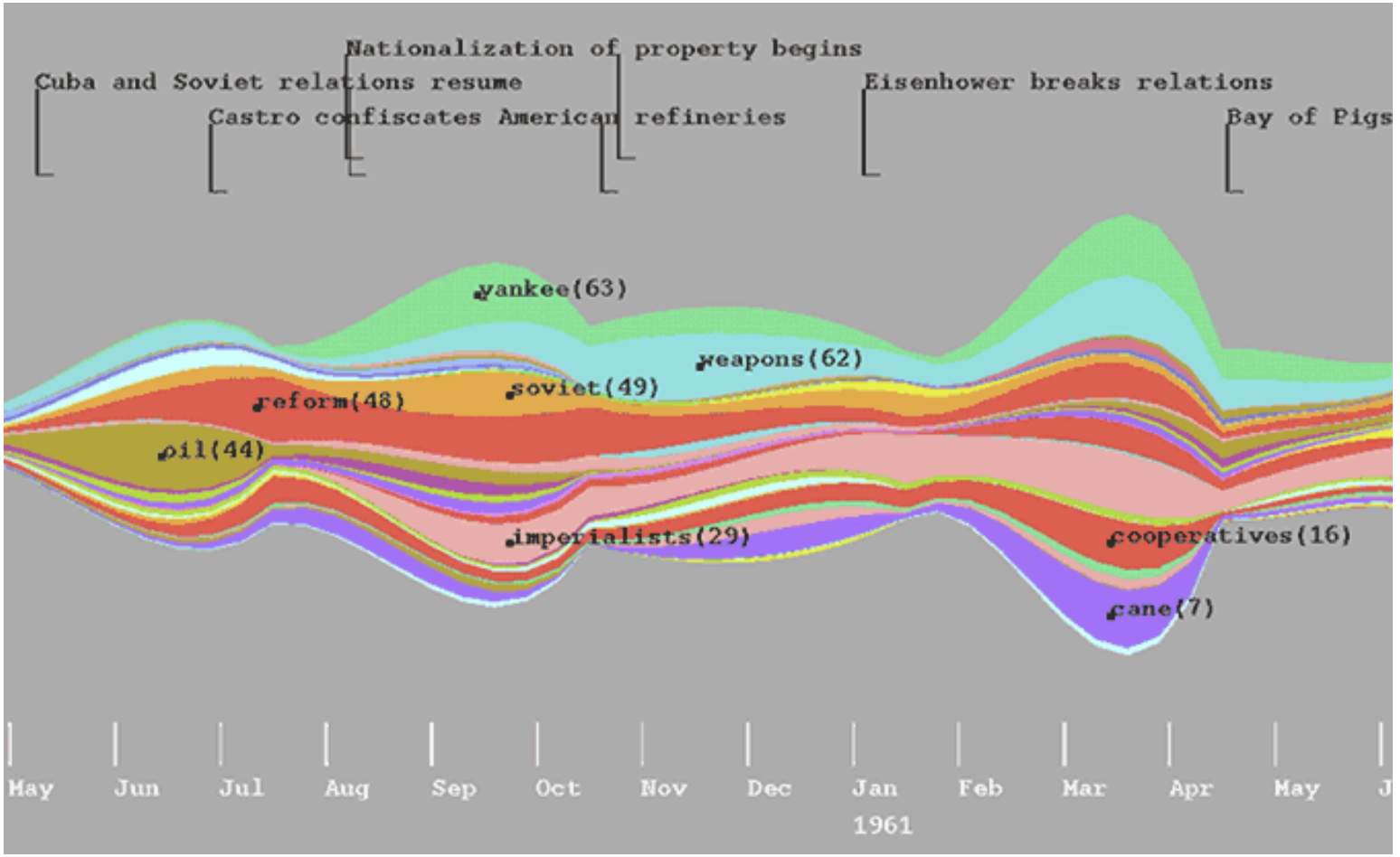
http://in-spire.pnl.gov/

Pacific Northwest National Laboratory.

1D visualisation of news articles

A ‘Theme River’ shows the relative importance of themes over the course of a year from press articles.

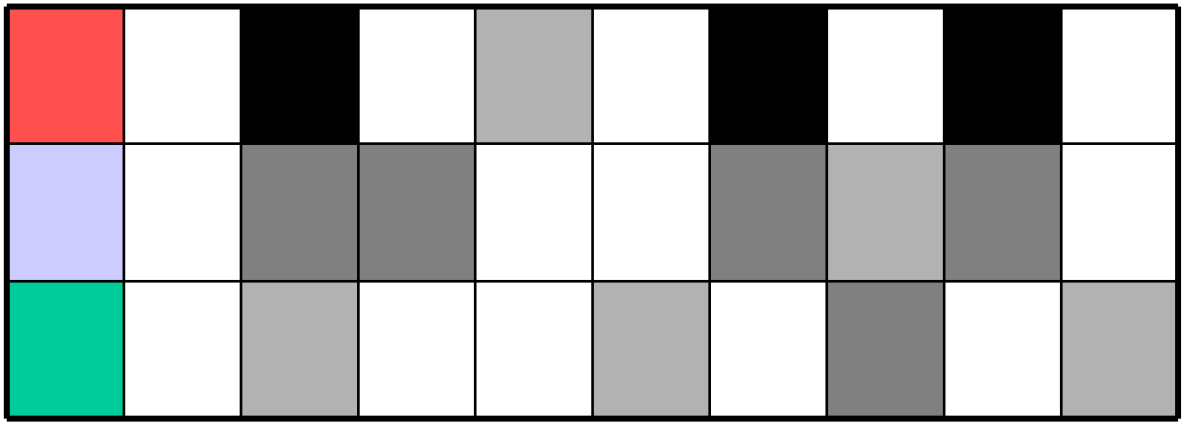
Pacific Northwest National Laboratory.



Document Querying

- We are interested in
  - **distribution of keywords** in the document
  - **related articles** to the keyword entered
- **Title bar scheme** (Hearst 1995)
  - display a list of documents with a **title bar**
  - title bar **shows the occurrence of keywords** in document

Title Bar Method



Columns **represent** paragraphs or pages in a document.
Shade indicates relevance shown by word occurrence.

- Visualisation - Use of document topology / colour-mapping

Example : Title Bar Query / Result

Query terms: DBMS (Database Systems)

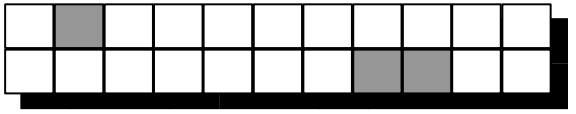
Reliability

What roles do they play in retrieved documents?

Mainly about both DBMS & reliability

Mainly about DBMS, discusses reliability

Mainly about, say, banking, with a subtopic discussion on DBMS/Reliability

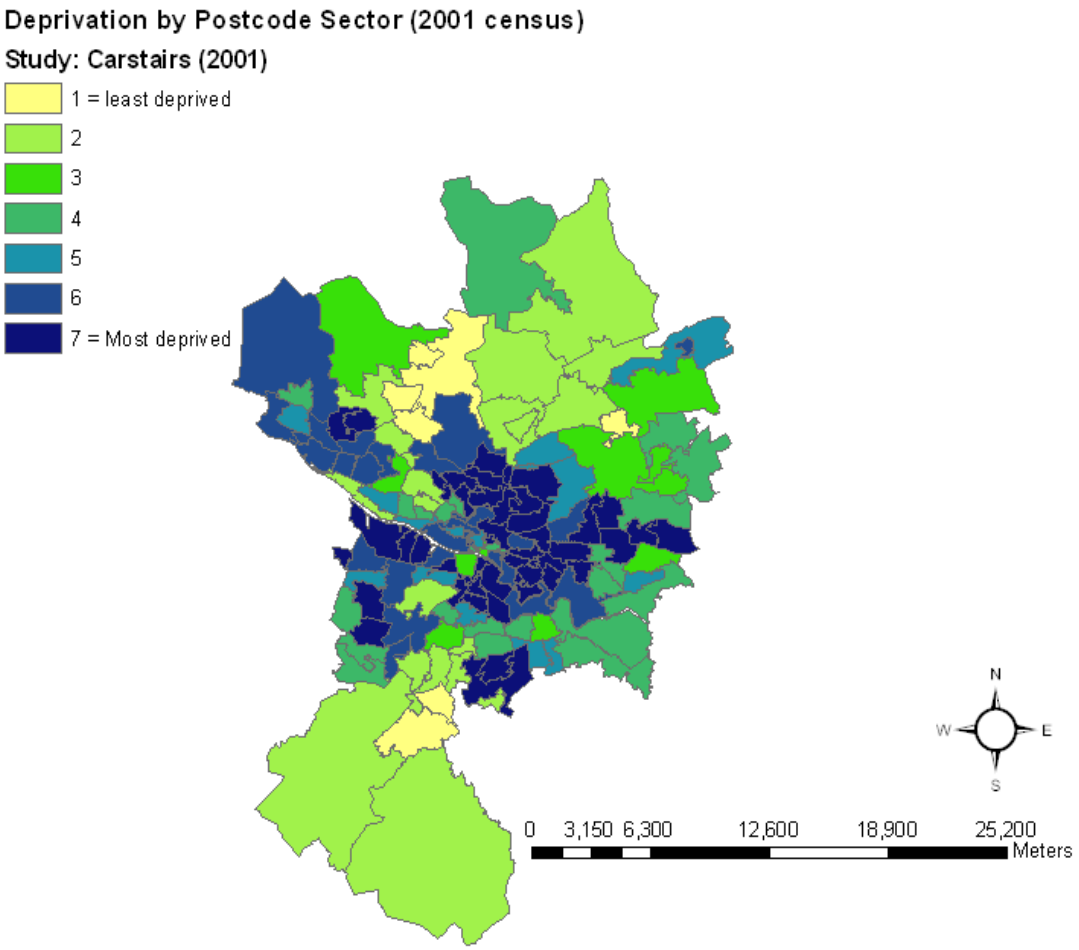


Mainly about something different

GIS : Geo-Spatial Visualisation

- visualisation of information clustered or placed relative to geographic location
- Increasing Area
  - increasing **availability of data & cheap compute power**
  - using fundamentals of computer based visualisation
    - colour-mapping

Spatial Visualisation of Deprivation in Glasgow (Census Data 2001)



Summary

- Visualisation of **Connectivity**
  - identify complex relationships – **"visual data mining"**
- **Document** Visualisation
  - keyword query, **visualisation of document relevance**
- **Geospatial** Visualisation – GIS
  - representation of data with **geographical interpretation**