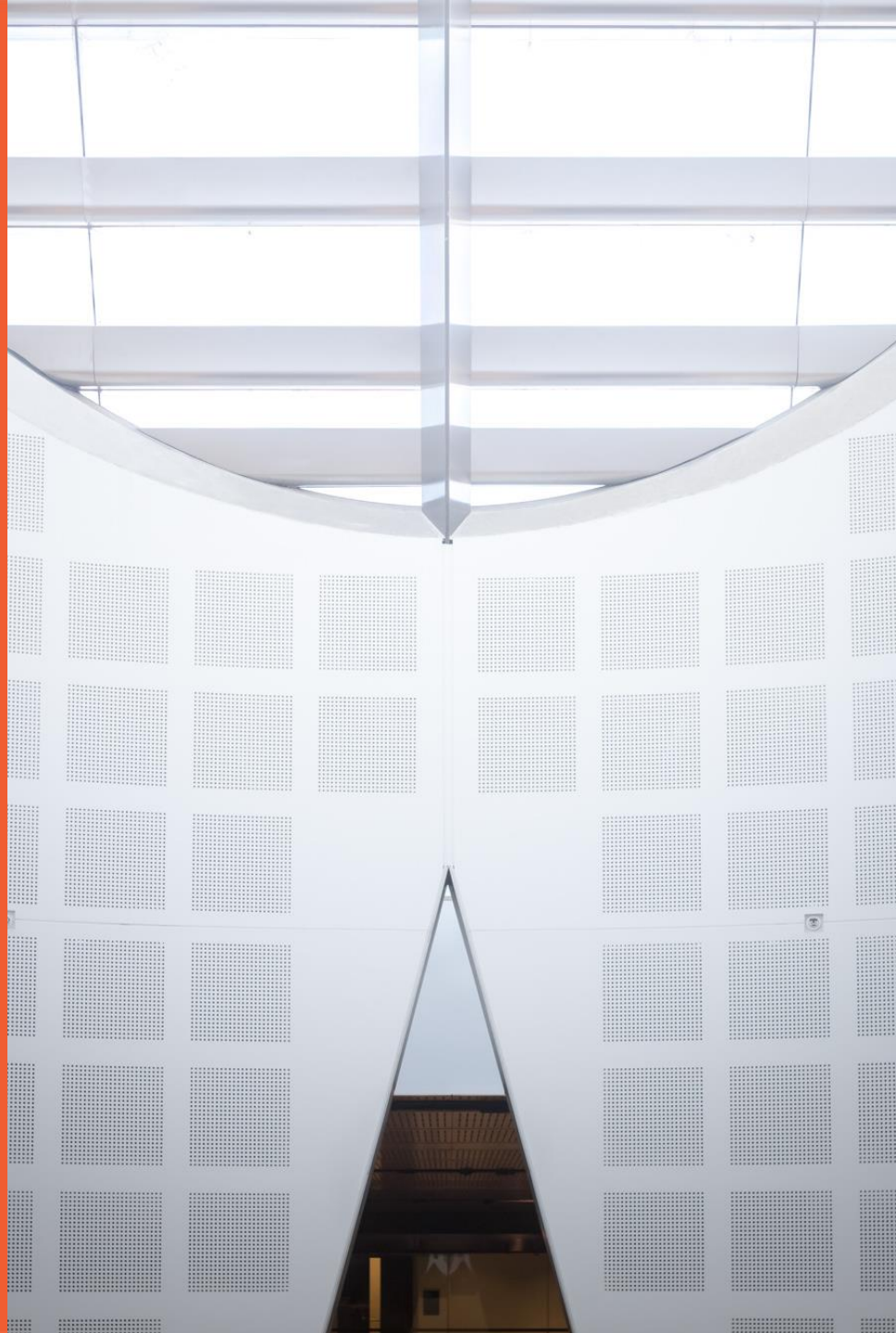
COMP5048
Visual Analytics

Week 6: Visual Analytics of Complex Data 2

Professor Seokhee Hong

School of Information Technologies

THE UNIVERSITY OF SYDNEY



Copyright warning

Visual Analytics

- *the science of analytical reasoning facilitated by interactive visual interfaces*
- Good visualisation can support ***Tasks:***
  - understand the *structure*
  - discover *new knowledge/insight*
  - find *regular/abnormal patterns*
  - generate/confirm/reject *hypothesis*
  - confirm *expected* and discover *unexpected*
  - predict the *future*

Visualisation Challenge

1. Computational complexity

Efficiency

Runtime

We need more efficient algorithms

2. Visual complexity

Effectiveness

Readability

We need better untangling

- http://www.visualcomplexity.com/vc/

Scalability: Main Approaches

Over the past 20 years, there have been several ideas put forward to solve the scale problem:

1. **Cluster the data**
2. **Multi-level approach**
3. **Use 3 dimensions**
4. **Reduce Visual Complexity**
5. **Integration with Analysis**
6. **Integration with Interaction**

Complex Data Visualisation

1. **Multi-dimensional/Multi-variate Data (Table data)**
   - **Multiple attributes: ordinal, nominal, categorical, image etc**

2. **Spatial Data**
   - **Data with geometry (map, longitude/latitude)**

3. **Temporal/Dynamic Data**
   - **Data with time stamps: changing over time**

4. **Data with Constraints (Week 2-4)**
   - **Relations, Hierarchy, Clusters, Directions**

5. **Multi-relational Data**
   - ***Multiple clusters, multiple relationships etc***

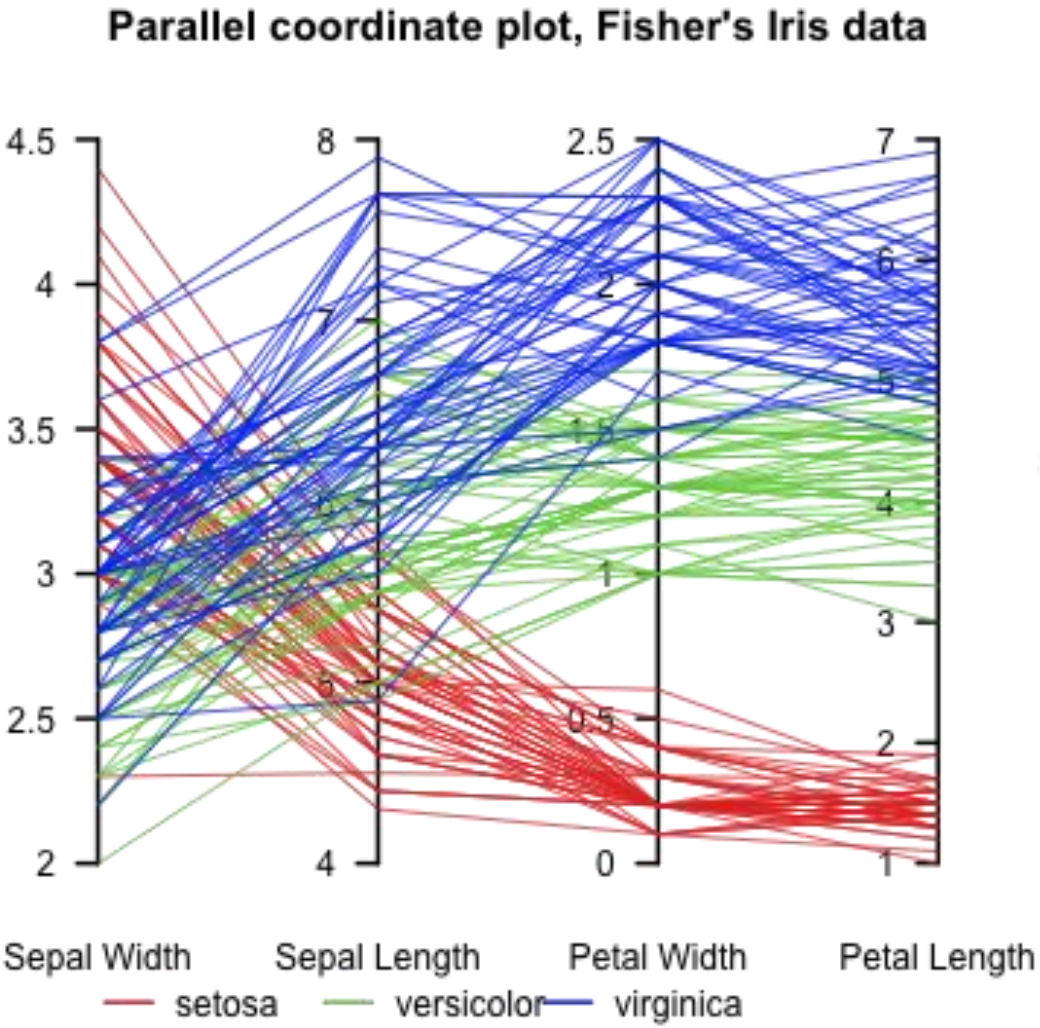
(1) Multi-dimensional/Multi-variate Data Visualisation

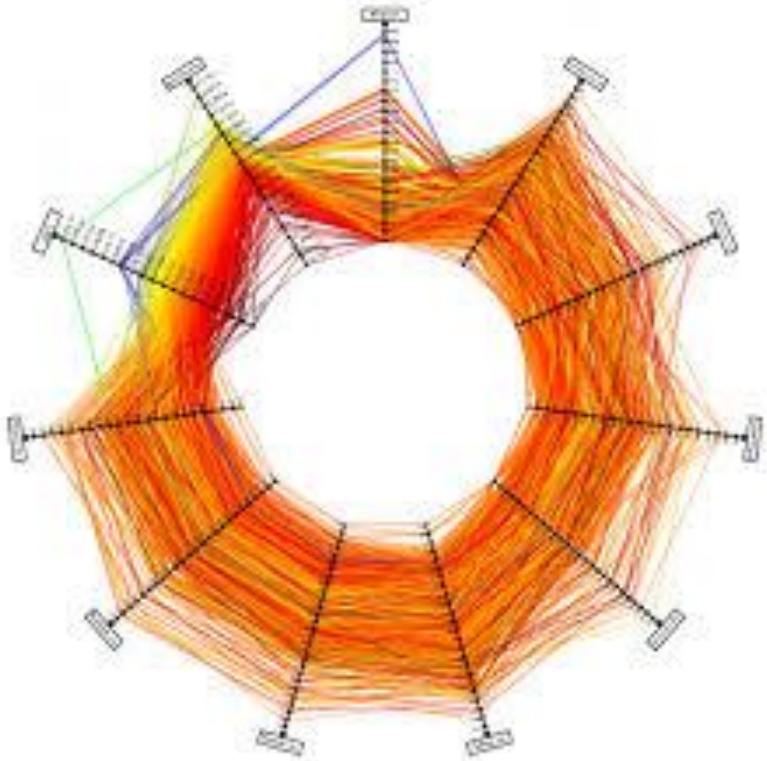
1. Parallrel Coordinates

2. MDS (Multi-Dimensional Scaling)

3. Glyph (Visual variables)

1. Parallel Coordinates [Inselburg]

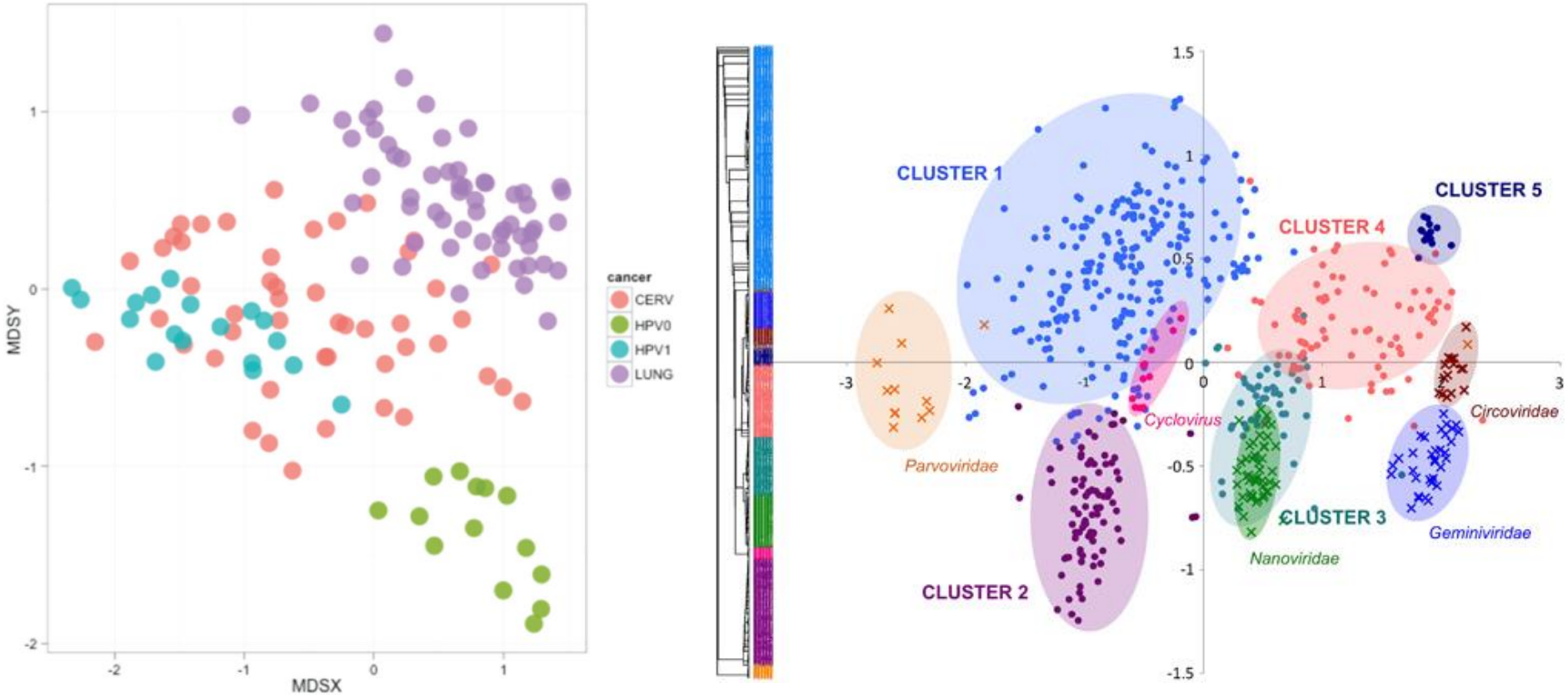




- Nominal values
- Change ordering to see patterns

2. MDS (Multi-Dimensional Scaling)

- Compute similarity between the data using distance matrix (n dimension)
- (e.g) PCA (Principal Component Analysis)
- Visualisation: 2D scatter plot

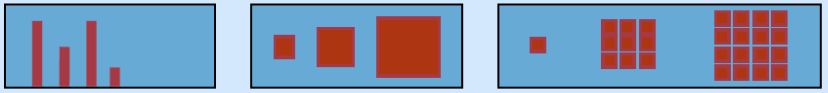


3. Glyph

- ✓ **Data-Ink-Ratio** [Tufte]: minimise ink to represent information
- ✓ **Marks (signs):** line, point, area
- ✓ **Visual Variables** [Bertin]

- **position**
  - changes in the x, y, (z) location

- **size**
  - change in length, area or repetition

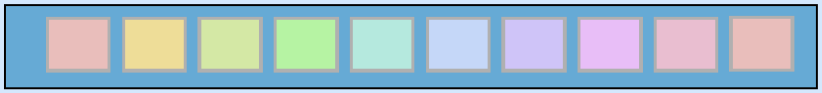


- **shape**
  - infinite number of shapes

- **value**
  - changes from light to dark

- **orientation**
  - changes in alignment

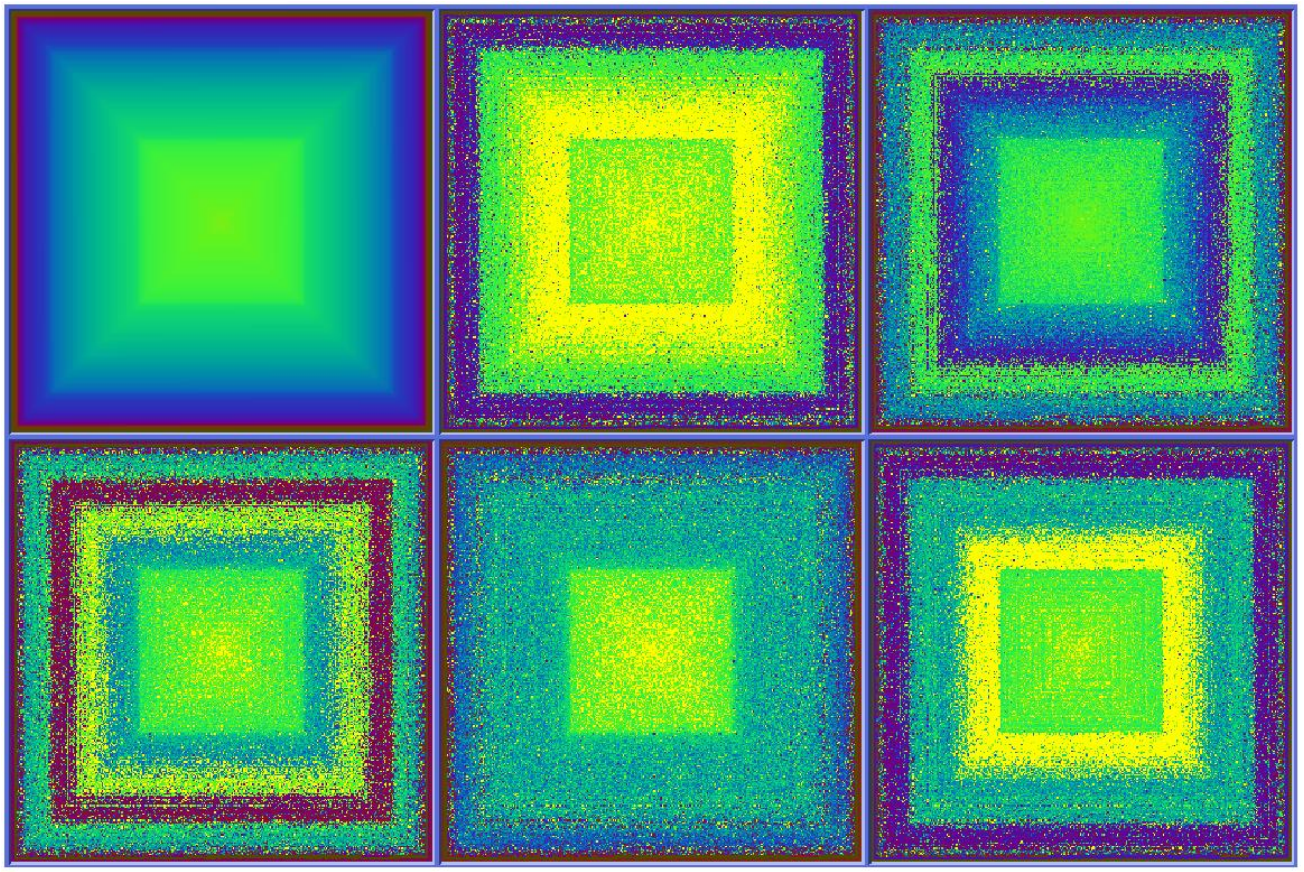
- **colour**
  - changes in hue at a given value



- **texture**
  - variation in pattern

- **motion**

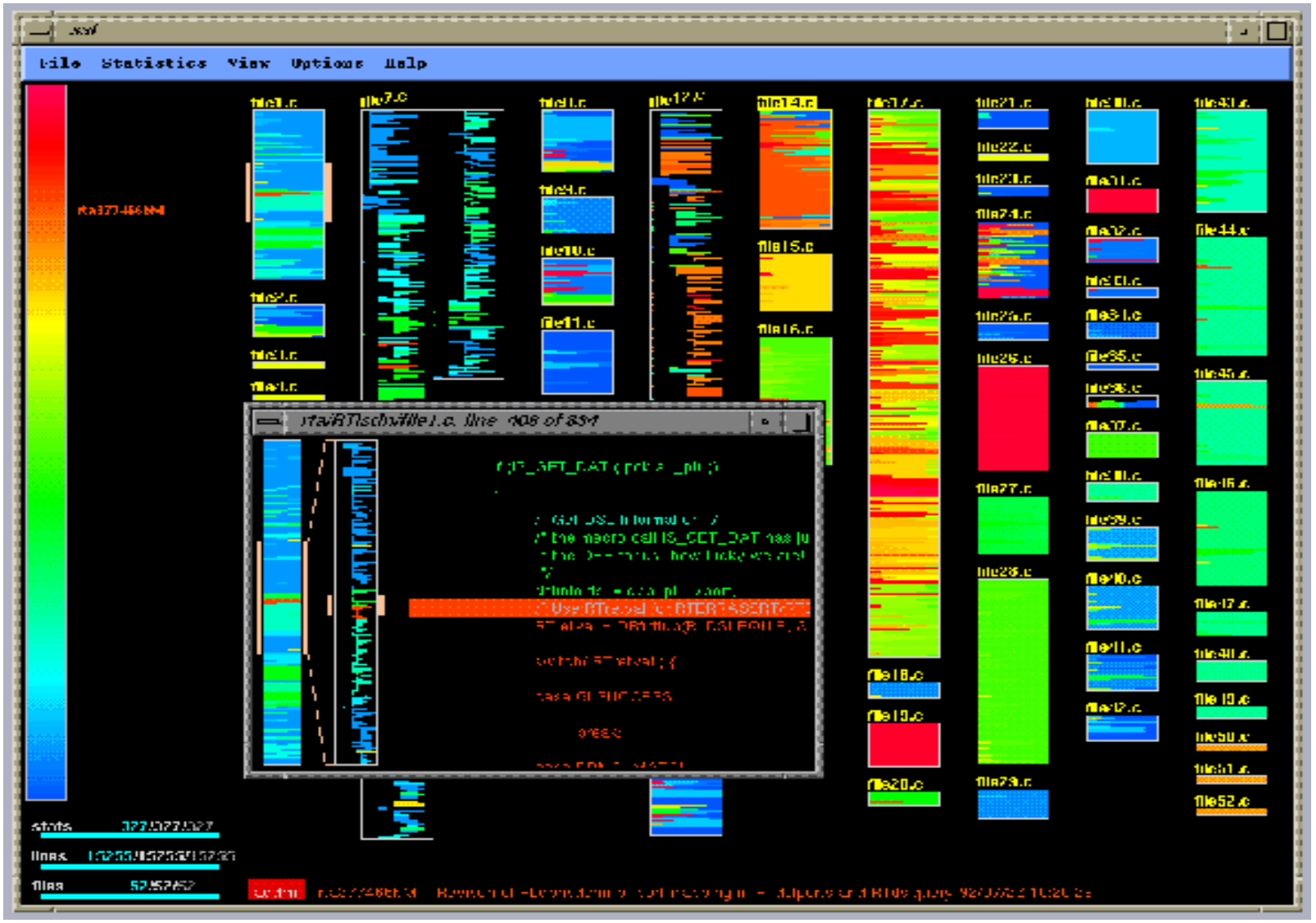
Database Visualisation: point (color)



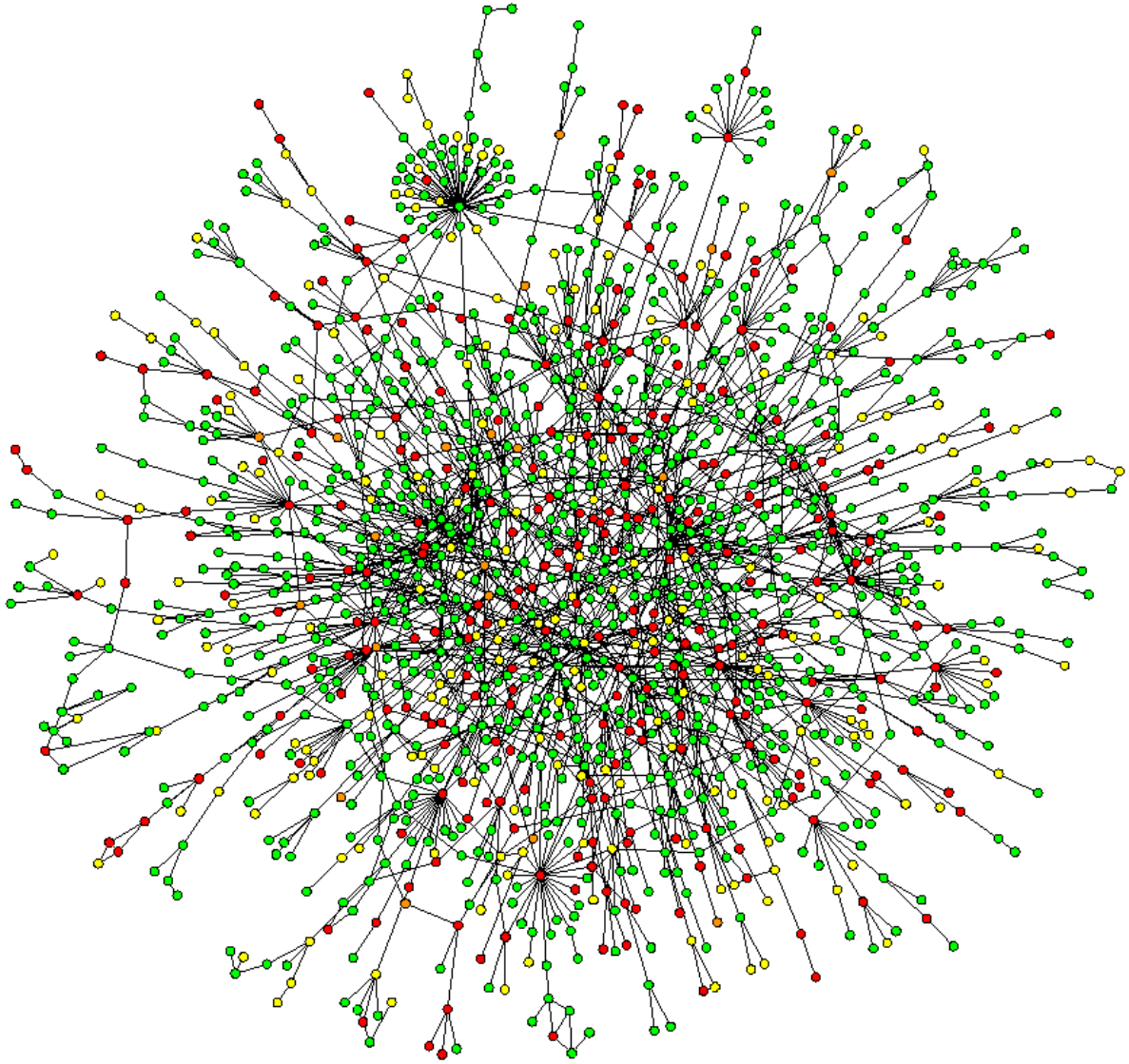
VISDB

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

Software Visualisation: line (color)



SeeSoft



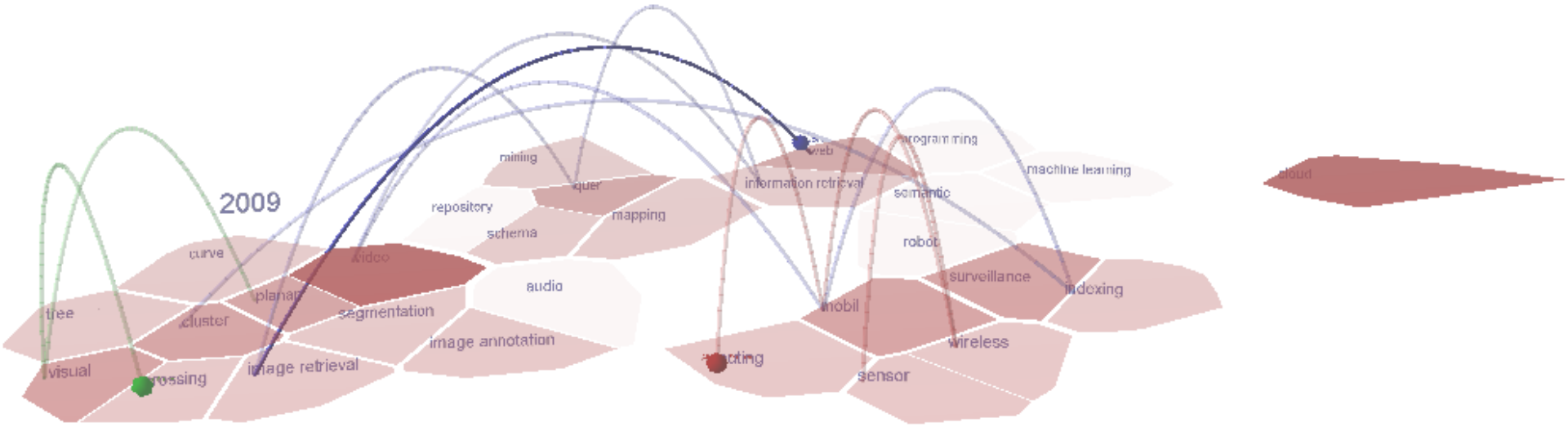
Visualization of Protein Interaction Network:

point+line (color)

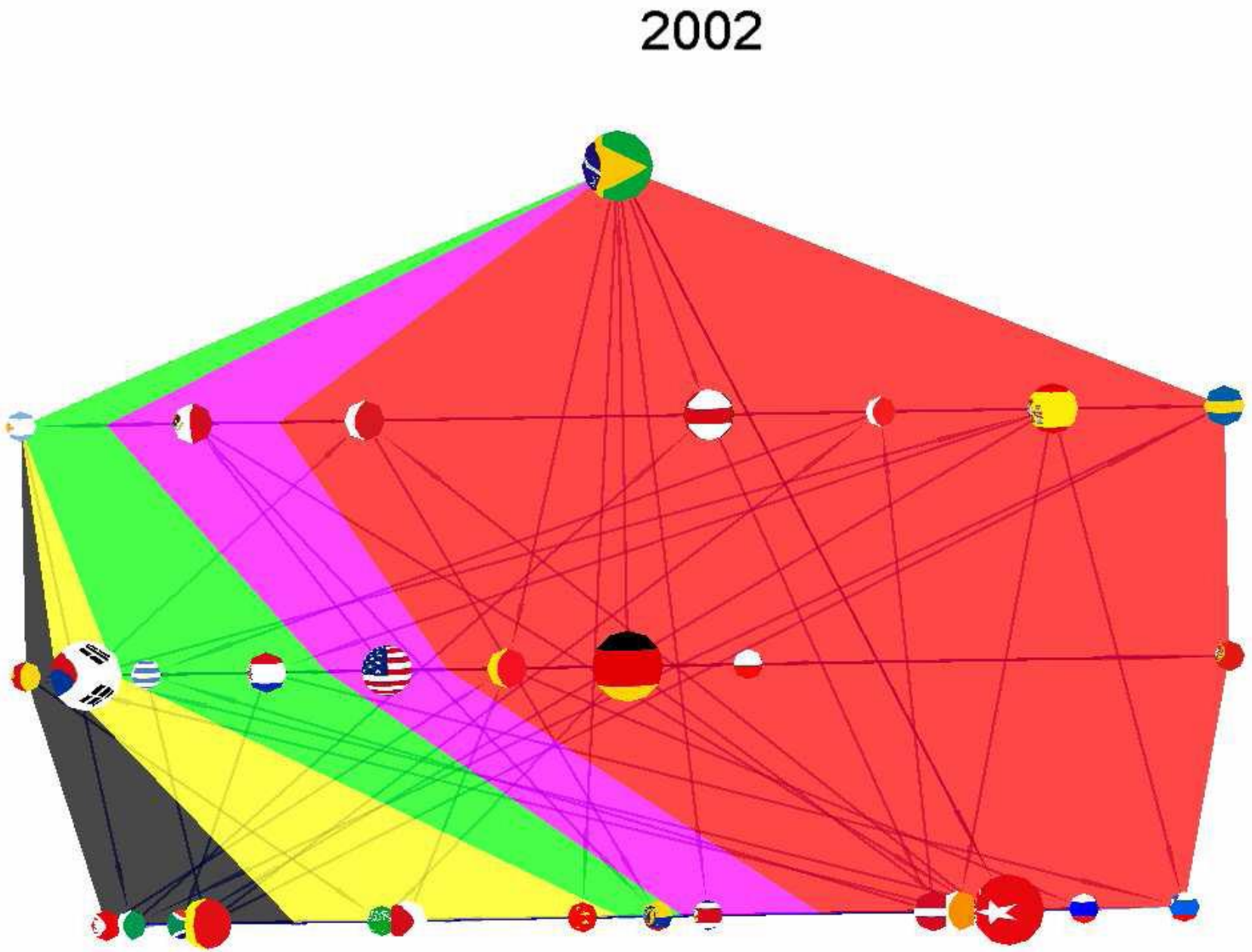
Map of protein-protein interactions. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-lethal; orange, slow growth; yellow, unknown). By **Hawoong Jeong**

Comparison of Trajectories:

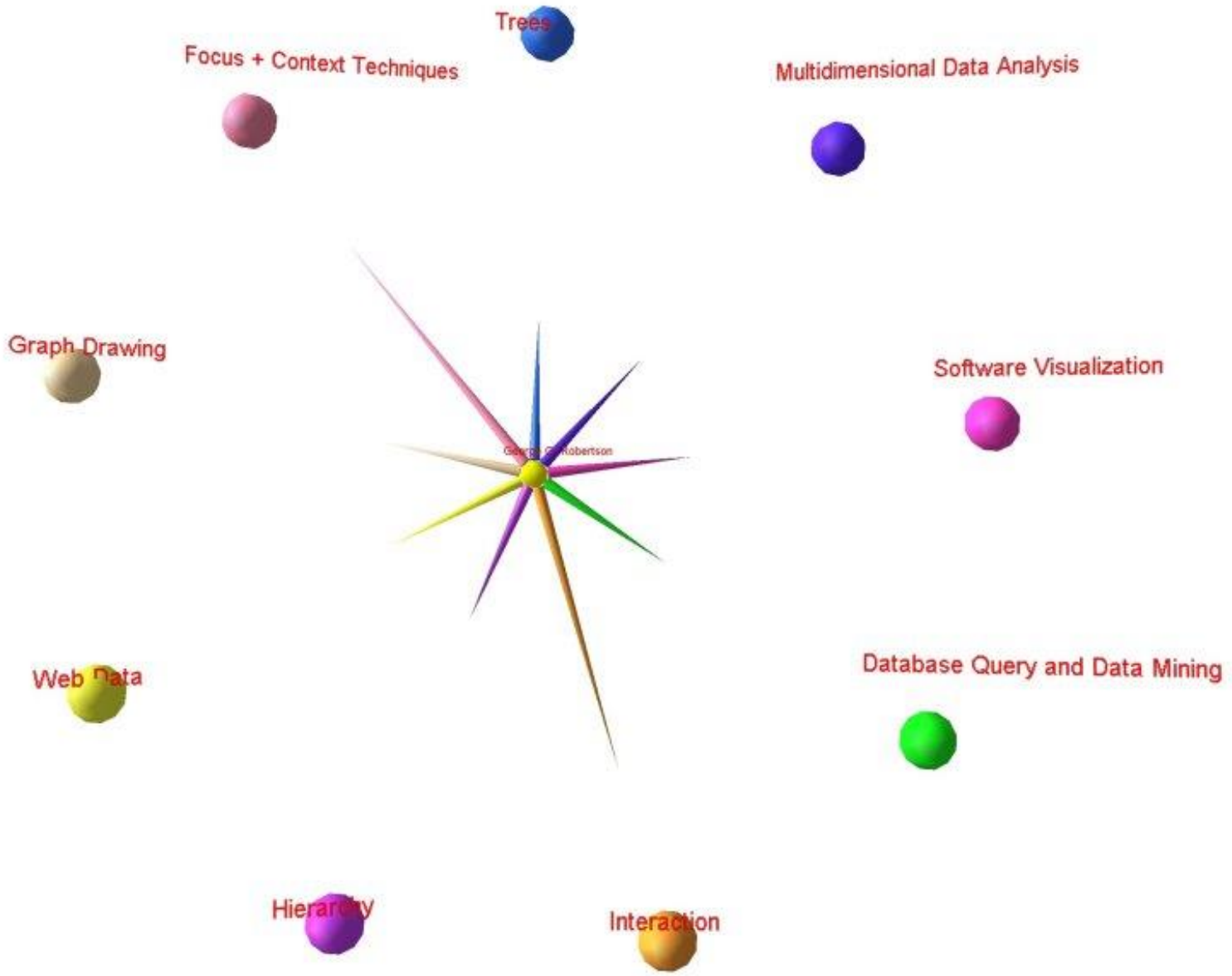
area, point, line (color, color hue)

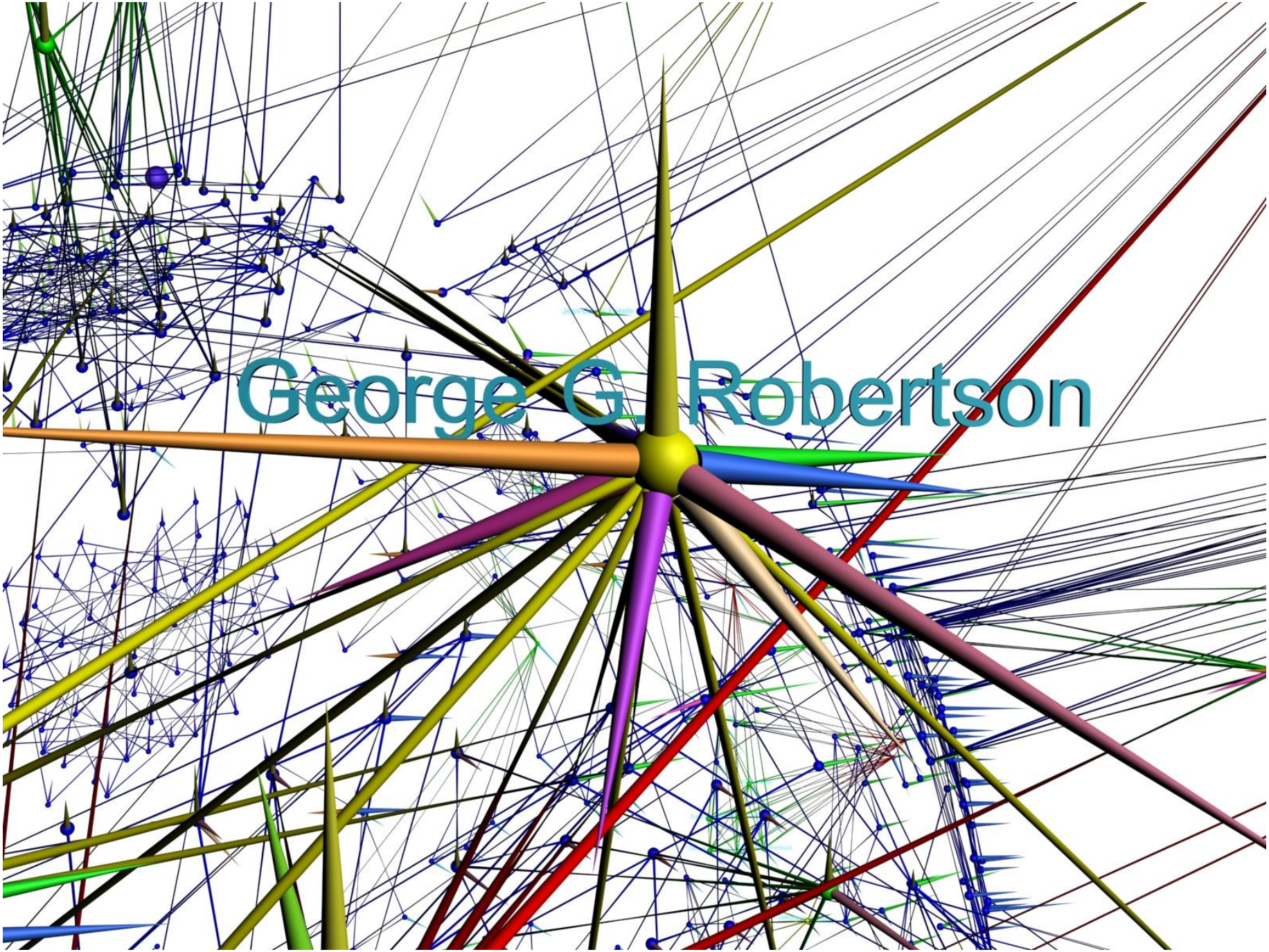


History of World Cup: area, point, line (color, texture, size)



Glyph: research area (point+line, color)

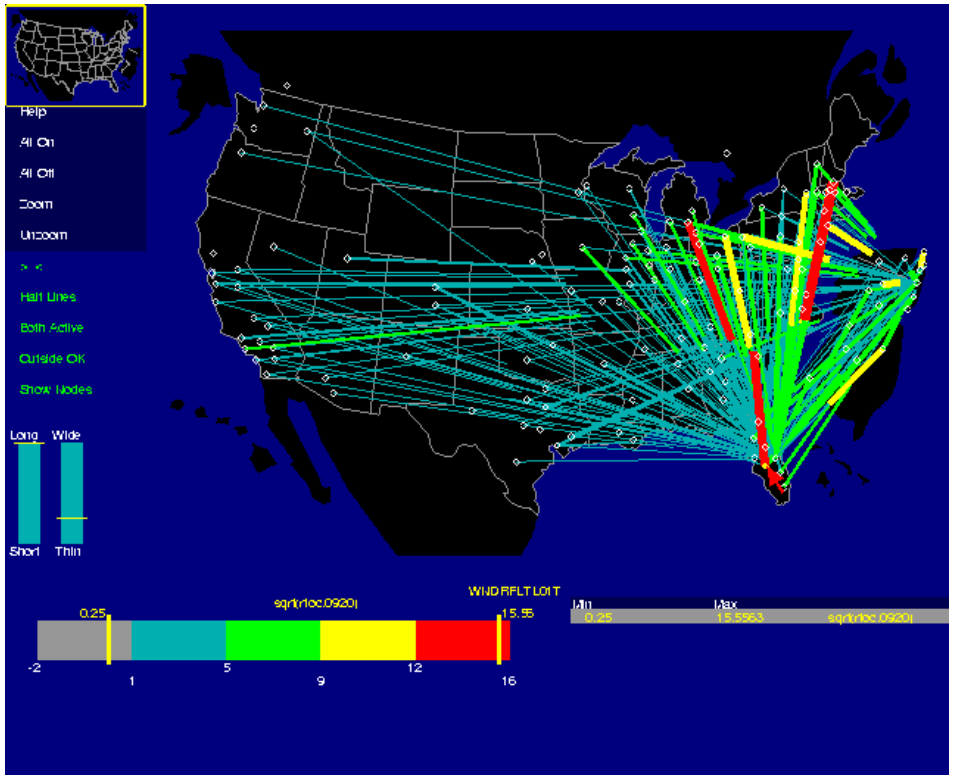



George G. Robertson

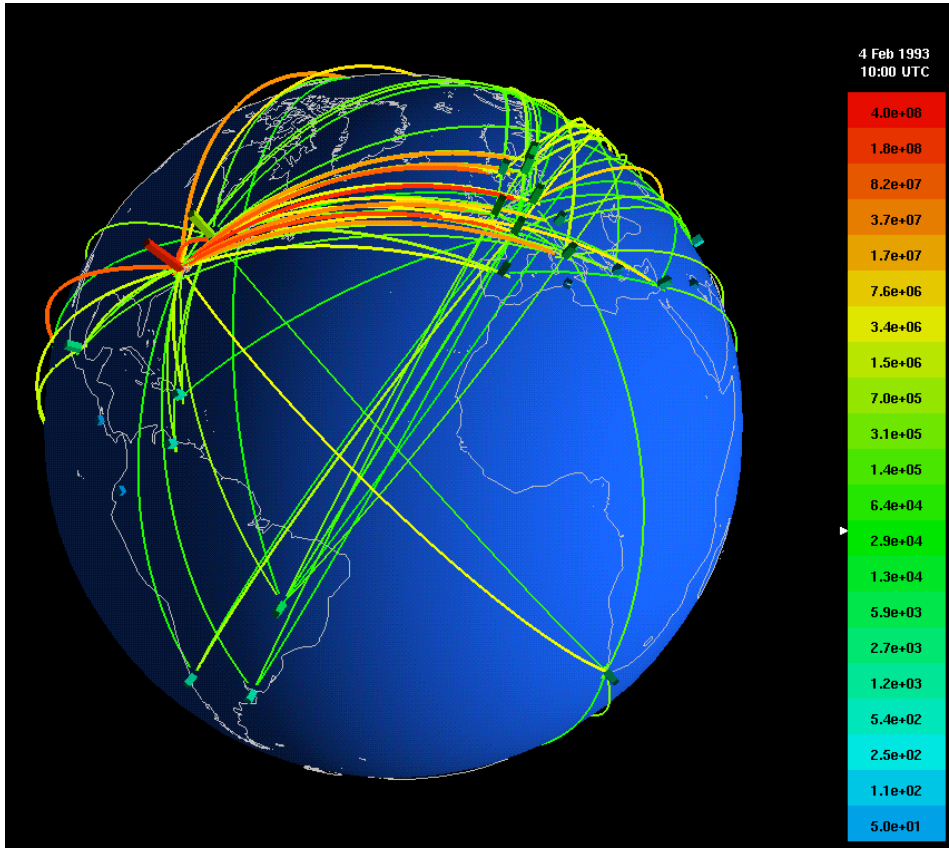
(2) Spatial Data Visualisation

Geo Visualisation (map)
Trajectory visualisation

Geographic/Spatial Data

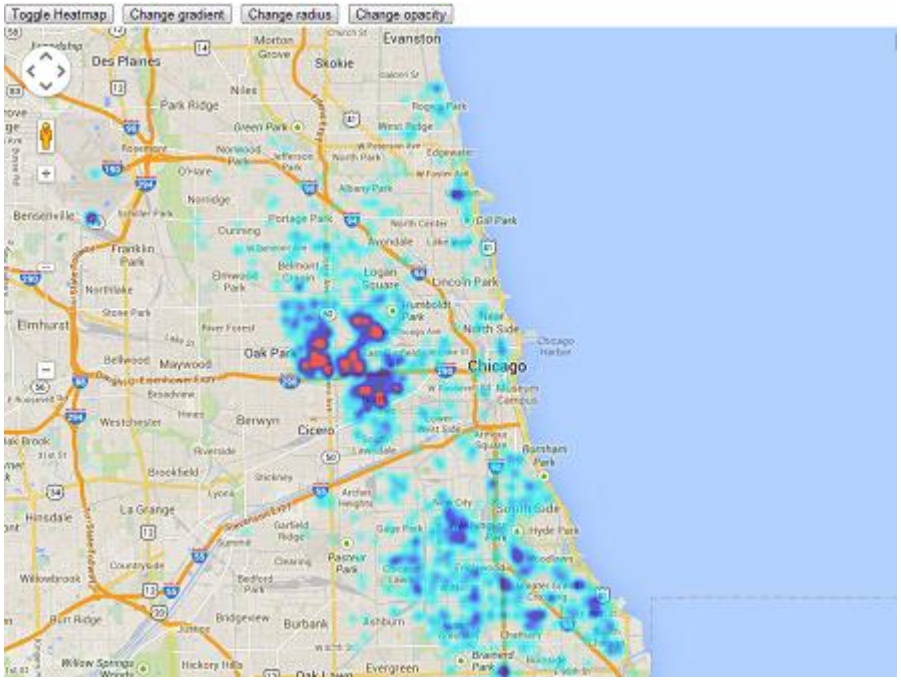


- [Becker, Eick, Wilks 95]
- SeeNet

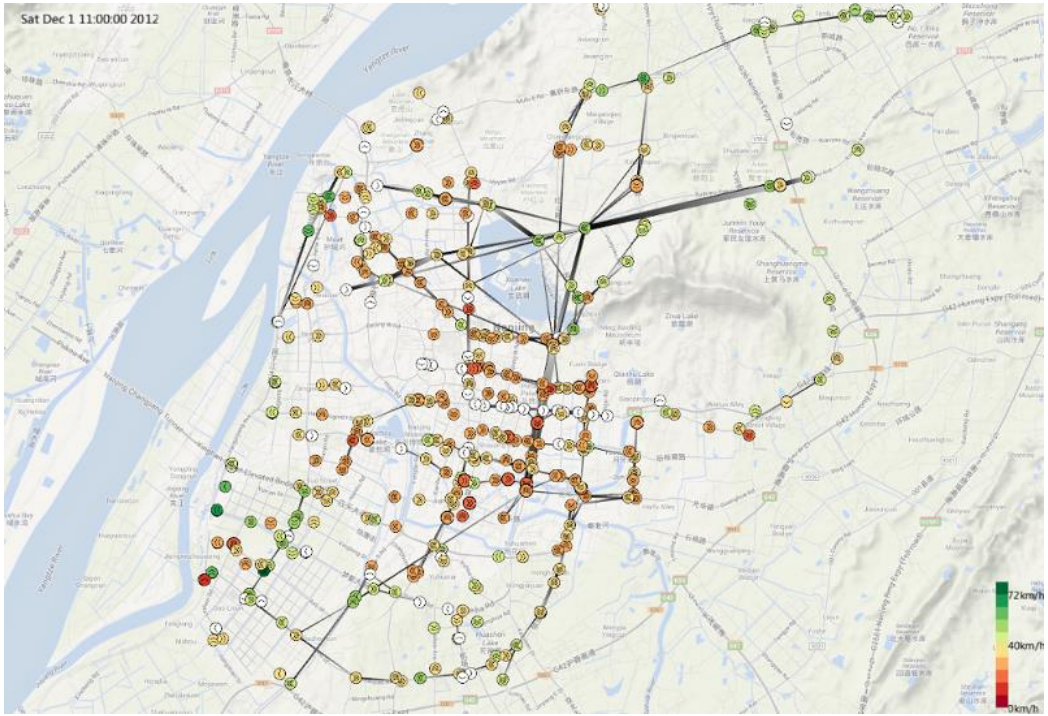


[Cox, Eick 95] 3D Displays of Network Traffic

SeeNet3D

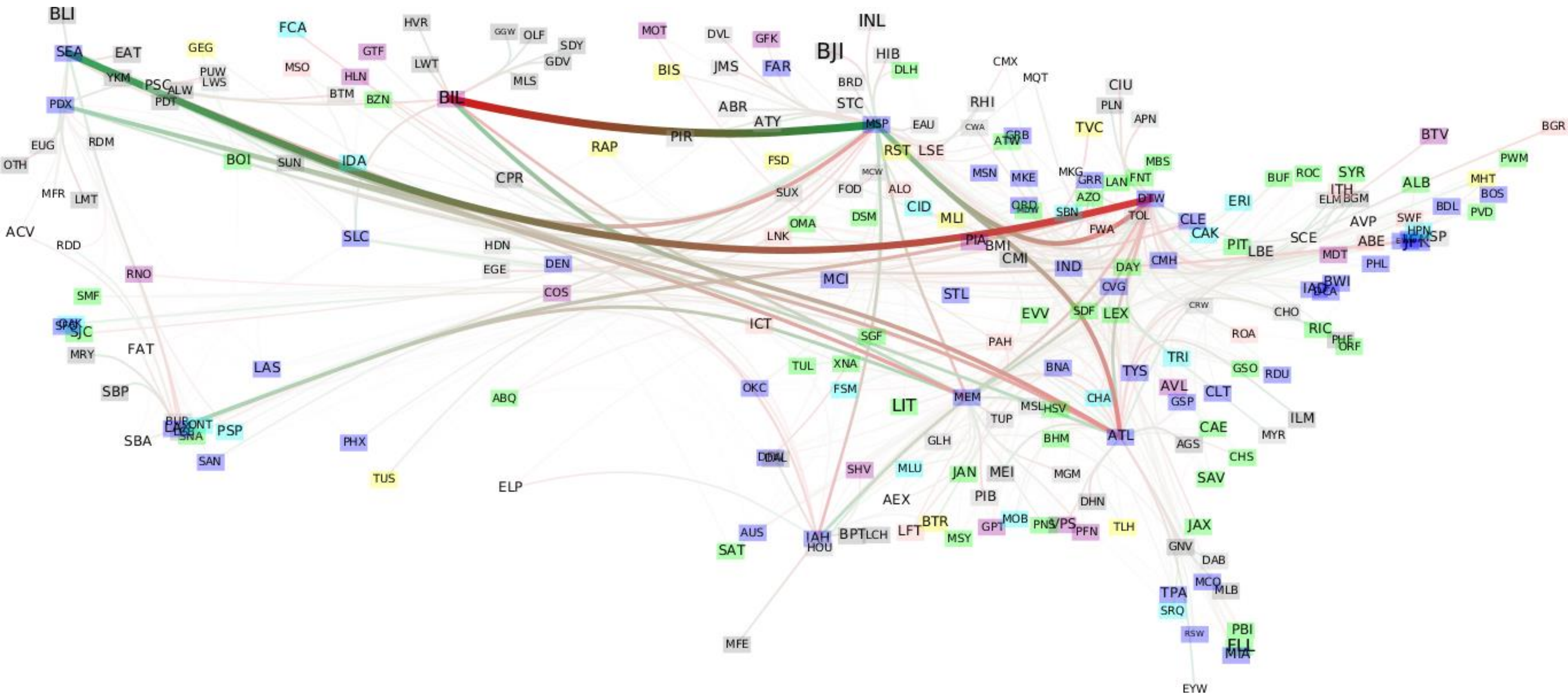


Crime Data Visualisation

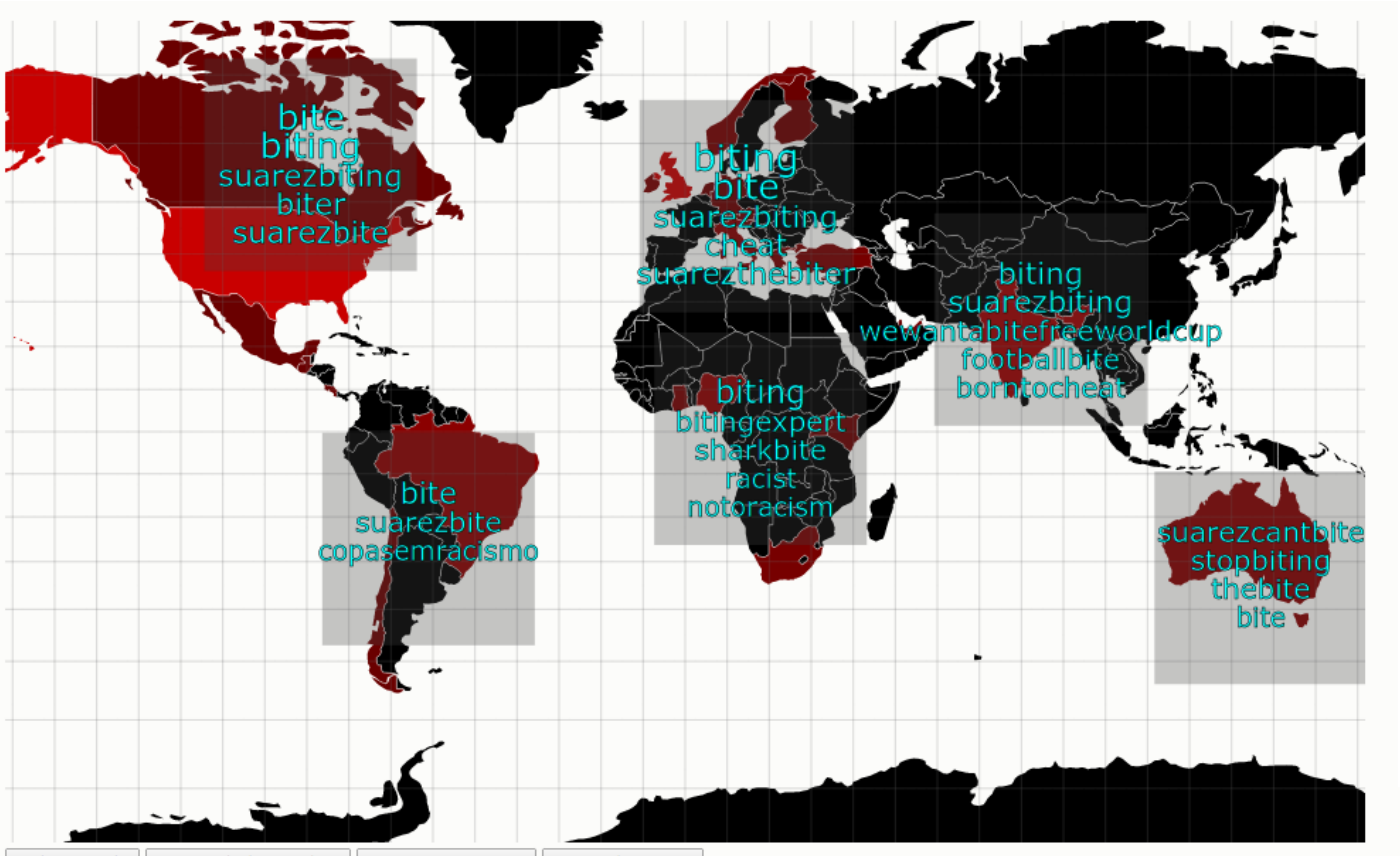


Traffice analysis using Trajectory

US Flight Network Traffic Analysis

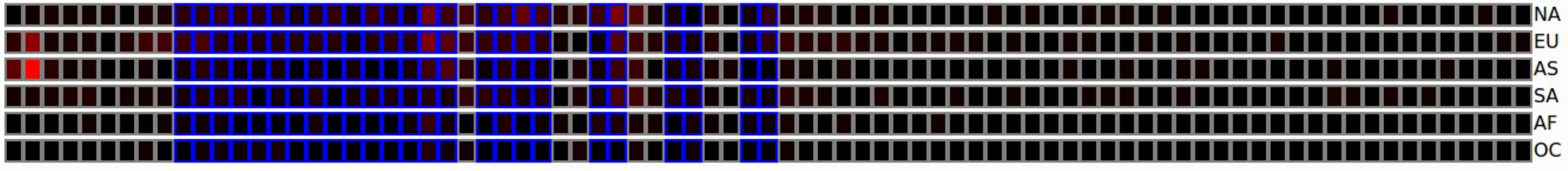


Twitter (WorldCup) Analysis



Auto-animate | Pause/start animation | Toggle map colour | Toggle tag cloud

Current date: 2014-6-25



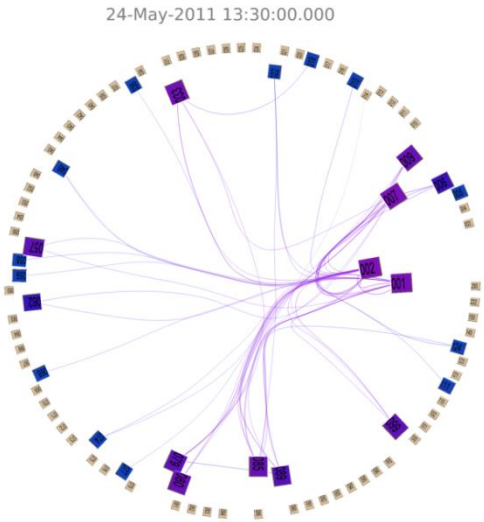
(3) Temporal/Dynamic Data Visualisation

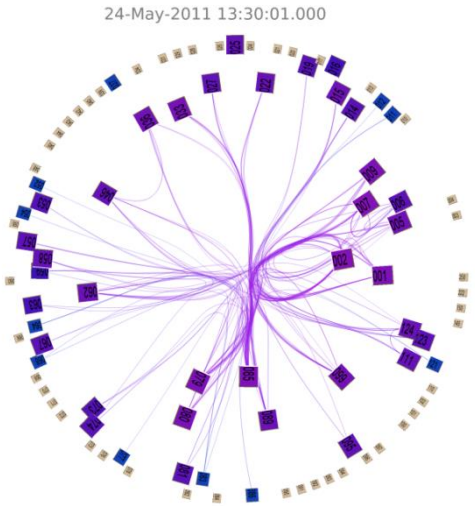
1. Stream Graph
2. Theme River
3. Storyline
4. Small Multiples
5. 2.5D Visualisation (stacked graph)
6. Animation

Criteria (Design consideration)

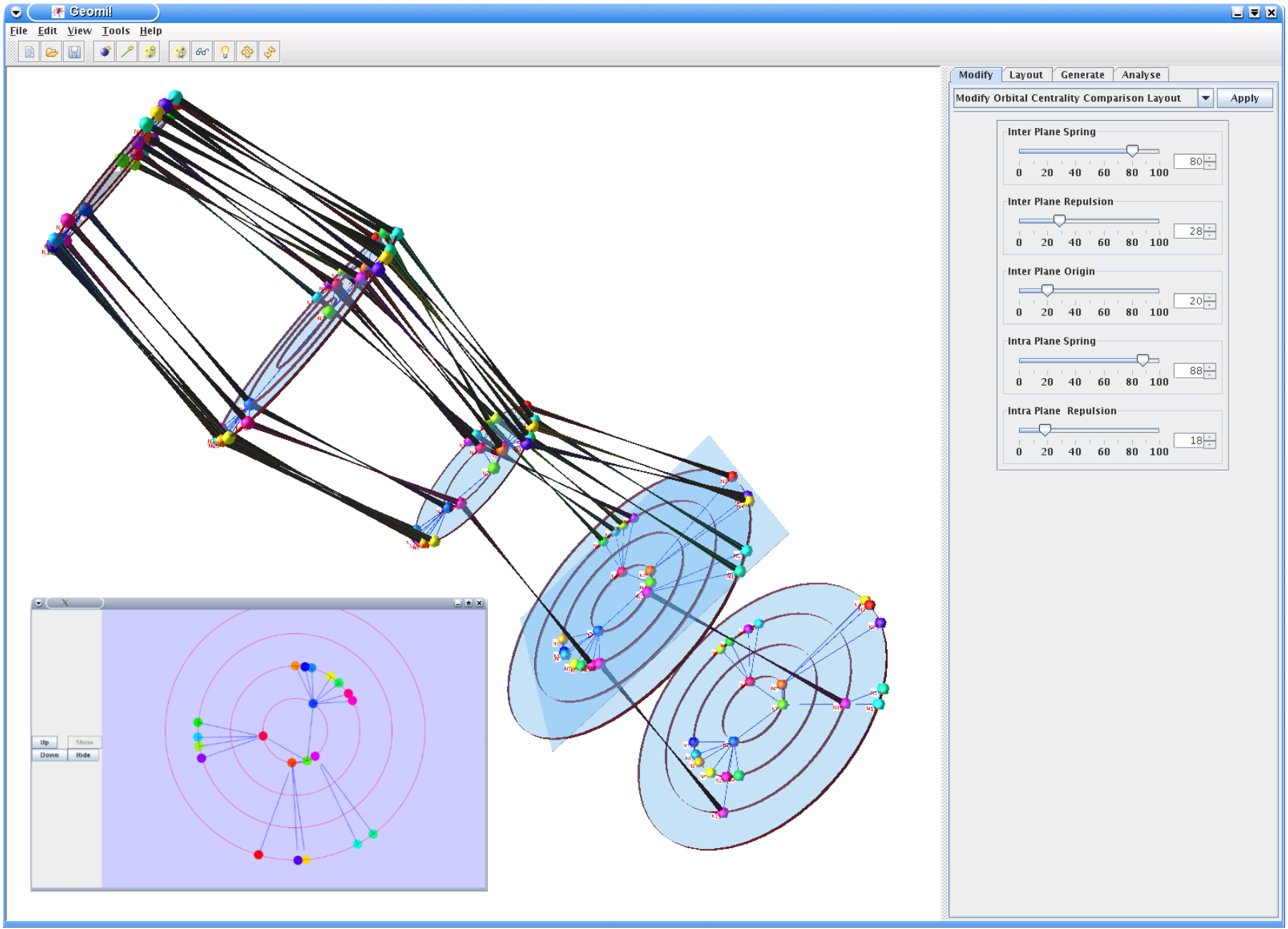
1. **Preserving Mental Map** [Eades, Lei, Misue, Sugiyama 1995]
   - Orthogonal ordering
   - Proximity
   - Topology

2. **Change Faithfulness** [Nguyen, Eades, Hong 2013]
   - Change in Visualisation should be proportional to the change in the data

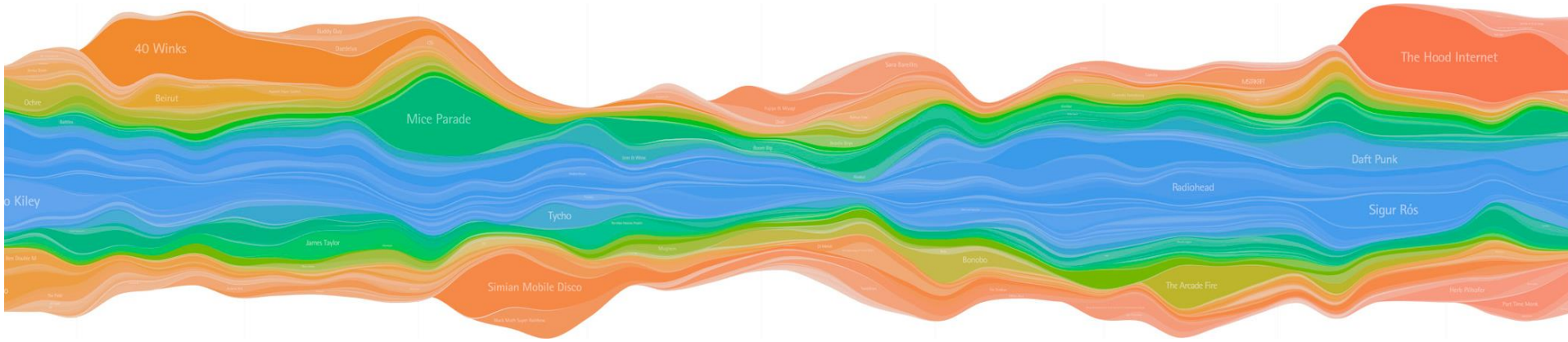




Comparison of Network Centralities

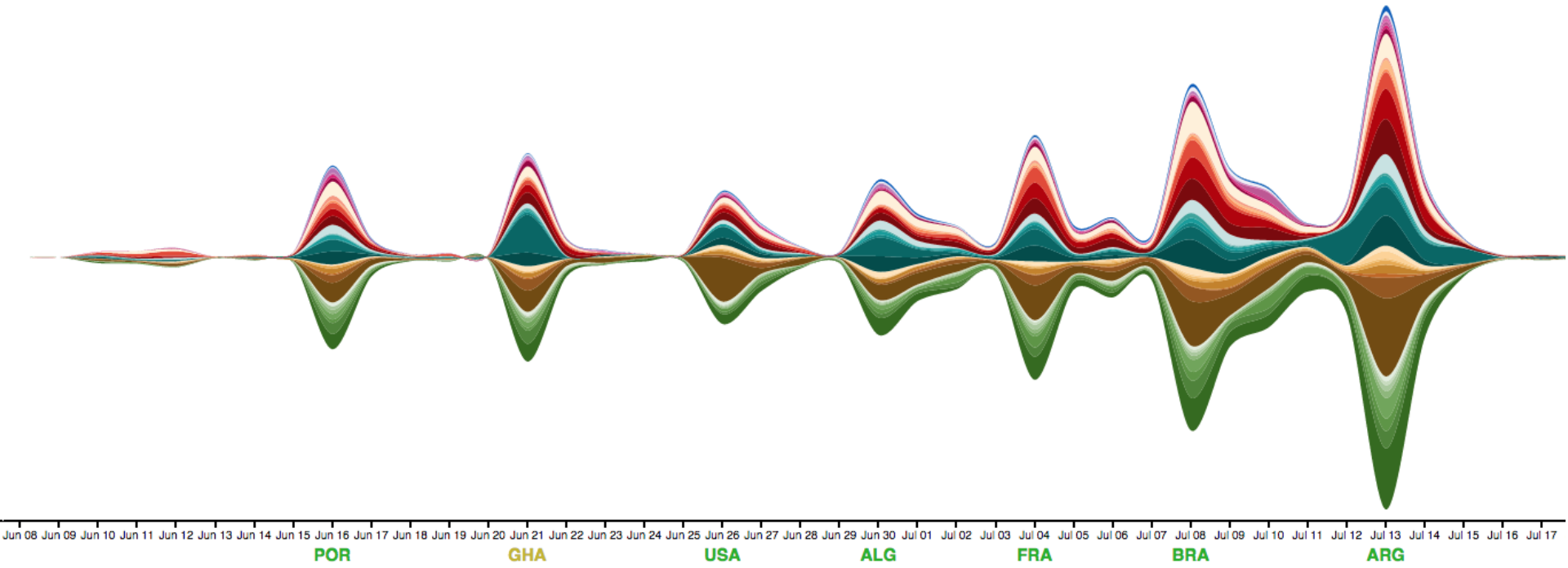


1. Stream Graph



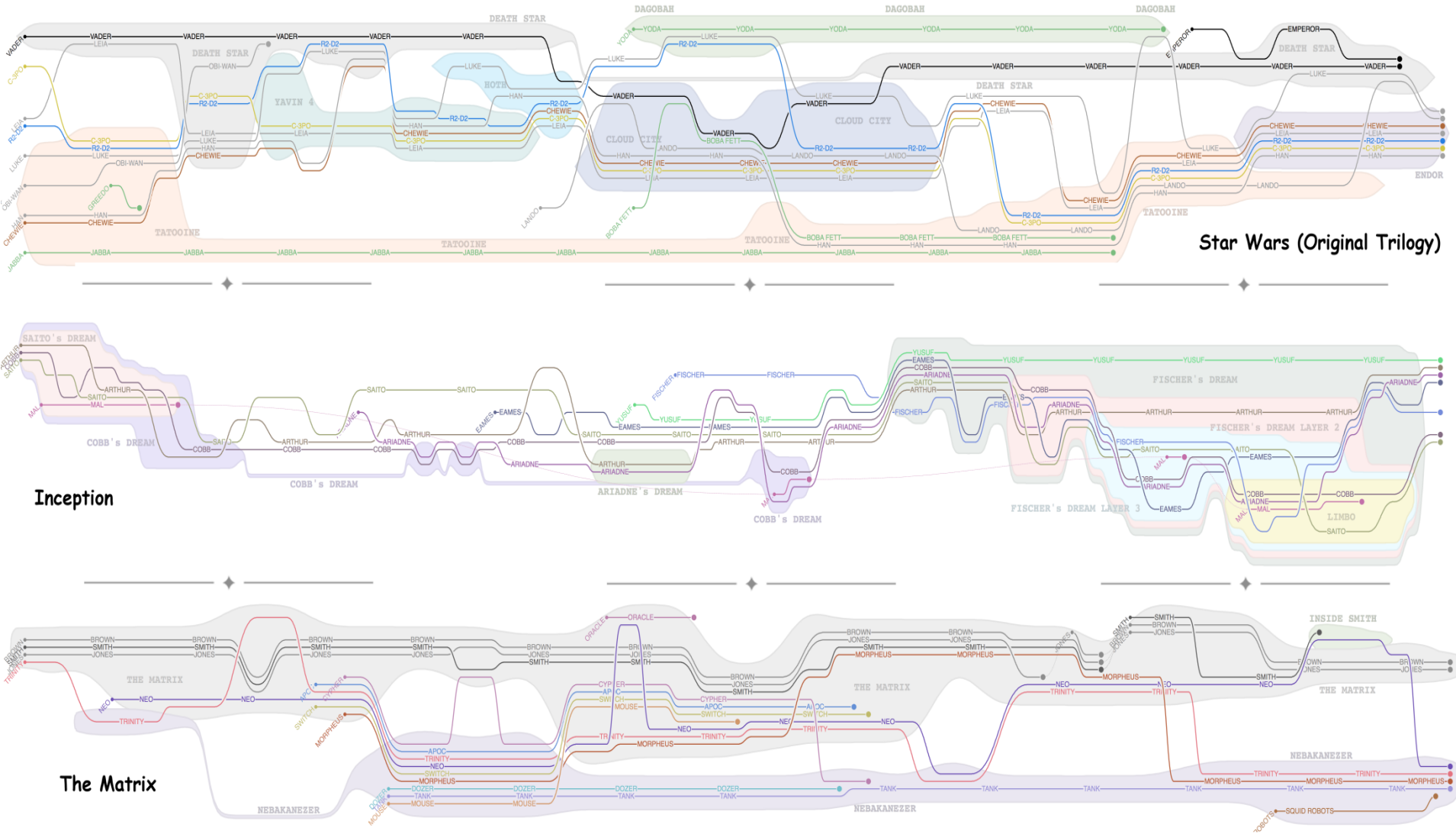
Shows overview
Trend analysis

2. Theme River



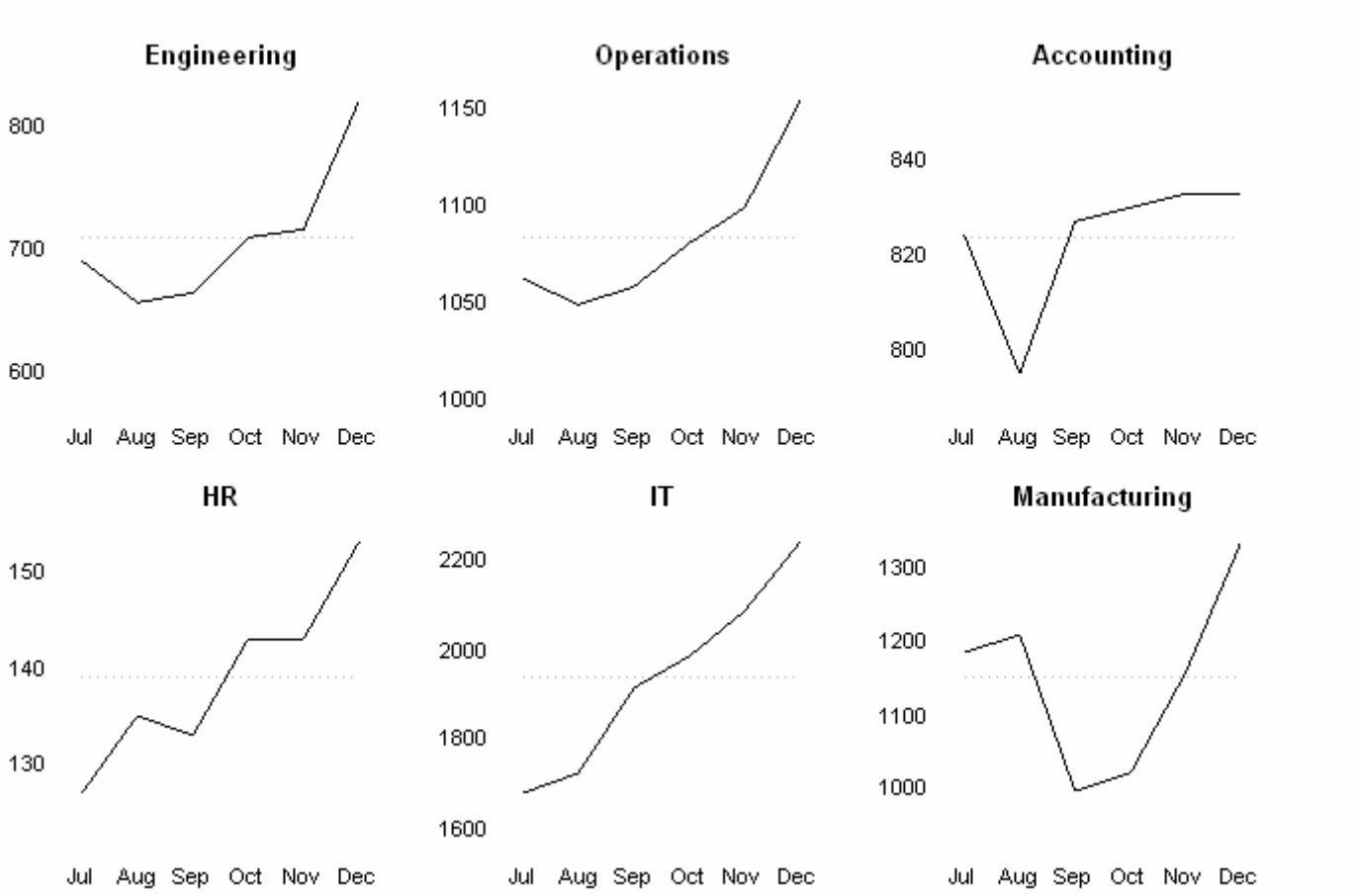
Shows overview
Trend analysis

3. Storyline



Shows overview
Dynamic clustering

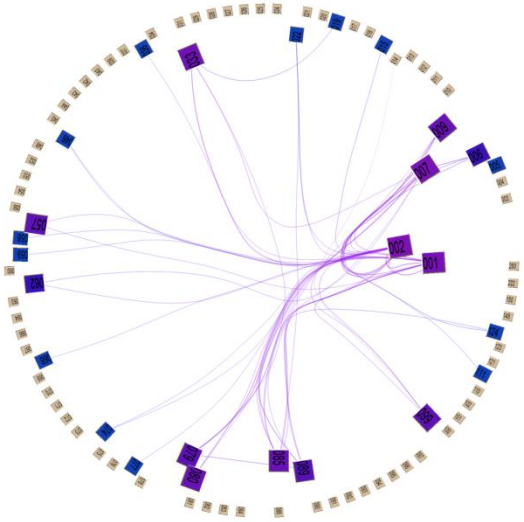
4. Small Multiples



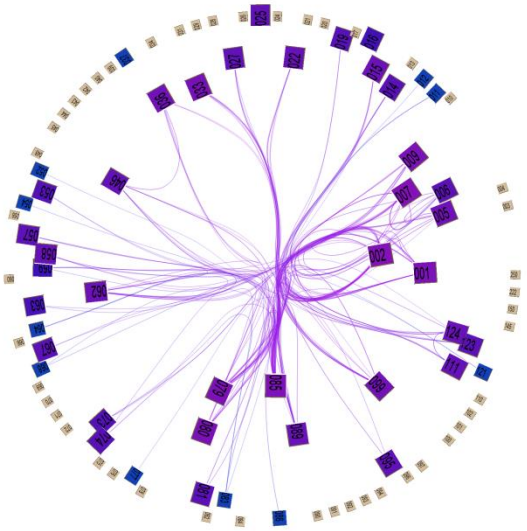
Shows overview
Visual Comparison

Stock Trading Network Analysis

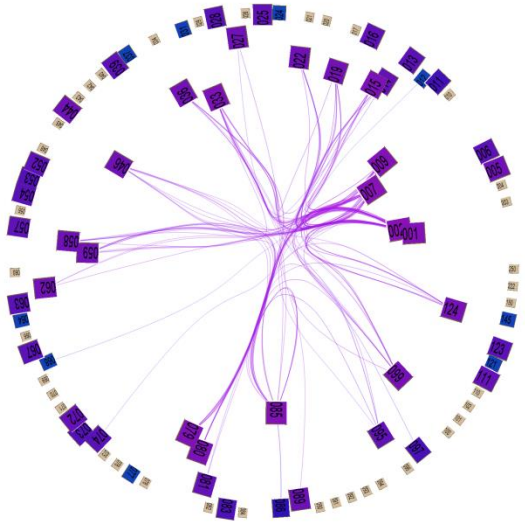
24-May-2011 13:30:00.000



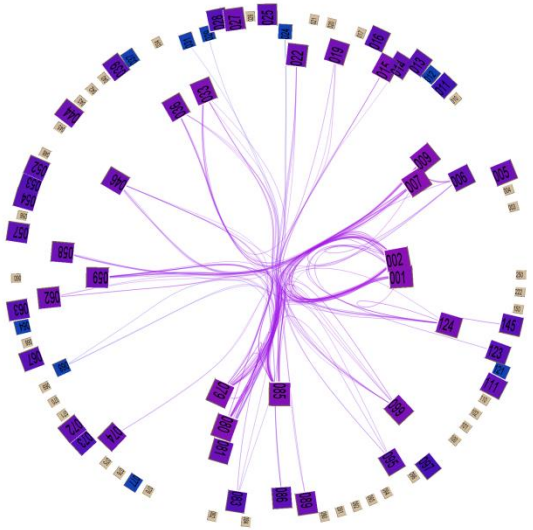
24-May-2011 13:30:01.000



24-May-2011 13:30:02.000

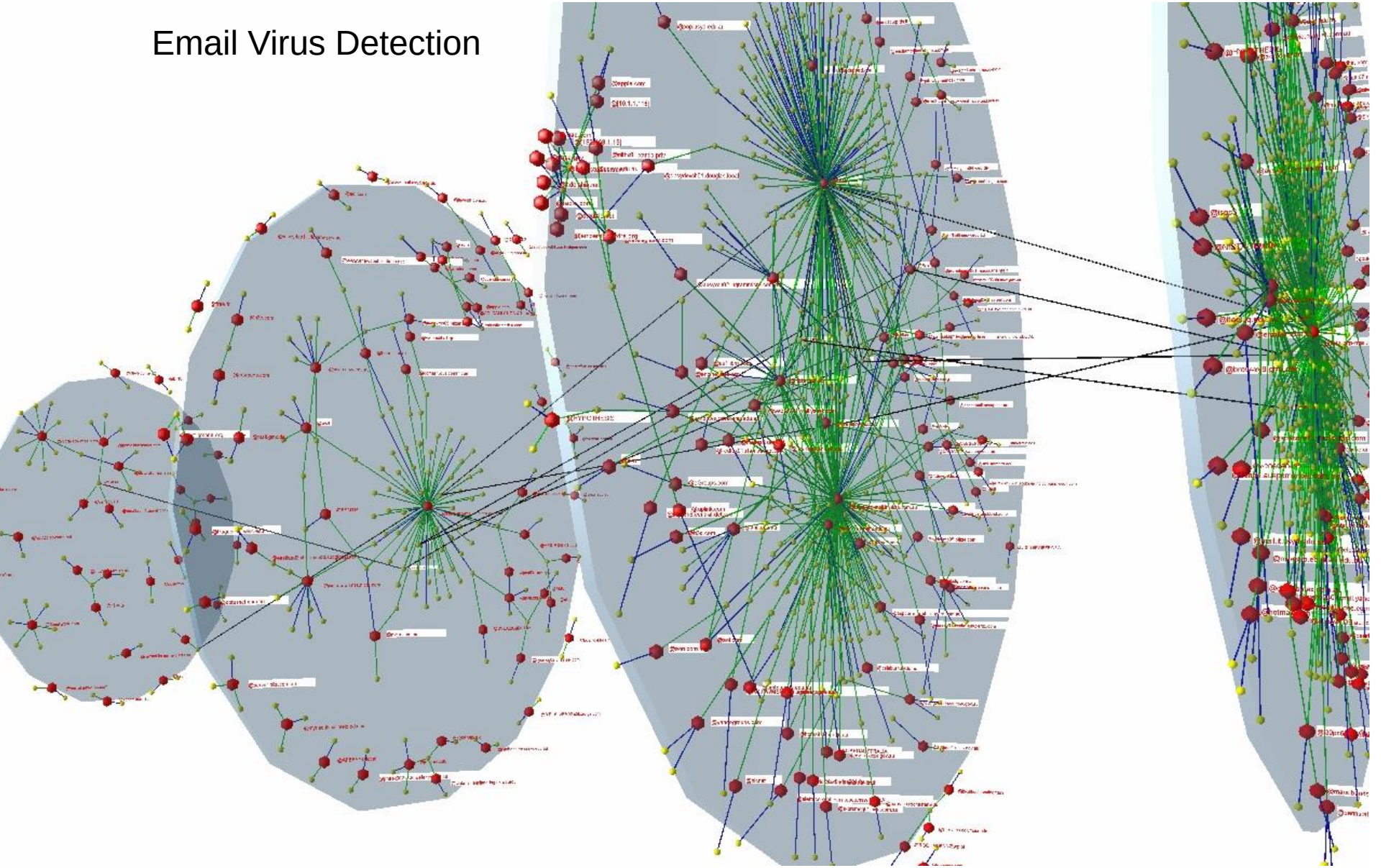


24-May-2011 13:30:03.000

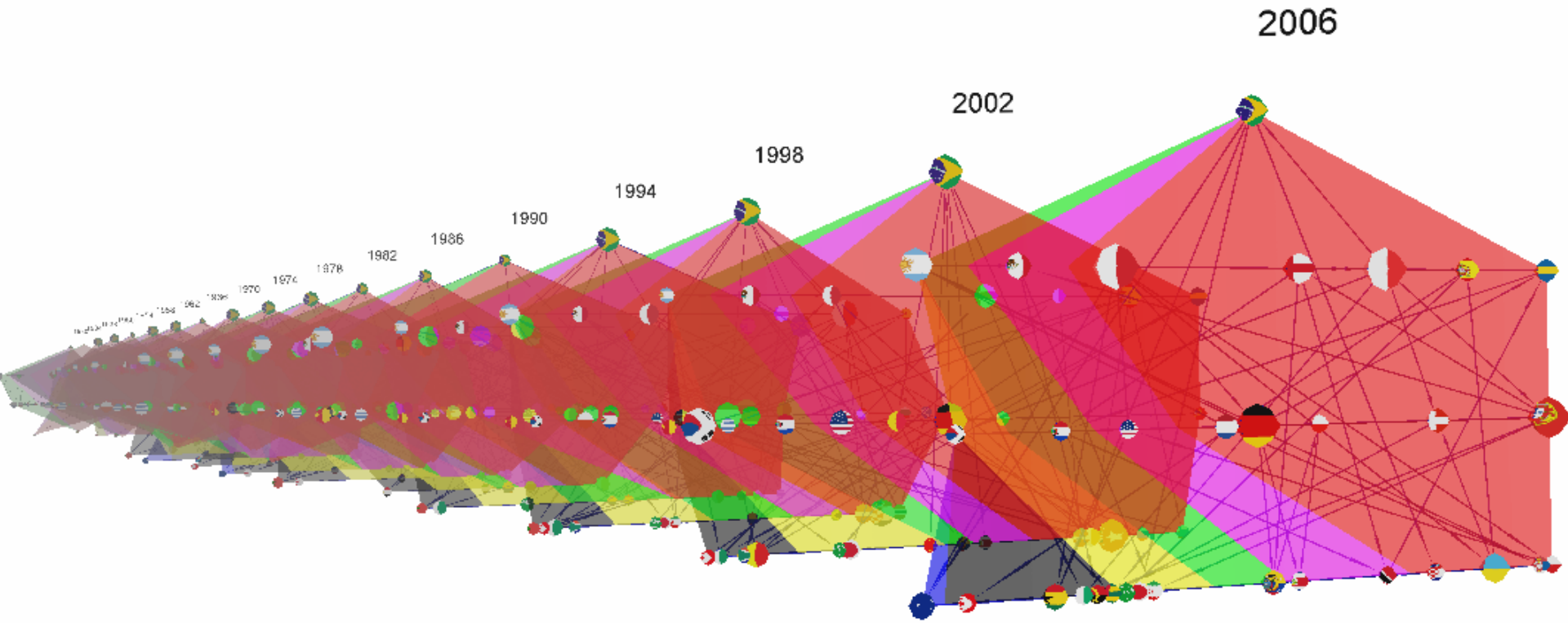


5. 2.5D Visualisation

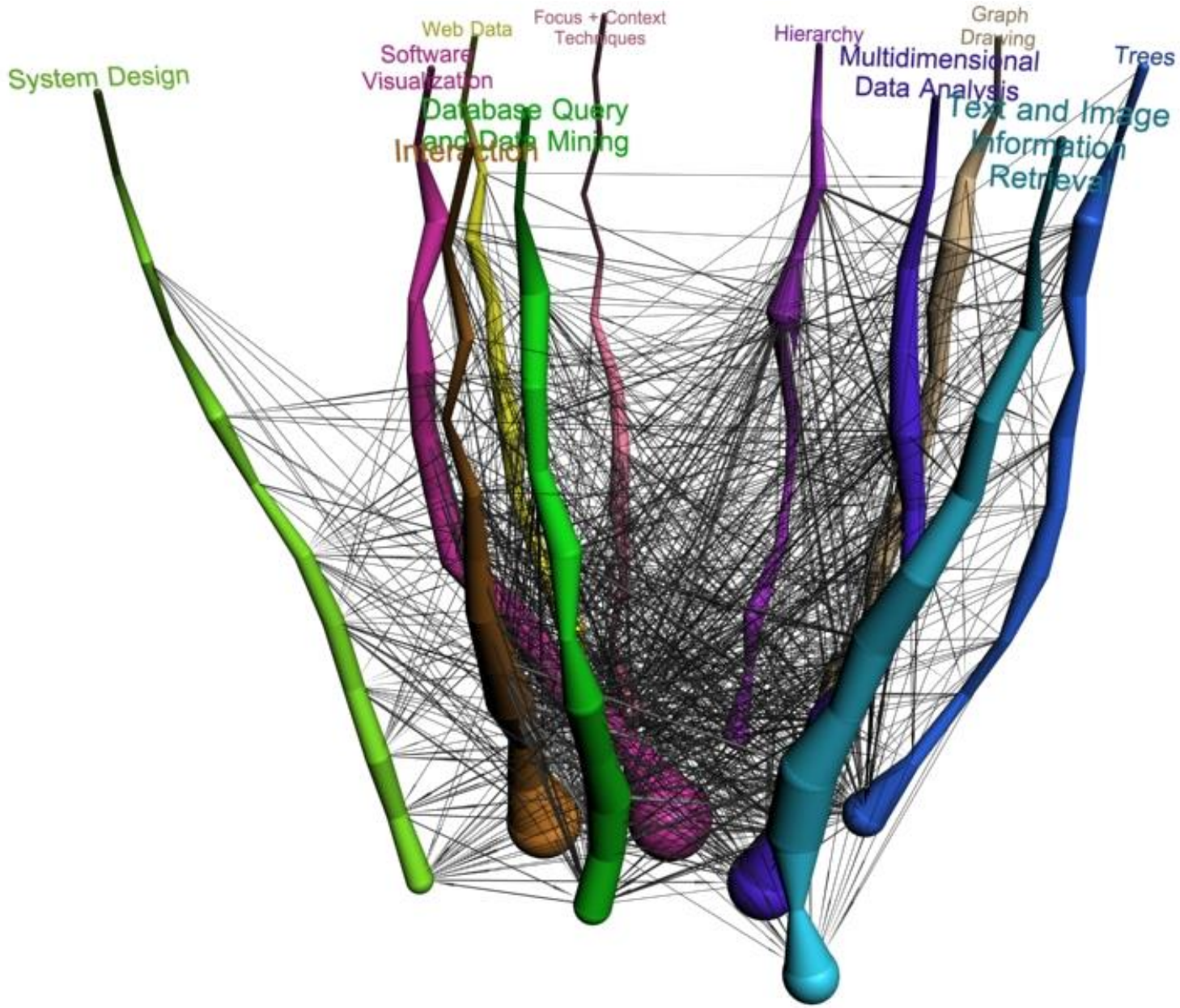
Email Virus Detection



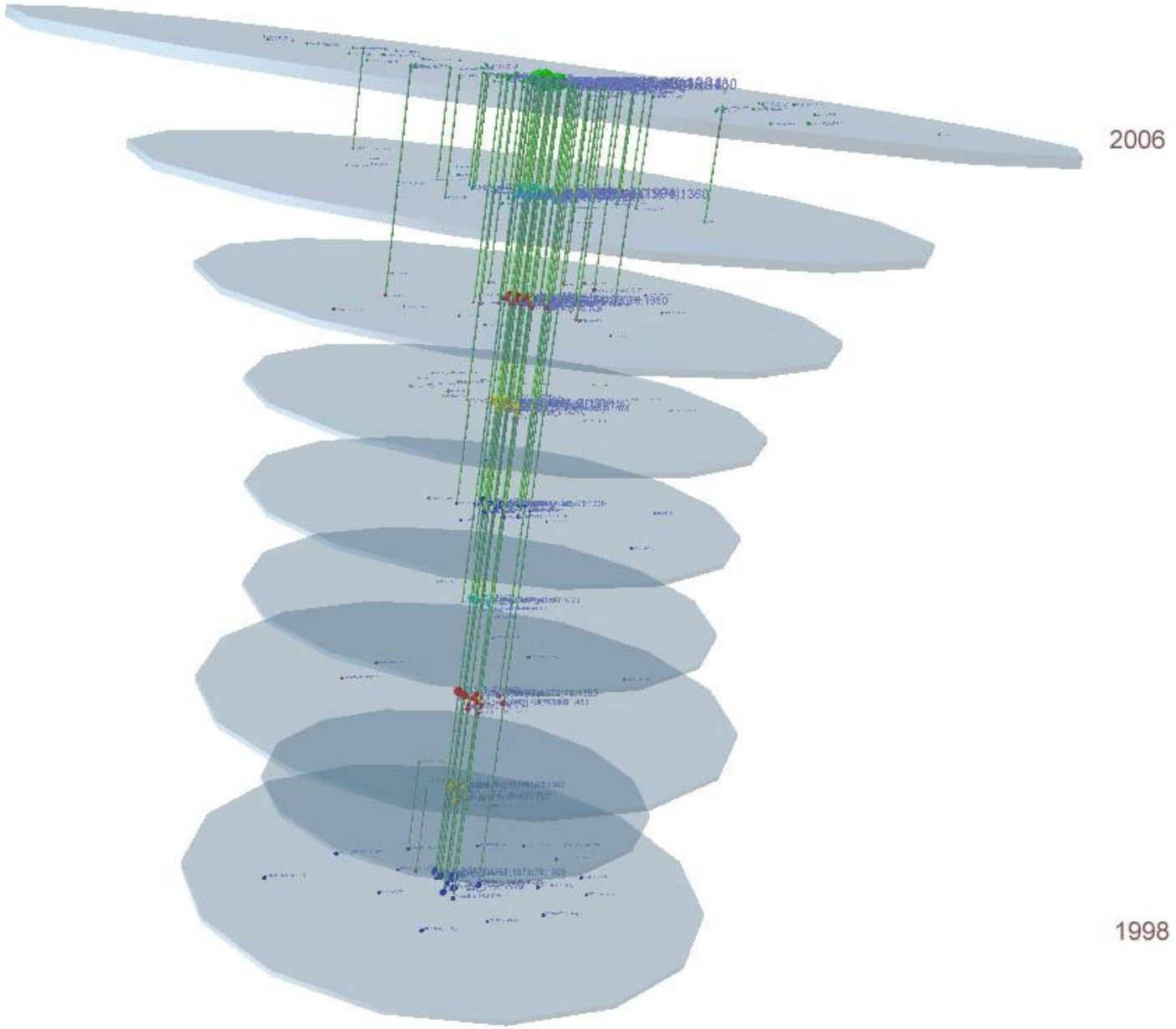
History of Worldcup



Evolution of Topics in Info Vis Community

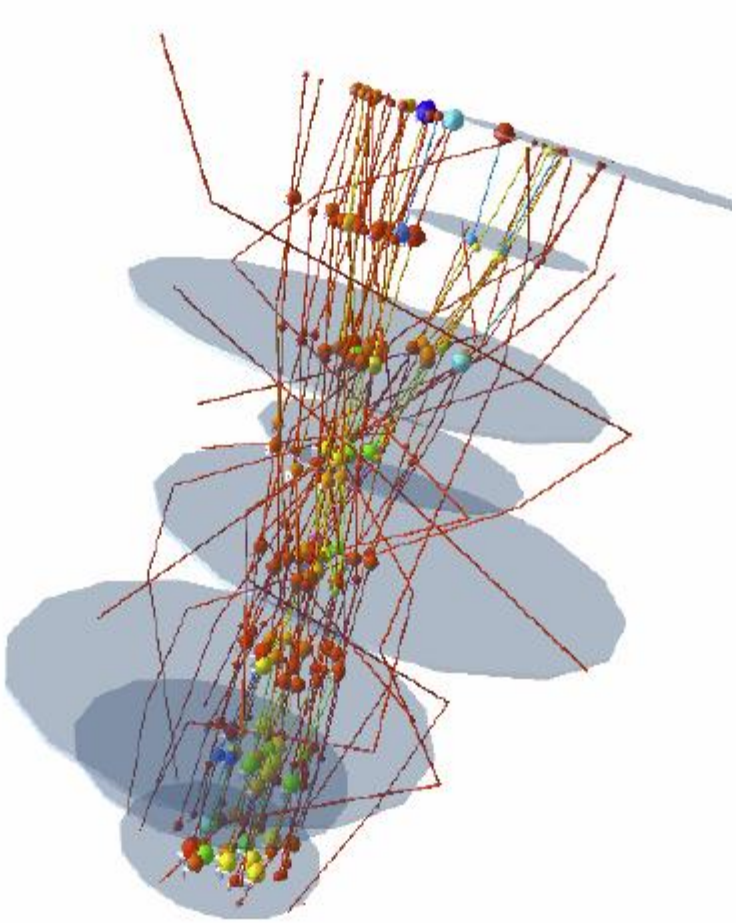


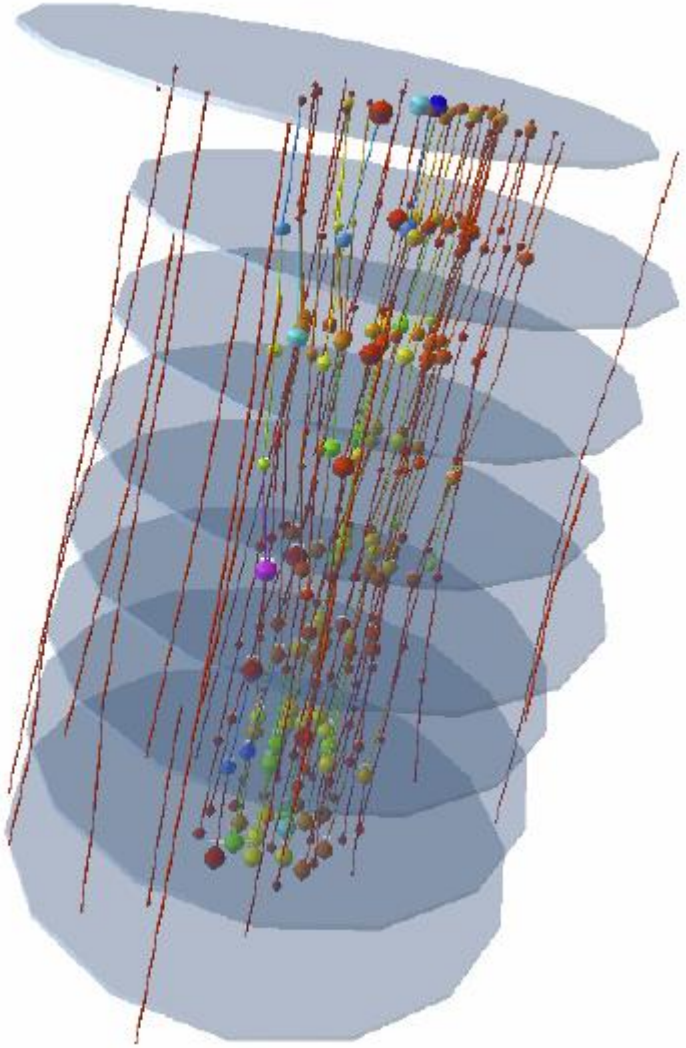
Evolution of Co-citation Network in WOS (social network)



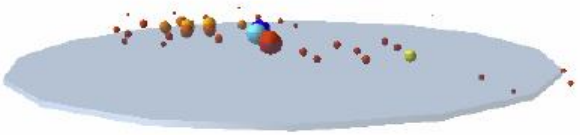
(A) Union graph approach

Union graph G= G1+G2+ G3+….Gk

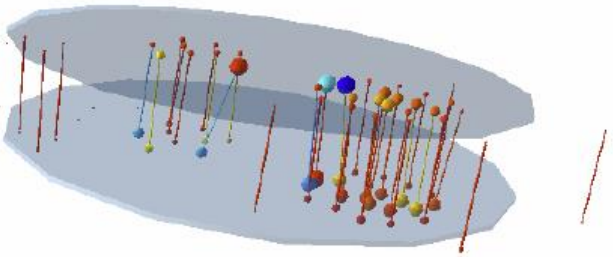




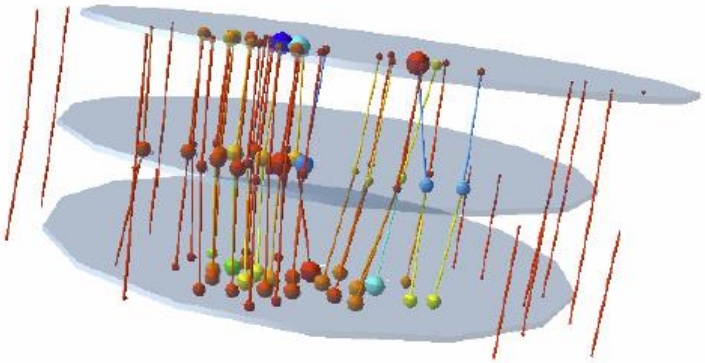
(B) Use good initialisation



(a) The first plate.



(b) The second plate.



(c) The third plate.

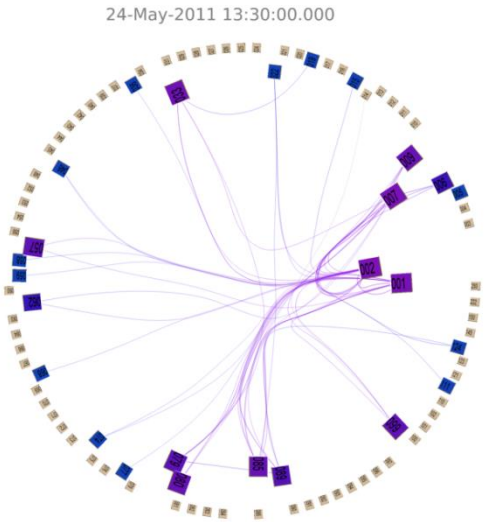
6. Animation

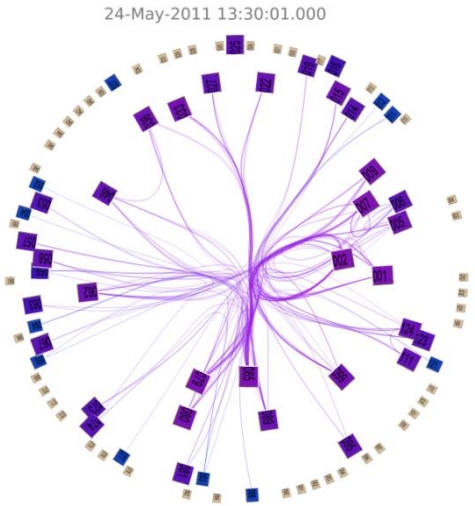
Criteria for Good Animation

- Preserving Mental Map
- Change faithful

Popular method: Linear interpolation between two visualisations to show smooth animation

* Short-term memory: 7 (+- 2)



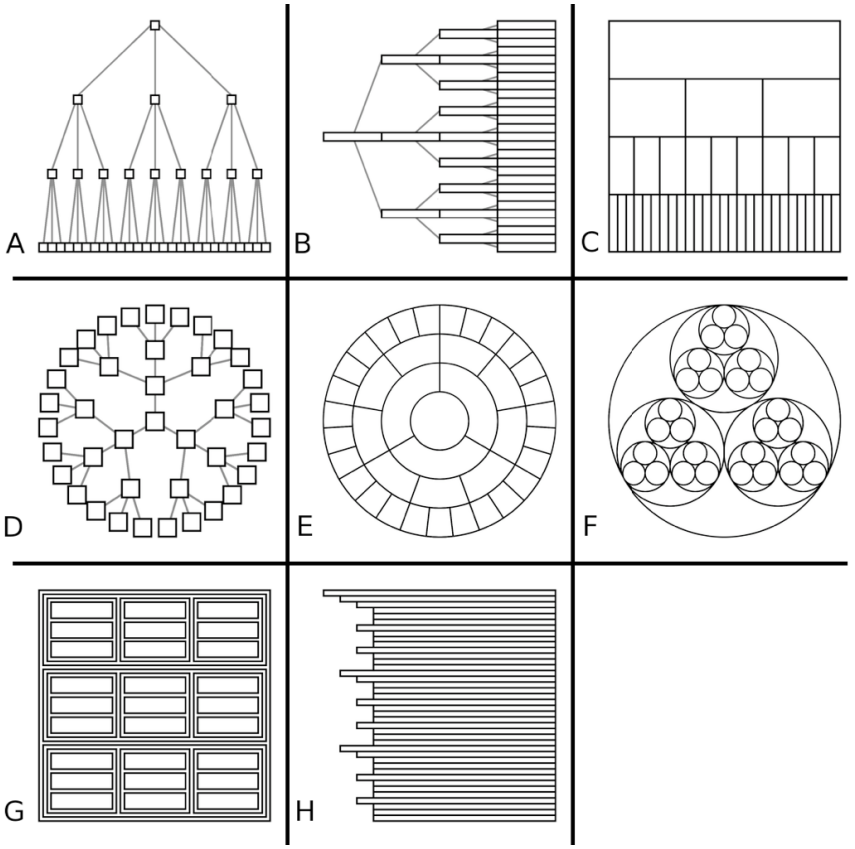


(4) Data with Constraints

1. Relations
2. Hierarchy
3. Clusters

Tree visualization: Hierarchical relation

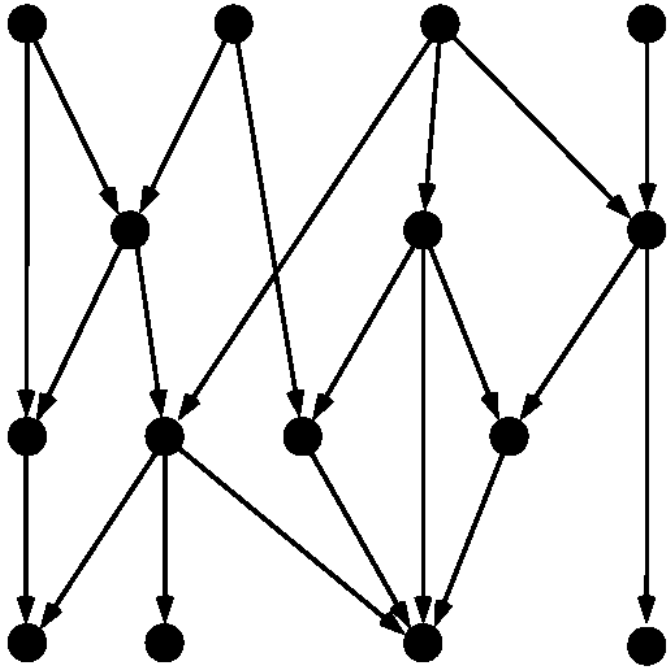
- Layered Drawing
- Radial Drawing
- HV-Drawing
- Inclusion Drawing



Sugiyama Method: Hierarchy, Direction

4 steps:

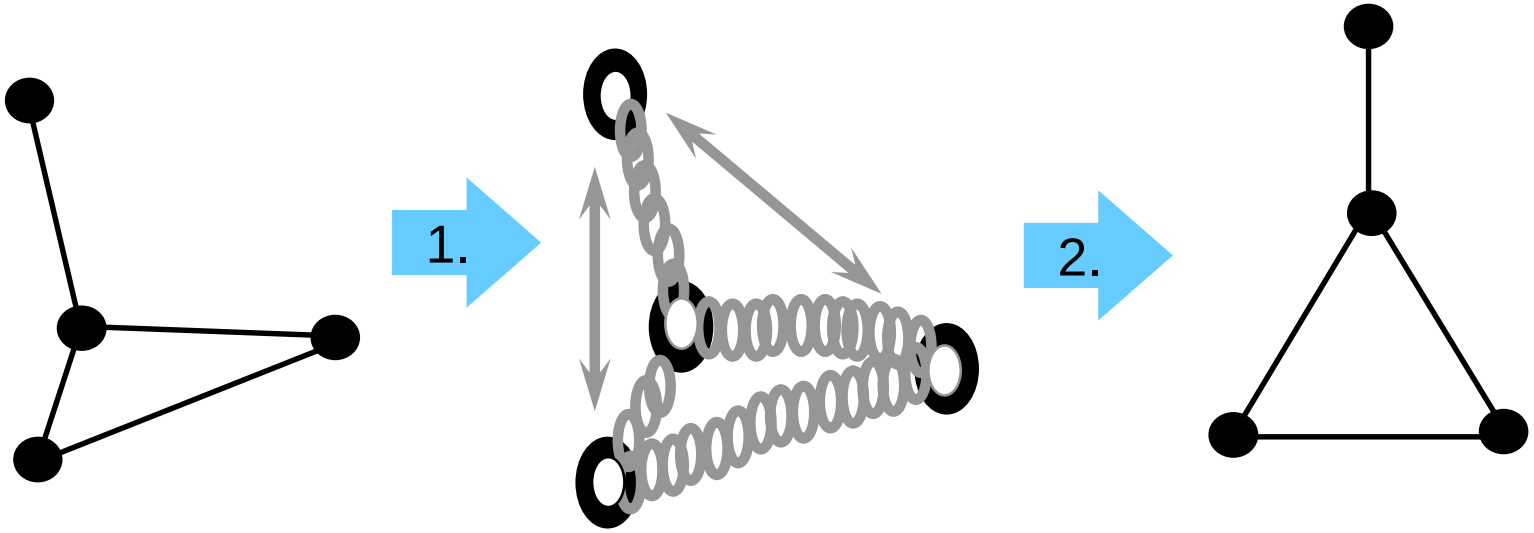
- step1. *Cycle removal*: make acyclic digraph
- step2. *Layer assignment*: assign y-coordinates
- step3. *Crossing reduction*: determine the order of vertices in each layer
- step4. *Horizontal coordinate assignment*: assign x-coordinates (Straighten the long edges)



Spring algorithm: Relationship

- *Spring algorithms* are the most popular.
  1. Place forces between pairs of vertices;
     - Attraction force (spring): for each edge
     - Repulsion force: between each pair of vertices
  2. Find a zero force configuration (equiliribrium).

Spring算法是最受欢迎的。
1.在顶点对之间放置力;
- 吸引力（弹簧）：每个边
- 排斥力：在每对顶点之间
2.找到零力配置（平衡）。



Spring algorithms: Constraints

- Force-directed methods can be extended to support several types of constraints

A. **Position constraints**

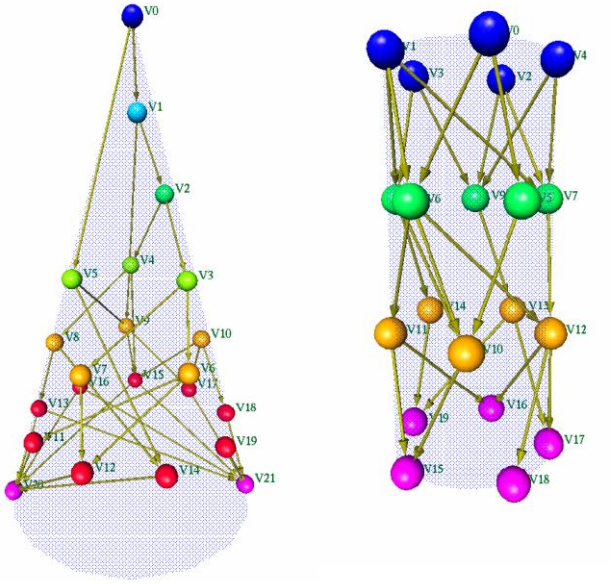
B. Fixed-subgraph constraints

C. Constraints that can be expressed by force/energy

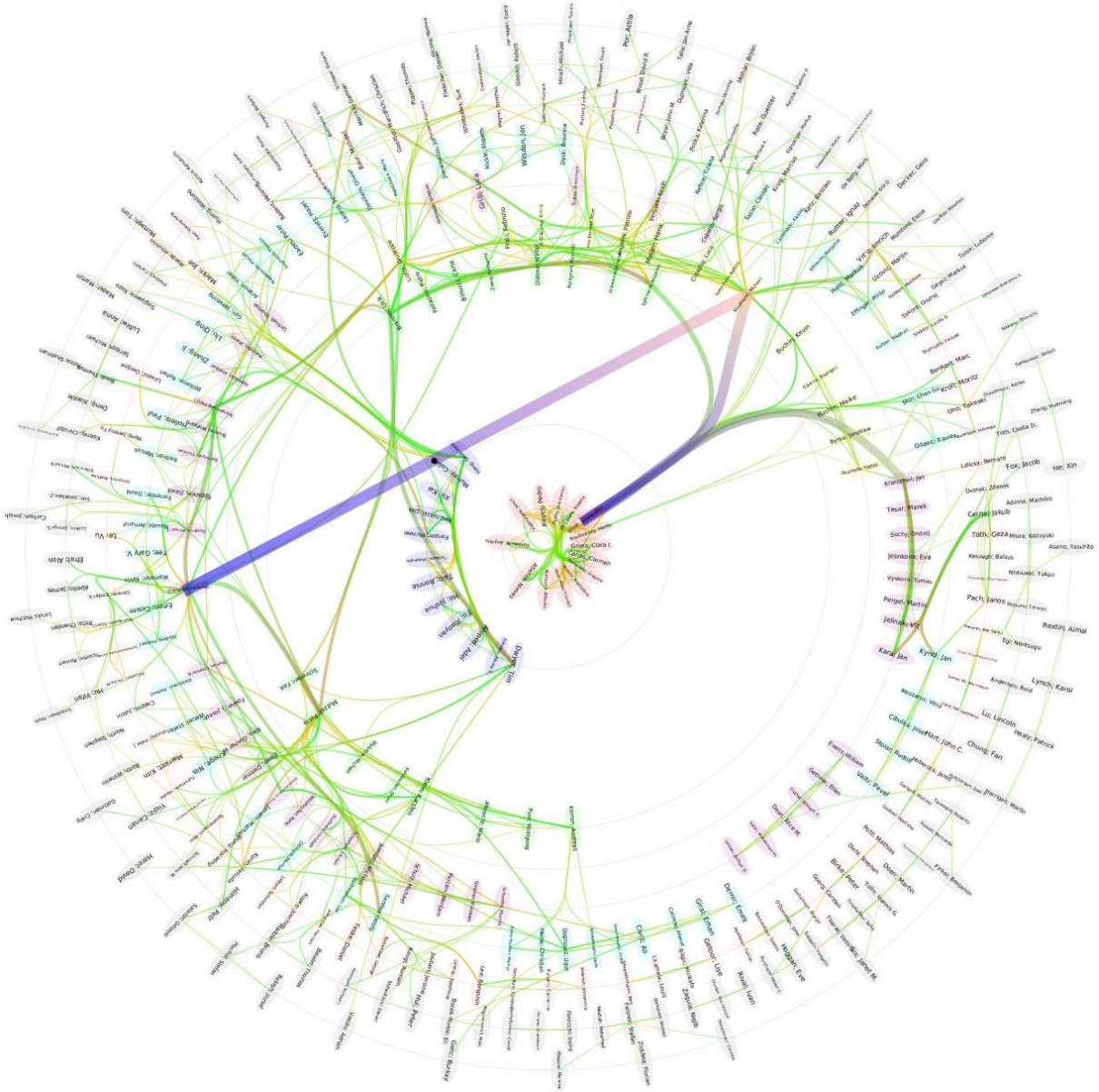
A. Position constraints

- assign to a vertex a topologically connected region where the vertex should remain
  - Single point: a vertex nail down at a specific location
  - Horizontal line: group of vertices arranged on a layer
  - Circle: set of vertices to be restricted to a distinct region

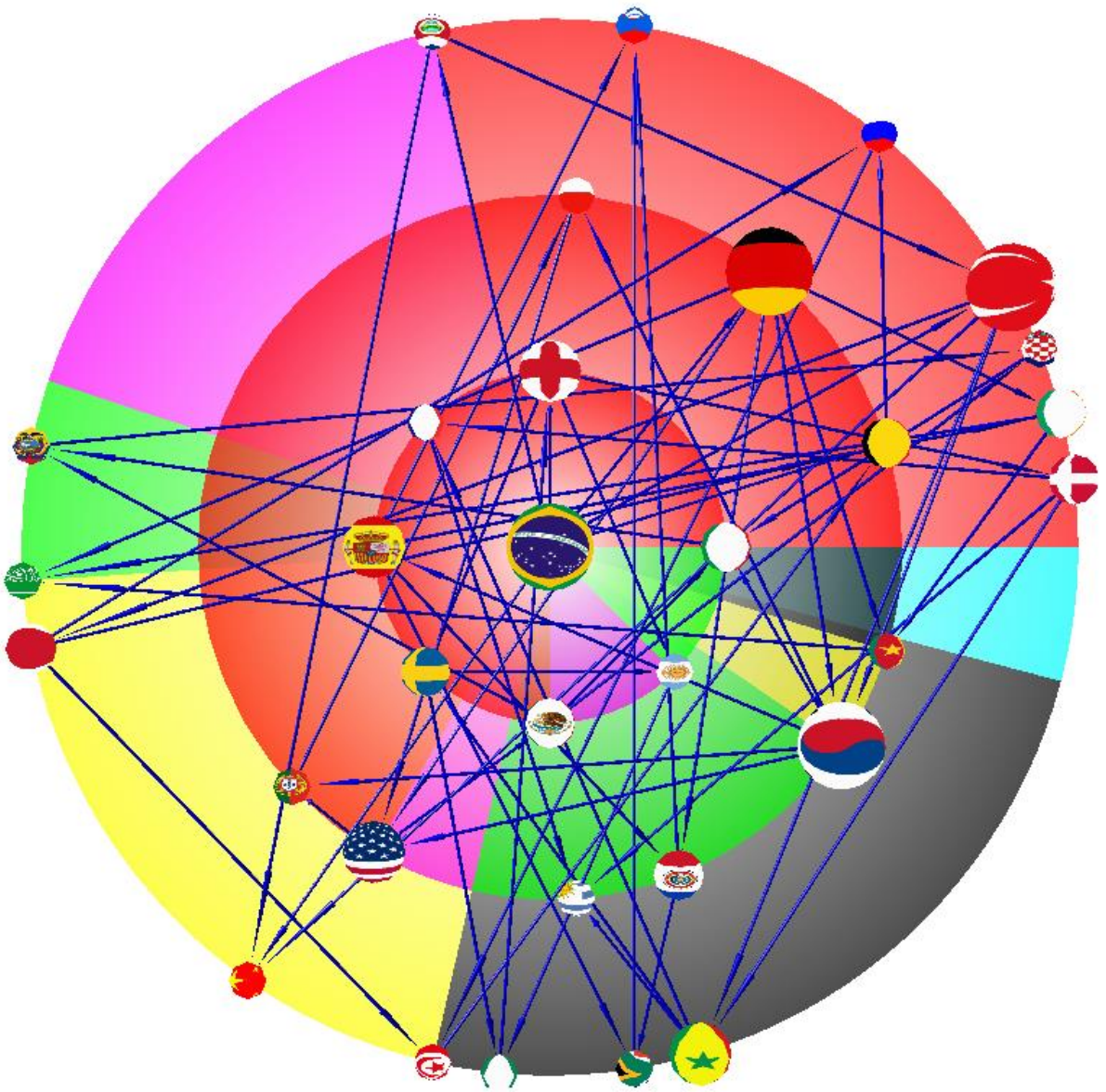
- [Ostry96]: constraints vertices to curves and 3D surfaces



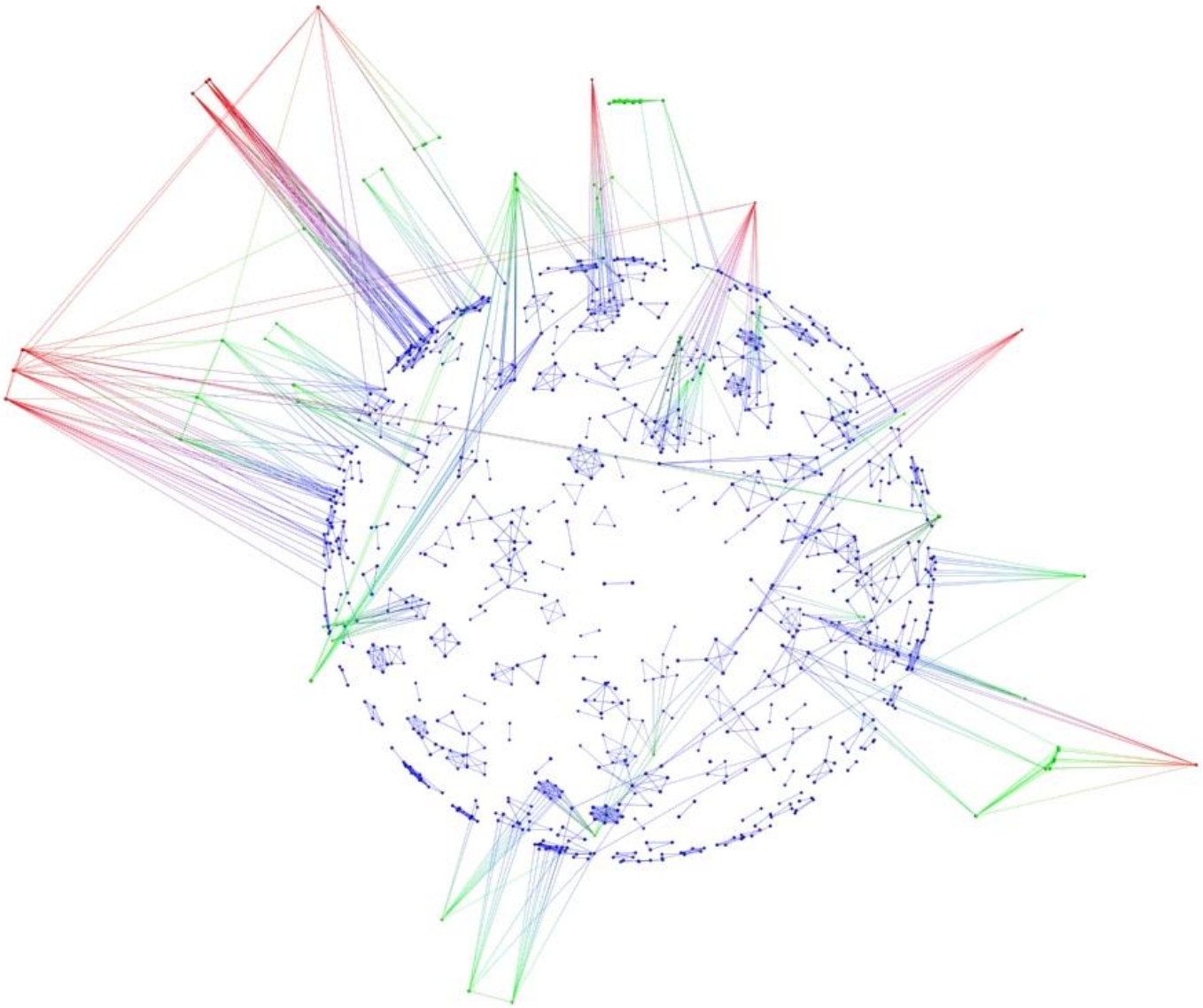
k-Core Visualisation: concentric circles



World Cup 2002: concentric circles

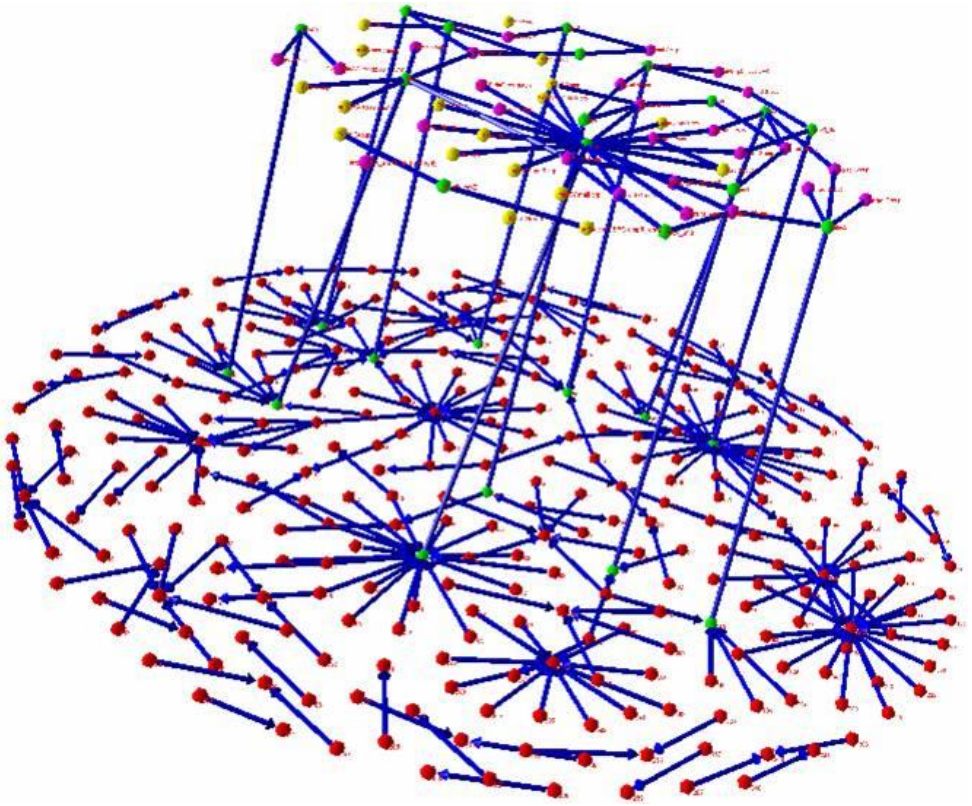


Collaboration Network: concentric sphere (sphere)

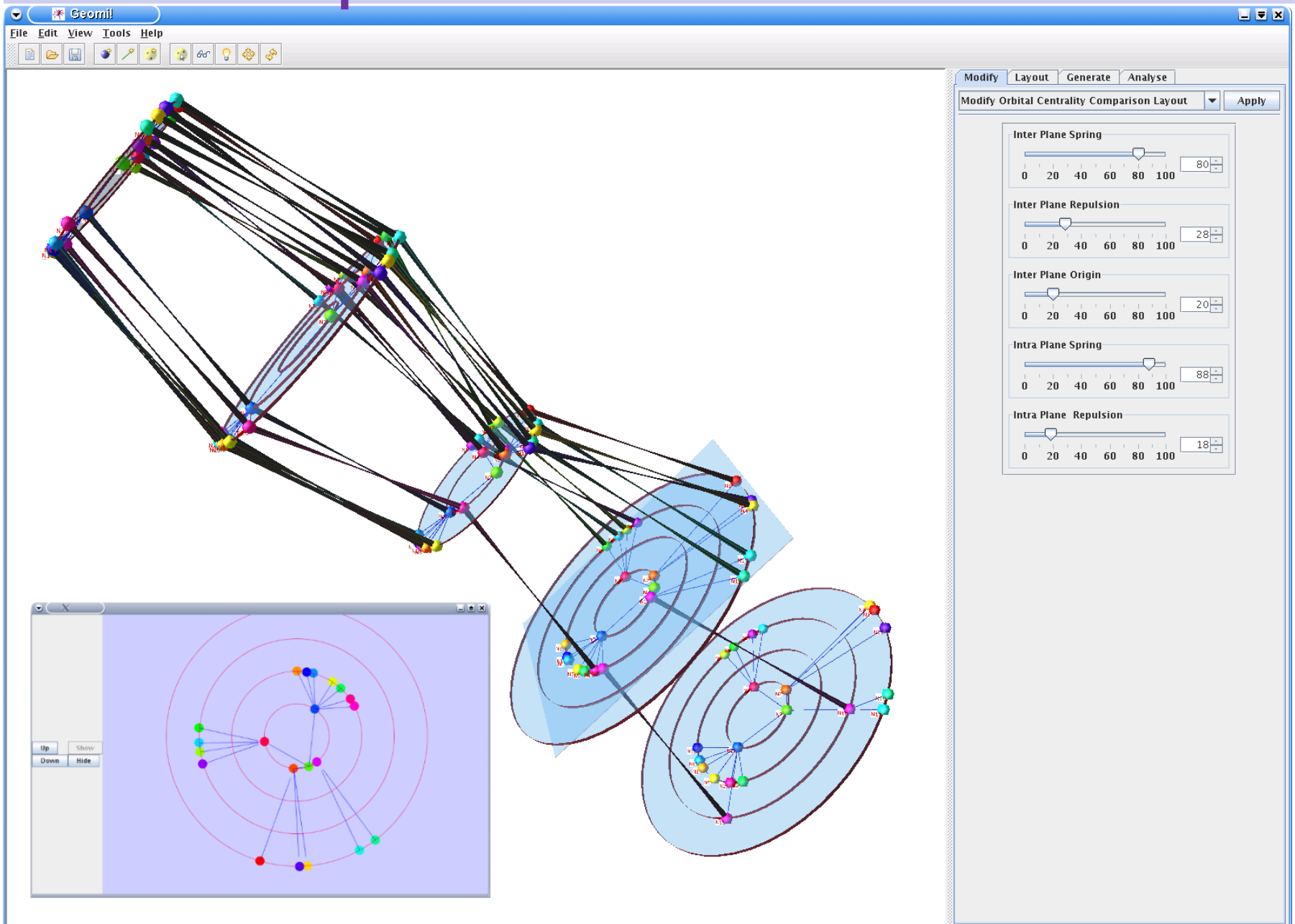


Visualisation of Network Motif: Plane constriants

- “Motifs”: small pattern in the network which occurs with significantly high frequency.
- Data: transcriptional regulation network of Escherichia coli.

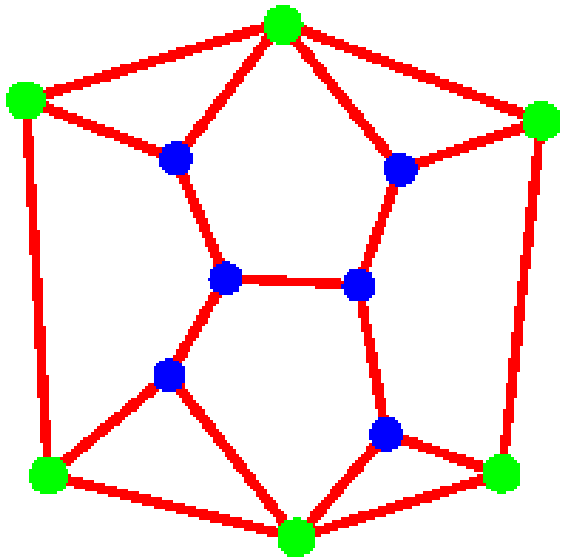


Comparison of Network Centralities: plane + concentric circle



B. Fixed subgraph constraints

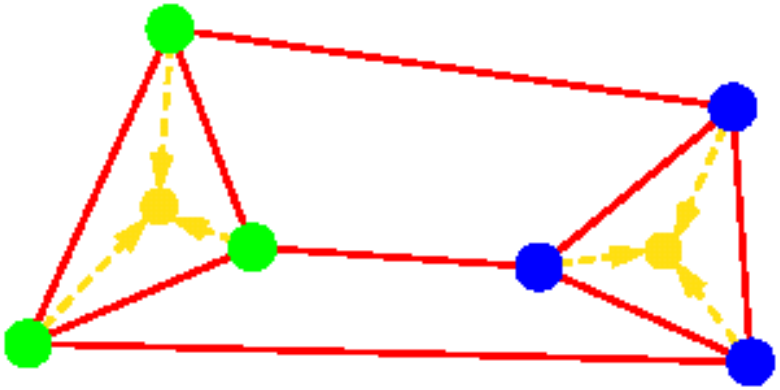
- Assign prescribed drawing to a subgraph .
- May be translated or rotated, but not deformed.
- Considering the subgraph as a rigid body.
- For example, barycenter method is a force-directed method that constrains a set of vertices (fixed external vertices) to a polygon.



C. Constraints expressed by forces

- Constraints expressed by forces
  - Orientation of directed edges: **magnetic spring**
  - Geometric clustering of special set of vertices
  - Alignment of vertices

- Visualise graph clustering (***clustered graphs***) [Eades et al. 97]
  - For each set *C* of vertices, add a dummy attractor vertex v_C .
  - Add attractive forces between an attractor v_C and each vertex in *C*.
  - Add repulsive forces between pairs of attractors and between attractors and vertices not in any cluster.



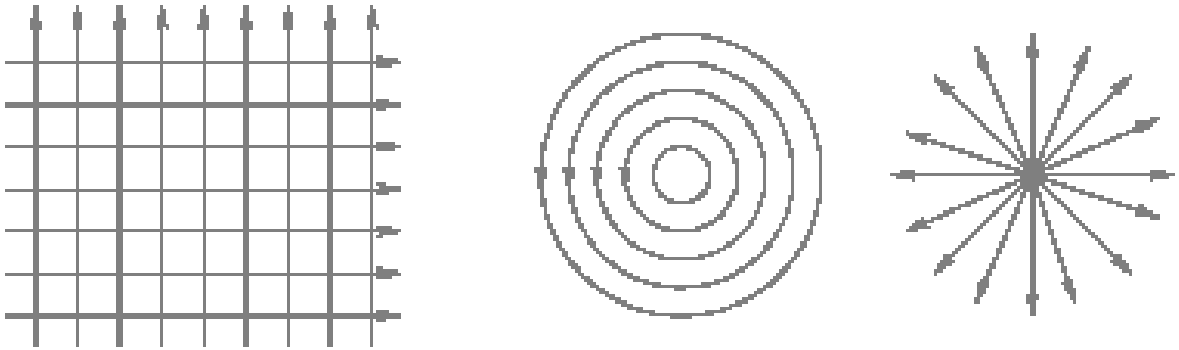
Magnetic Spring [Sugiyama Misue 95]

- Variations:
  - Some or all of the springs are magnetized
  - There is a global magnetic field that acts on the spring
  - Magnetic field can be used to control the orientation of edges

变化：
- 部分或全部弹簧被磁化
- 有一个作用在弹簧上的全球磁场
- 磁场可用于控制边缘的方向

- 3 types of magnetic fields
  - **Parallel**: all magnetic forces operate in the same direction
  - **Concentric**: the force operates in concentric circles
  - **Radial**: the forces operate radially outward from a point

•3种磁场
- 平行：所有磁力在同一方向上运行
- 同心：力以同心圆运行
- 径向：力从一个点径向向外操作



- The three basic magnetic fields can be combined
  - encourage orthogonal edges with a combination of parallel forces in the horizontal & vertical directions

- The springs can be magnetized in two ways:
  - Unidirectional: the spring tends to align with direction of the magnetic field
  - Bidirectional: the spring tends to align with the magnetic field, but in either direction

•可以组合三个基本磁场
- 在水平和垂直方向上鼓励具有平行力组合的正交边缘
•弹簧可以通过两种方式磁化：
- 单向：弹簧趋向于与磁场方向对齐
- 双向：弹簧倾向于与磁场对齐，但在任一方向上

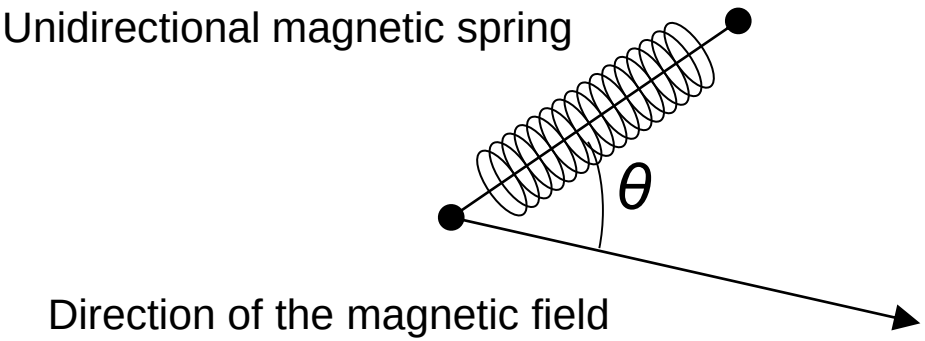
- The magnetic field induces a ***torsion*** or ***rotational force*** on the magnetic springs.
- For a unidirectionally magnetized spring representing (u,v), the force is proportional to $d(p_u, p_v)^{\alpha}\theta^{\beta}$

  $d(p_u, p_v)$: Euclidean distance between p_u and p_v

  θ: angle between the magnetic field and the line from p_u to p_v

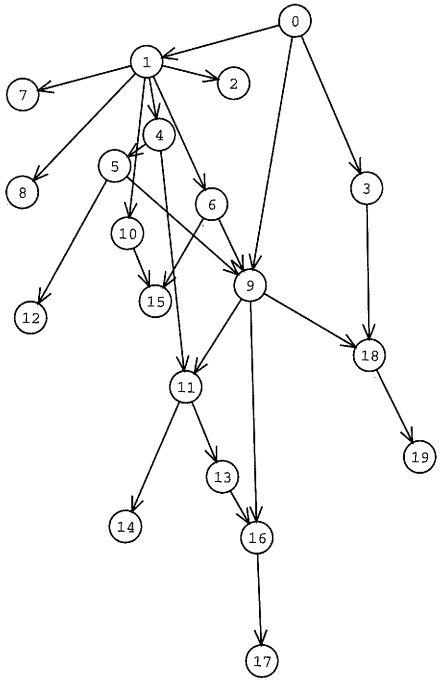
  α and β are constant

•磁场在磁弹簧上产生扭转力或旋转力。
•对于代表（u，v）的单向磁化弹簧，力与d（pu，pv）aq b成正比
d（pu，pv）：pu和pv之间的欧几里德距离
q：磁场与从pu到pv a和b的线之间的角度是恒定的

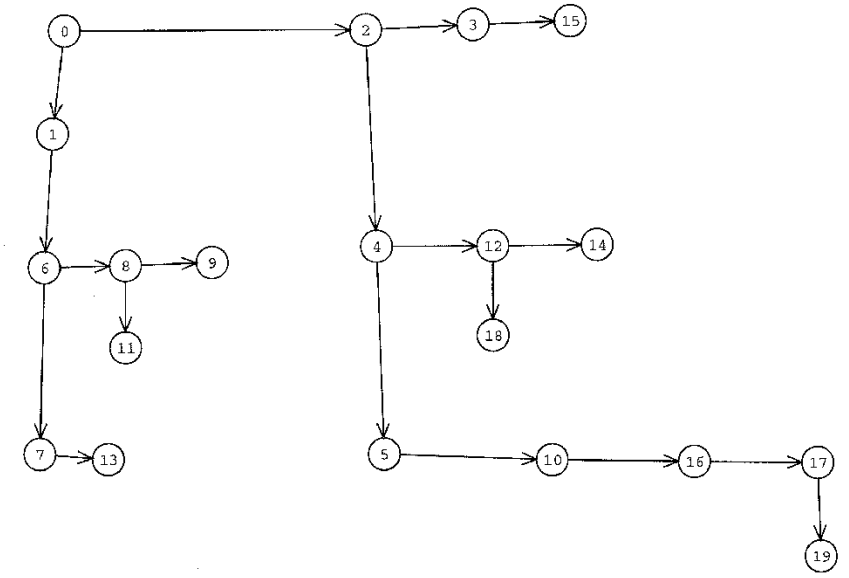


- The magnetic forces are combined with the spring & electrical force
- Algorithm to find equilibrium:
  - initially random position and at each iteration move the vertex to lower energy position
- Can handle ***directed graphs*** (unidirectional springs with one of the 3 fields)
  - arcs point downward: downward parallel field
  - Outward: radial field
  - Counterclockwise: concentric field
- Can be applied to *orthogonal drawings*: combined vertical & horizontal field with bidirectional springs
- Applied with success to *mixed graphs* (graph with both directed & undirected edges)

Examples



Vertical magnetic field



Vertical and horizontal magnetic field

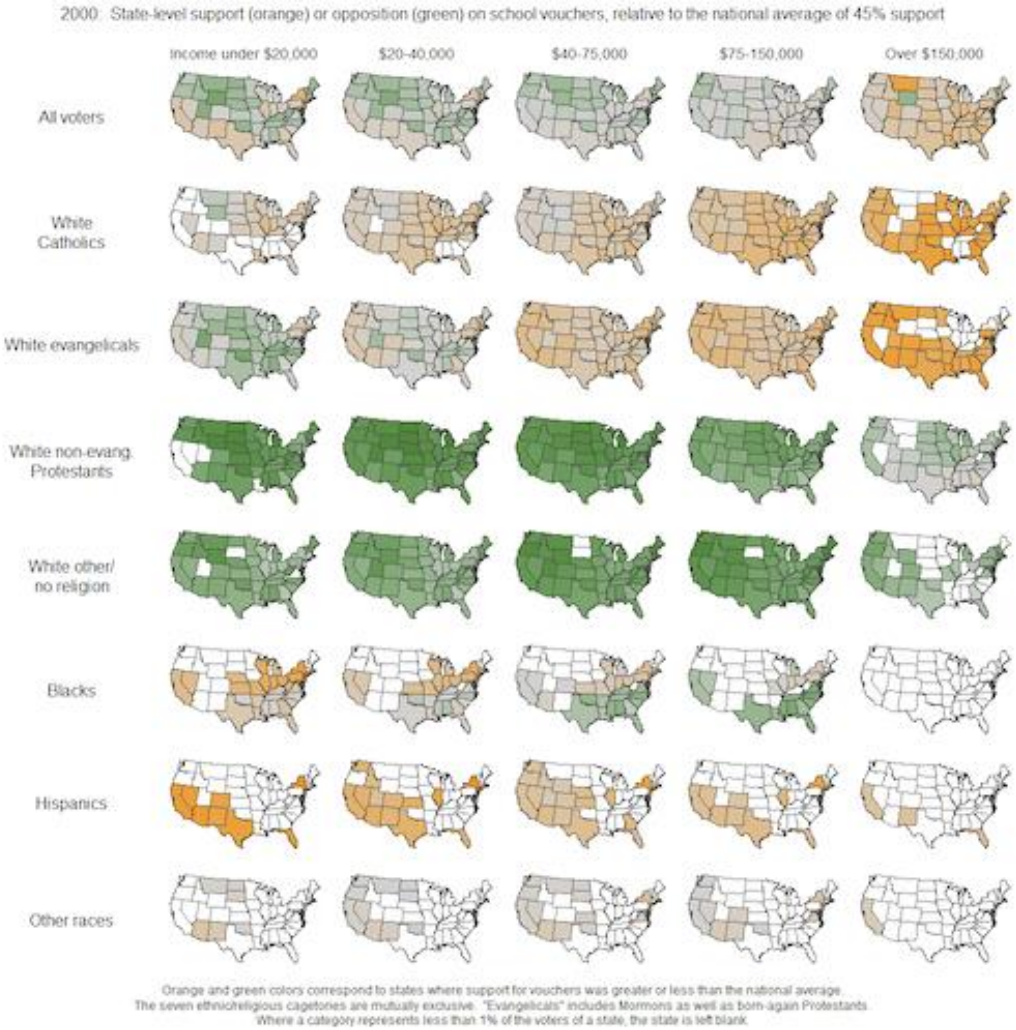
(5) Multi-relational Network Visualisation

1. Small Multiples

2. 2.5D Visualisation

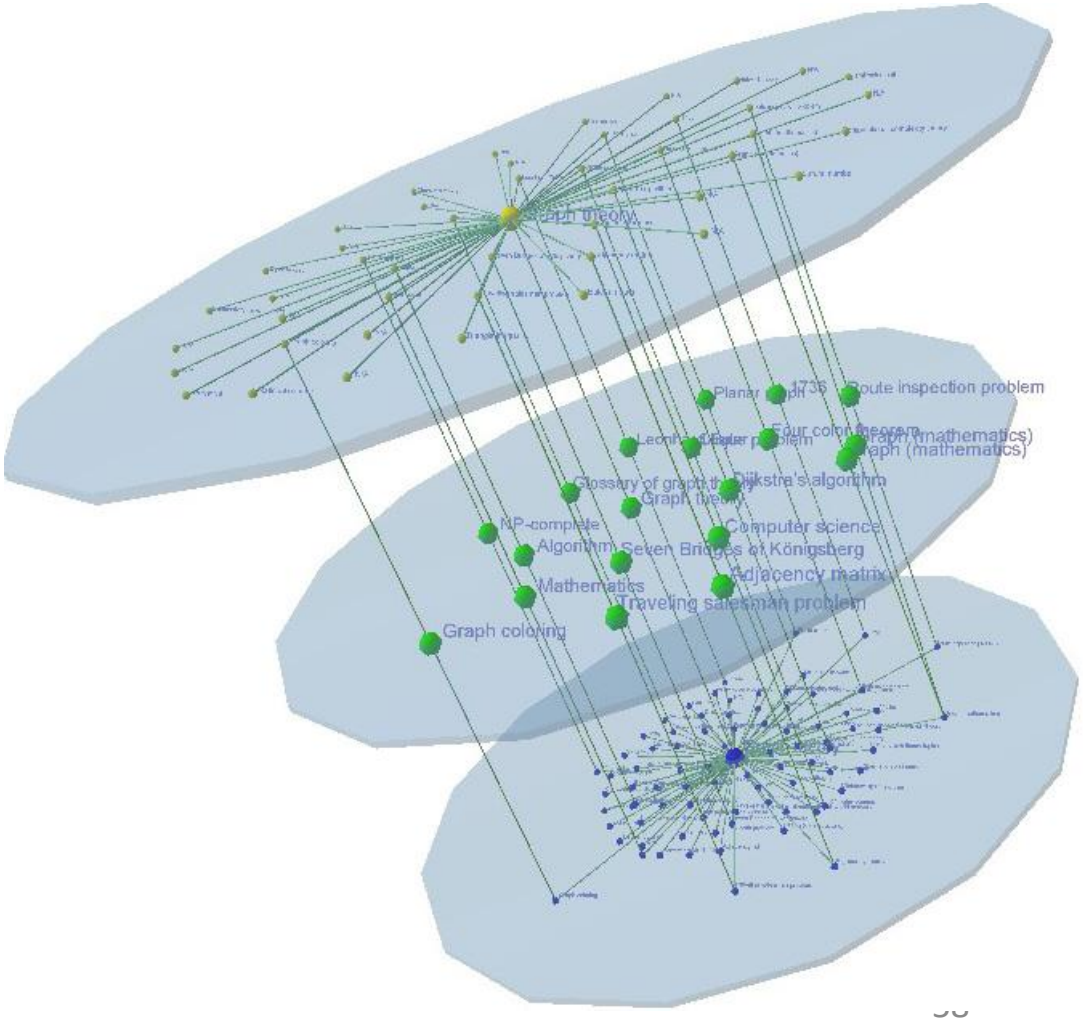
3. Multistory

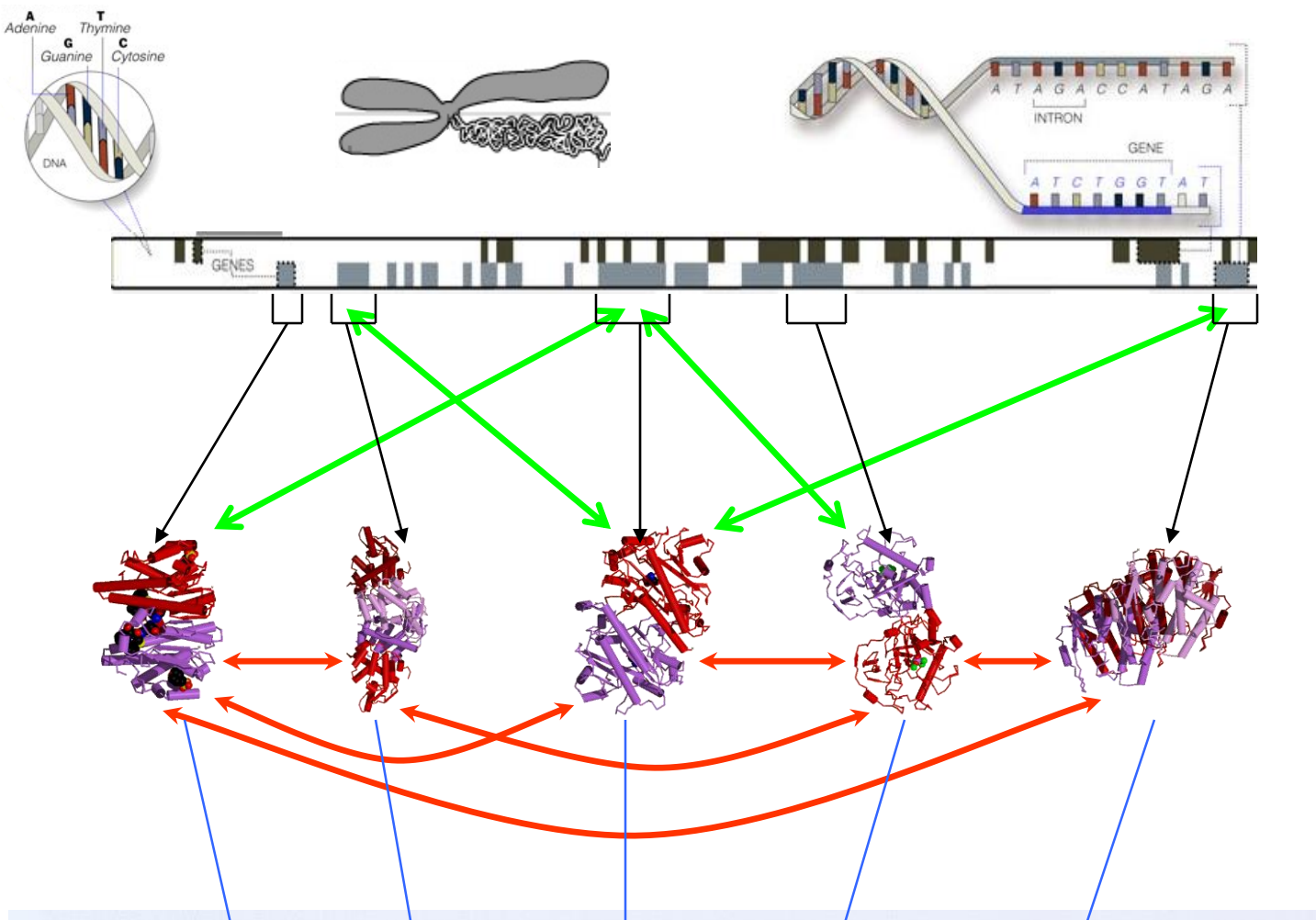
1. Small Multiples



2. 2.5D Visualisation

- Wikipedia networks
- Keyword – "*Graph Theory*"
- Top – German (47 nodes)
- Bottom – English (84 nodes)
- Middle – 19 overlappings





GENOME

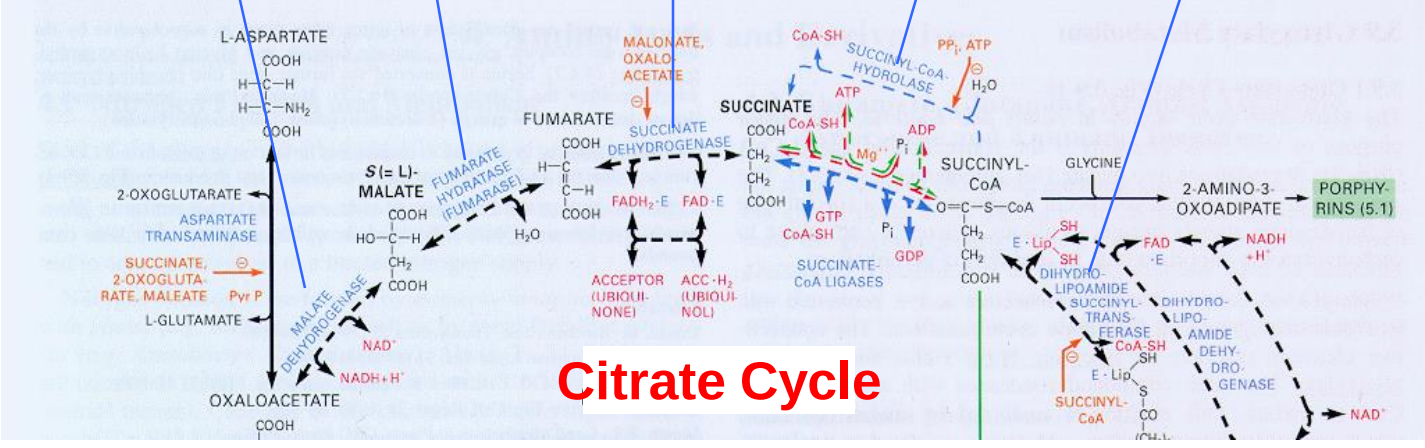
protein-gene interactions

PROTEOME

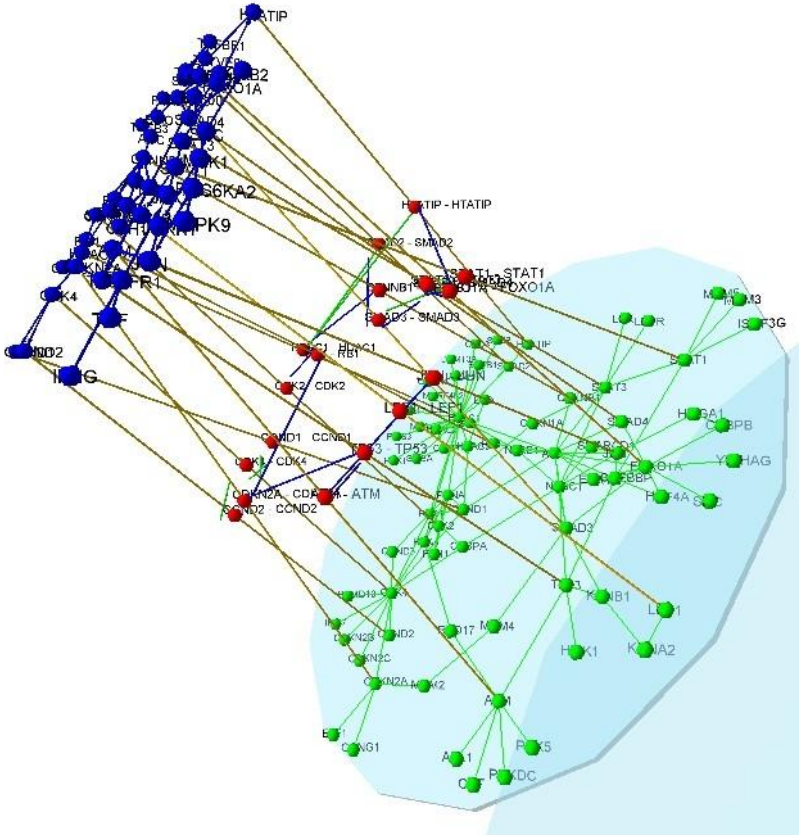
protein-protein interactions

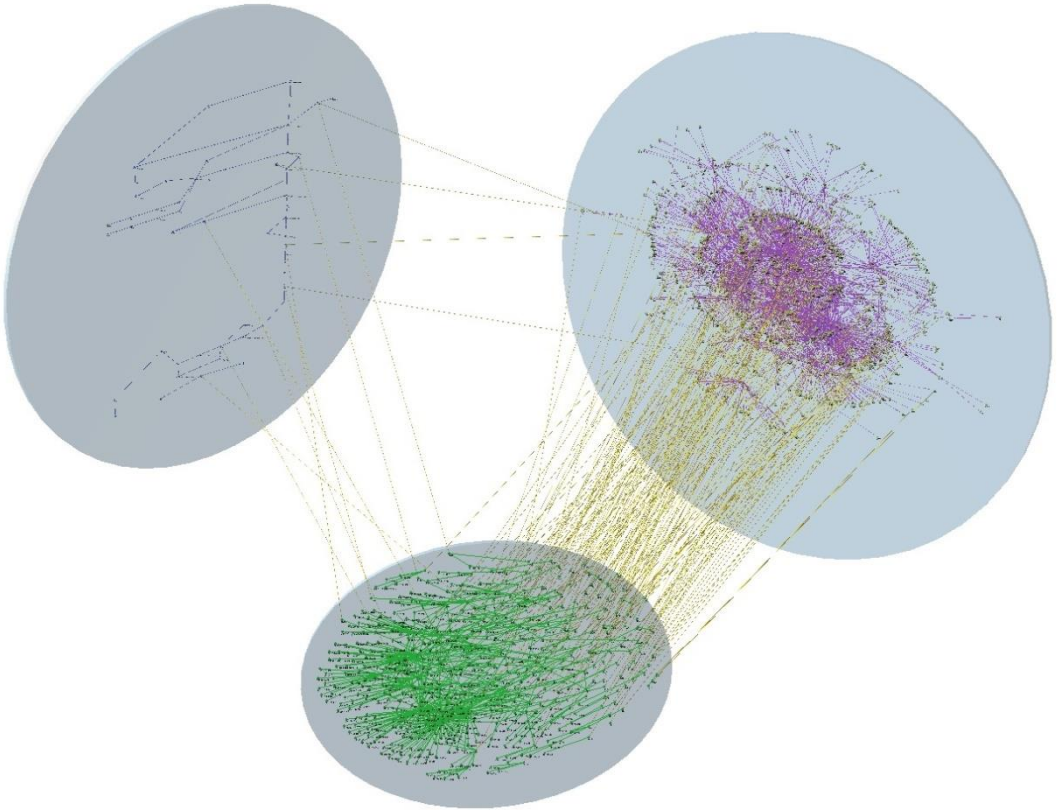
METABOLISM

Bio-chemical reactions



Multiple Biological Networks

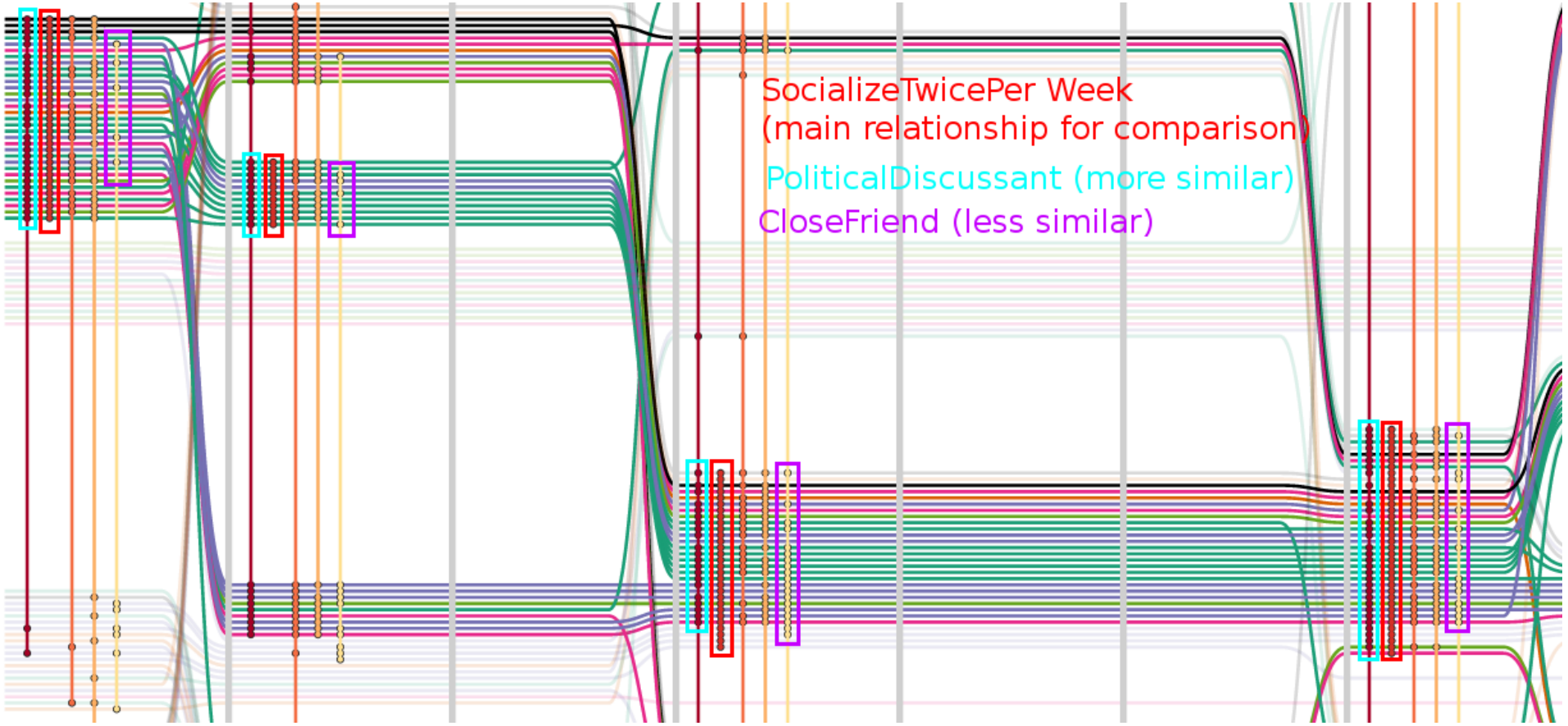




3. MultiStory: Dynamic Multi-relational Networks

数据:麻省理工学院现实挖掘数据
任务:动态多关系聚类相似性分析

Data: MIT reality mining data
Task : Dynamic multi-relationship cluster similarity analysis

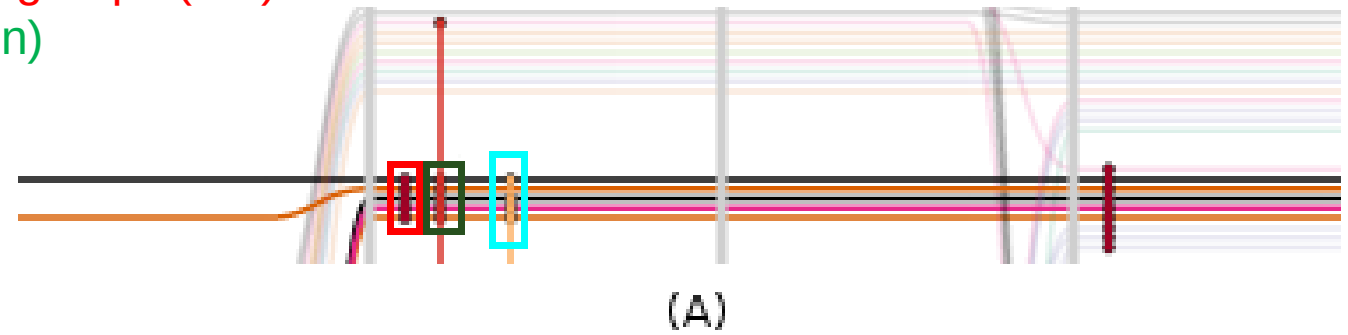


Compare similarity of dynamic clusters across different time steps

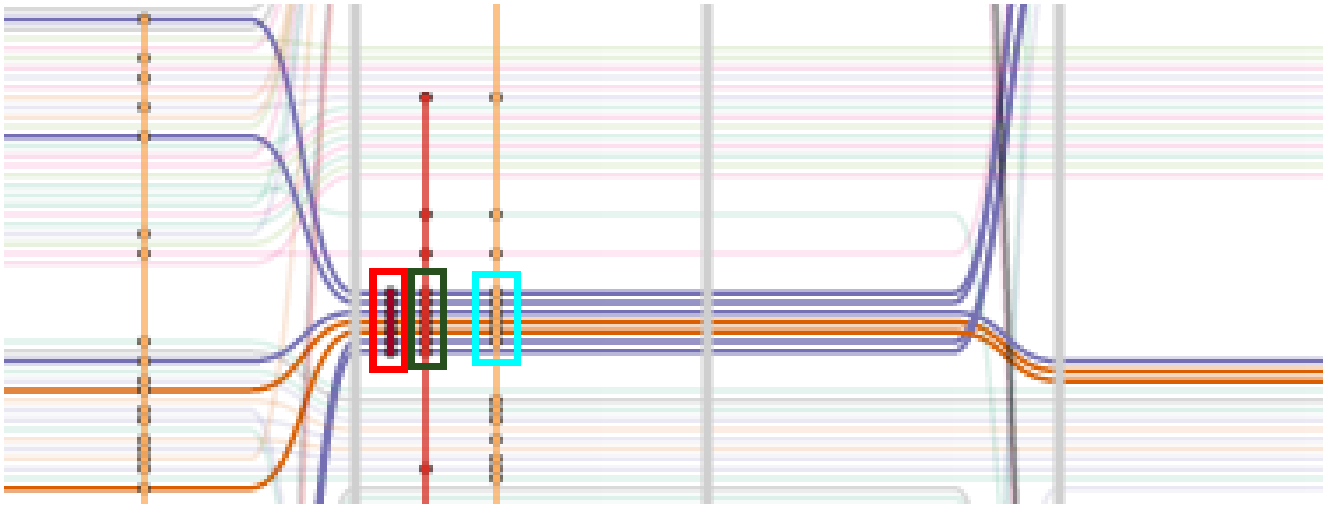
Task : Inter-relationship correlation/causal analysis

任务:相互关系相关/因果分析

small offline socialisation groups (red)
political discussant (green)
blogs (light blue)



(A)

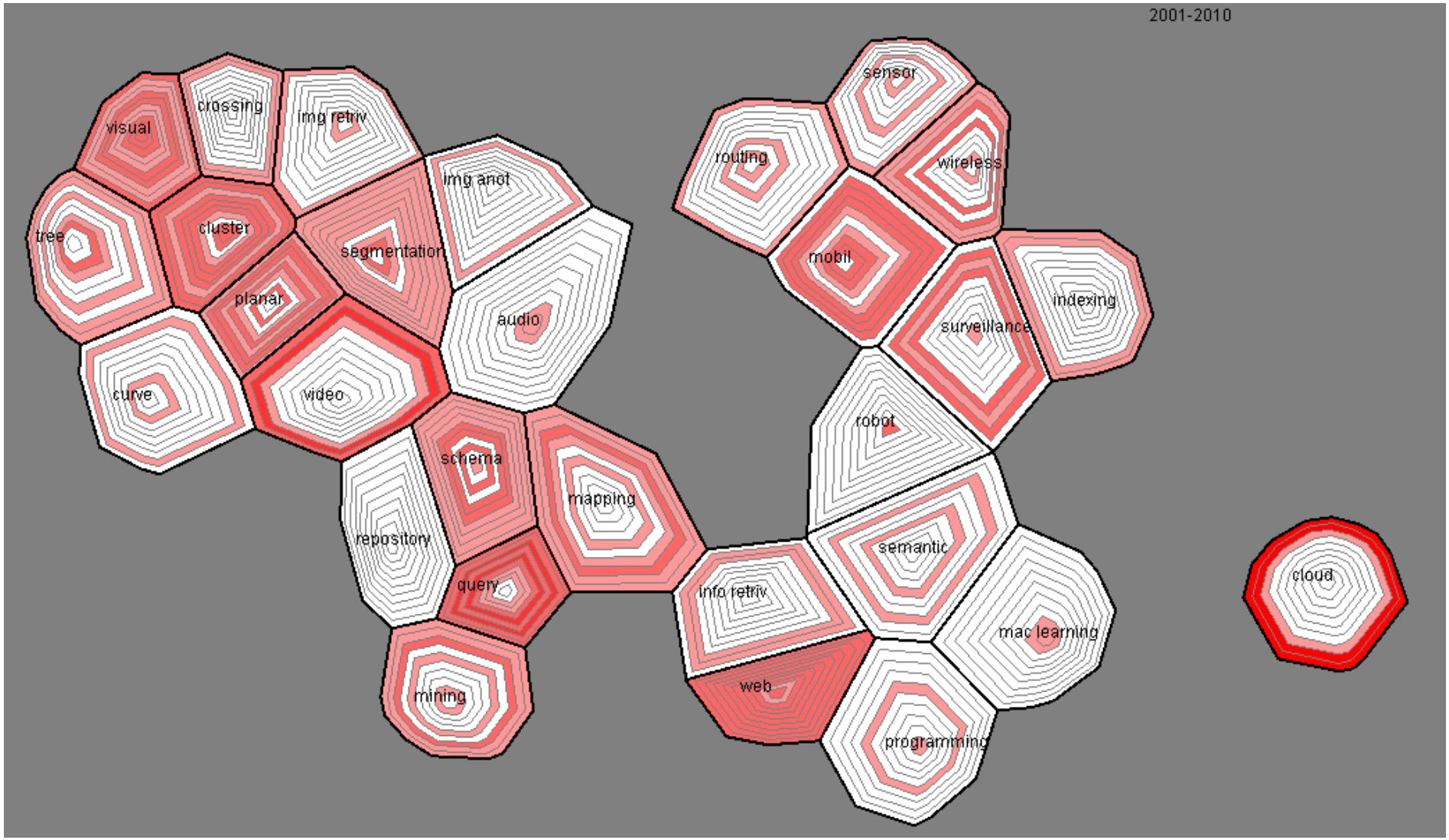


(B)

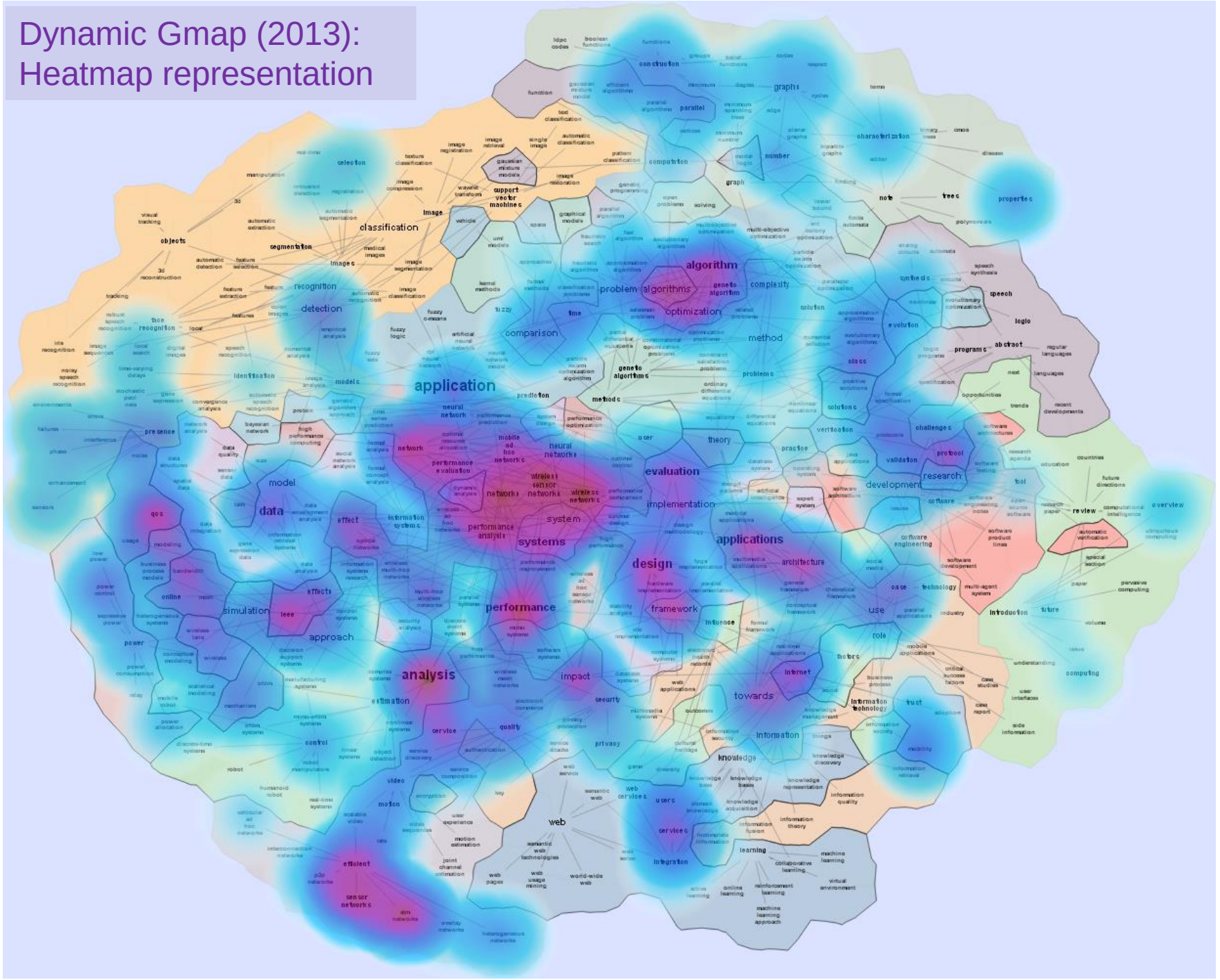
November 2008 during the US elections
online political discussions may prompt socialisation

Variations and Extensions: recent examples

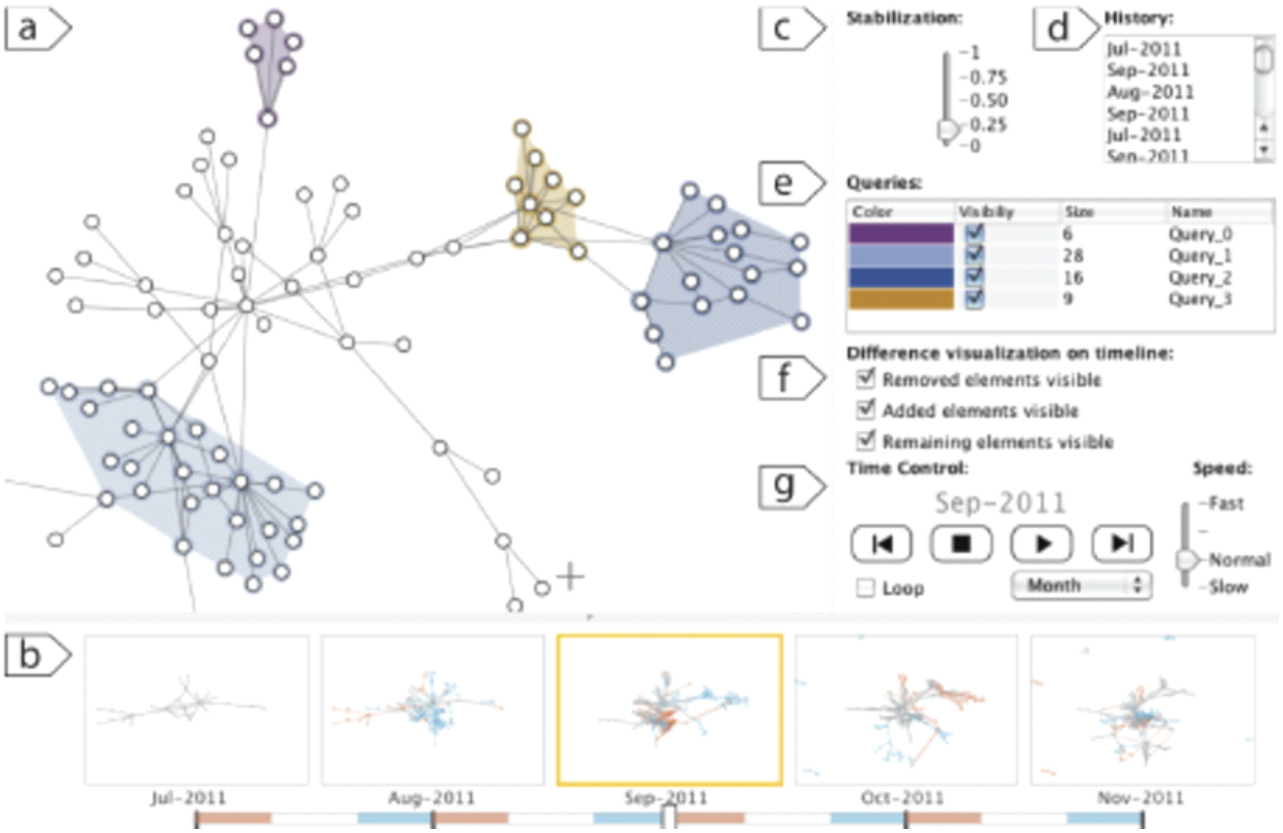
Temporal Map of DBLP (2012)



Dynamic Gmap (2013): Heatmap representation

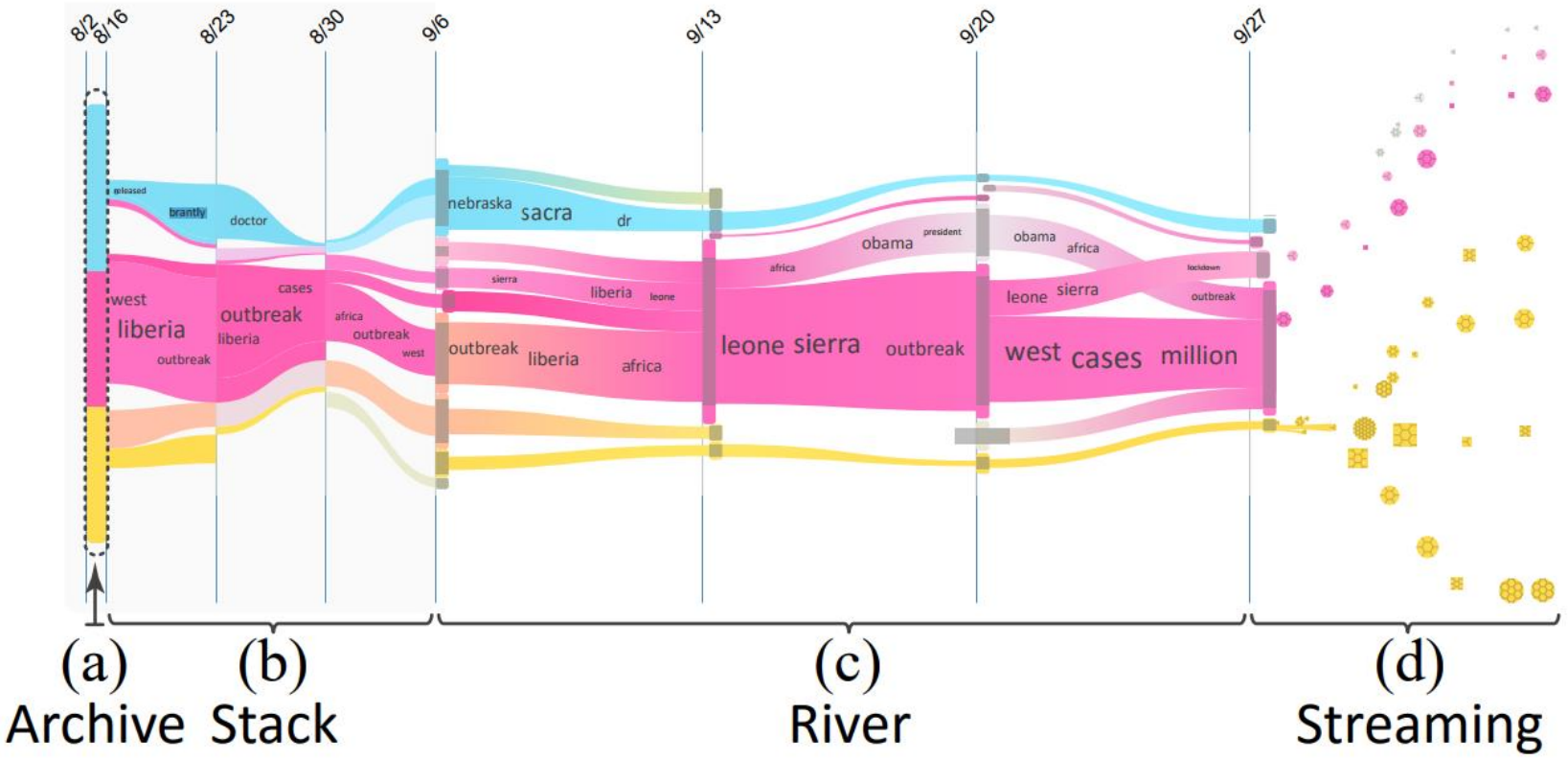


Graph Diaries (2014)



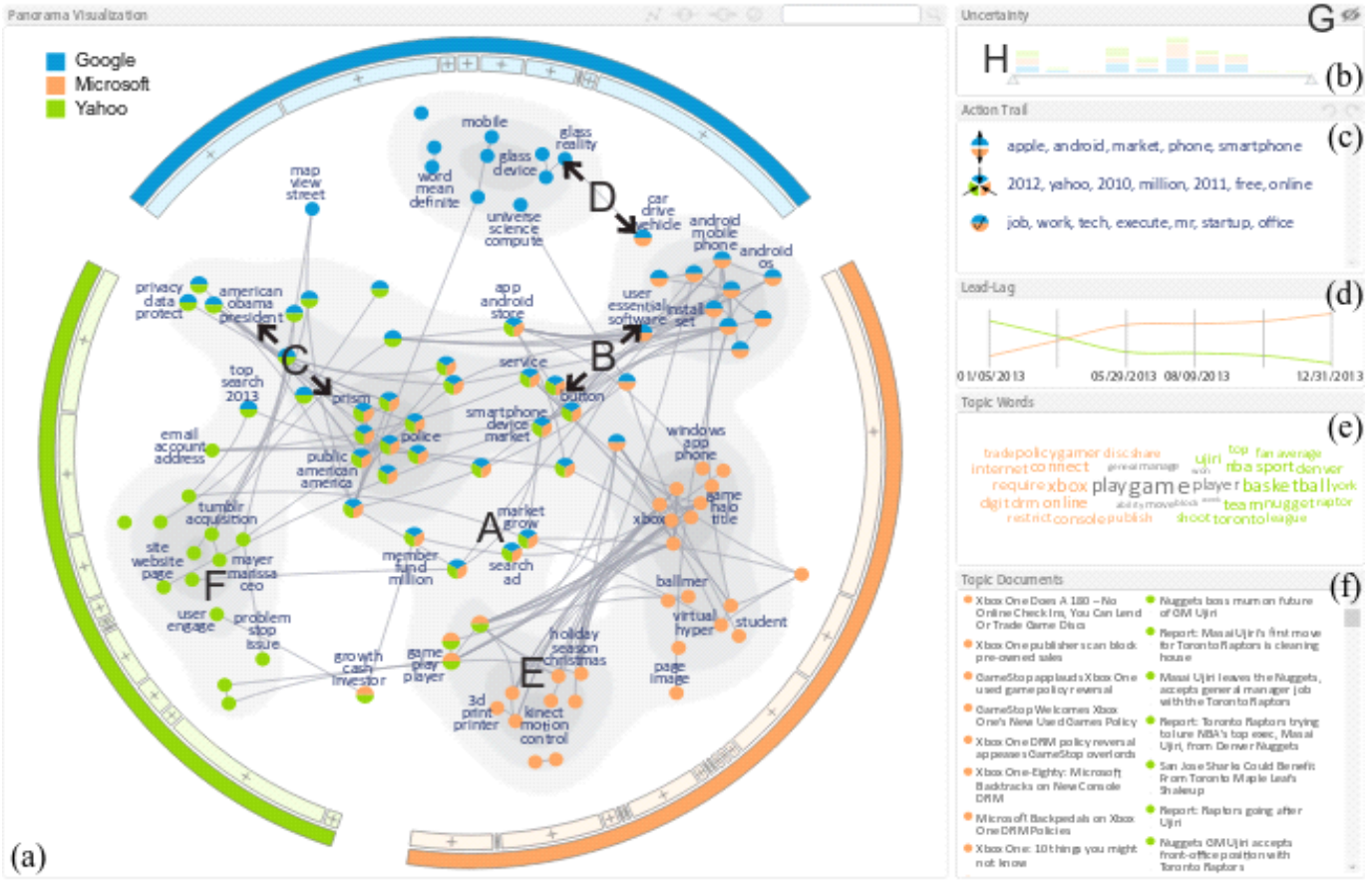
Source:
http://www.aviz.fr/~bbach/graphdiaries/Bach2013GraphDiaries.pdf

Online Visual Analytics of text streams (2016)



Source: http://shixialiu.com/publications/TopicStream/paper.pdf

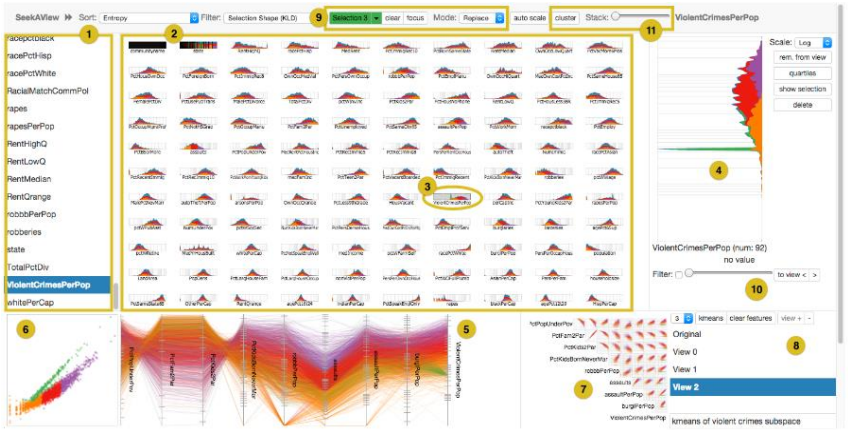
Topic Panorama (2016)



Source:
http://shixialiu.com/publications/TopicPanorama_TVCG/paper.pdf

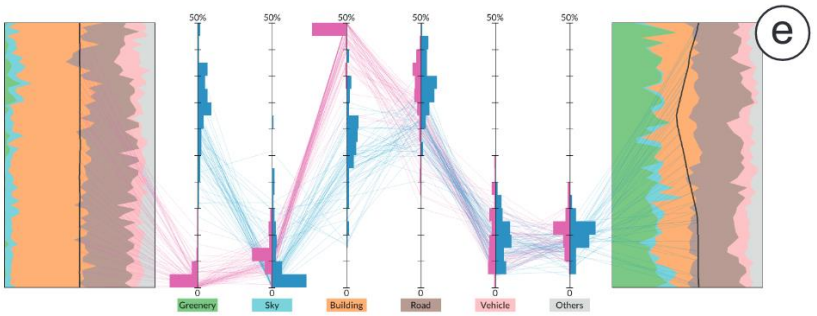
Parallel Coordinates (2016/17)

SeekAView: An Intelligent Dimensionality Reduction Strategy for Navigating High-Dimensional Data Spaces (2016)



Source: https://hal.inria.fr/hal-01377974/file/SeekAView-LDAV2016.pdf

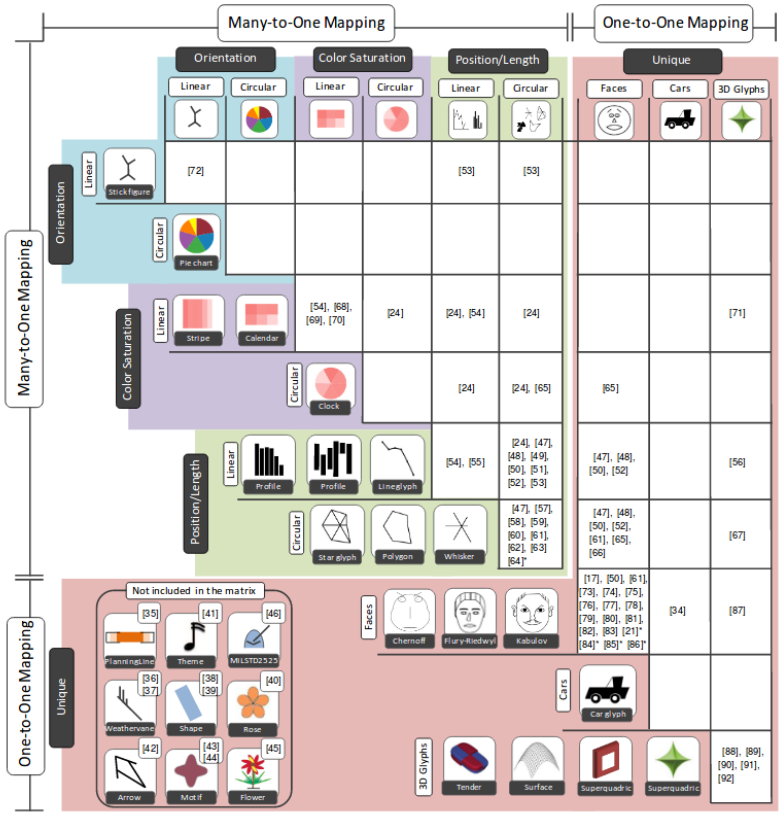
StreetVizor: Visual Exploration of Human-Scale Urban Forms Based on Street Views (2017)



Source: http://www.cse.ust.hk/~huamin/tvcg_qiaomu_streetvizor_2017.pdf

Glyphs (2016/17)

A systematic Review of Experiment Studies of data glyphs (2016)



Source: https://scibib.dbvis.de/uploadedFiles/StateOfTheArt.pdf

NEREx: Named-Entity Relationship Exploration (2017)

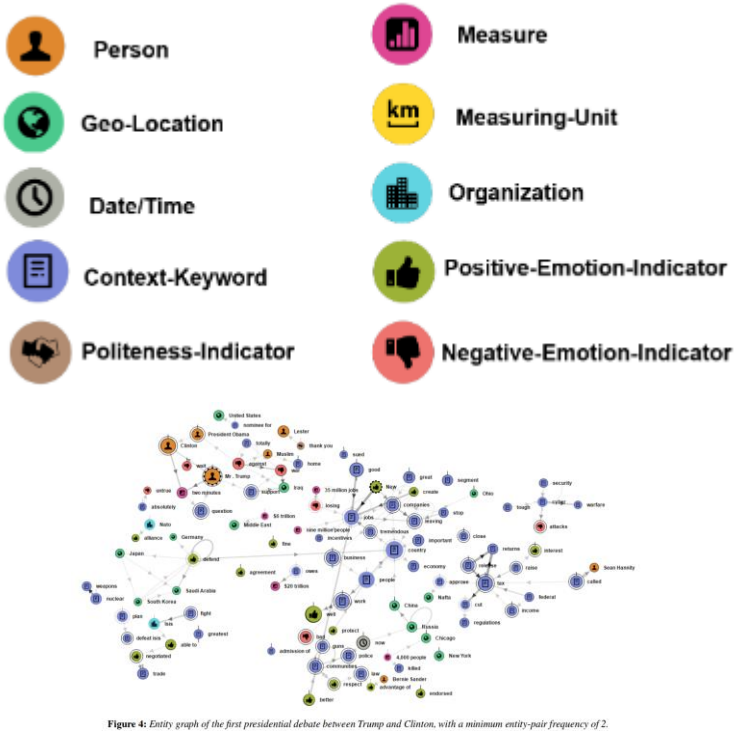
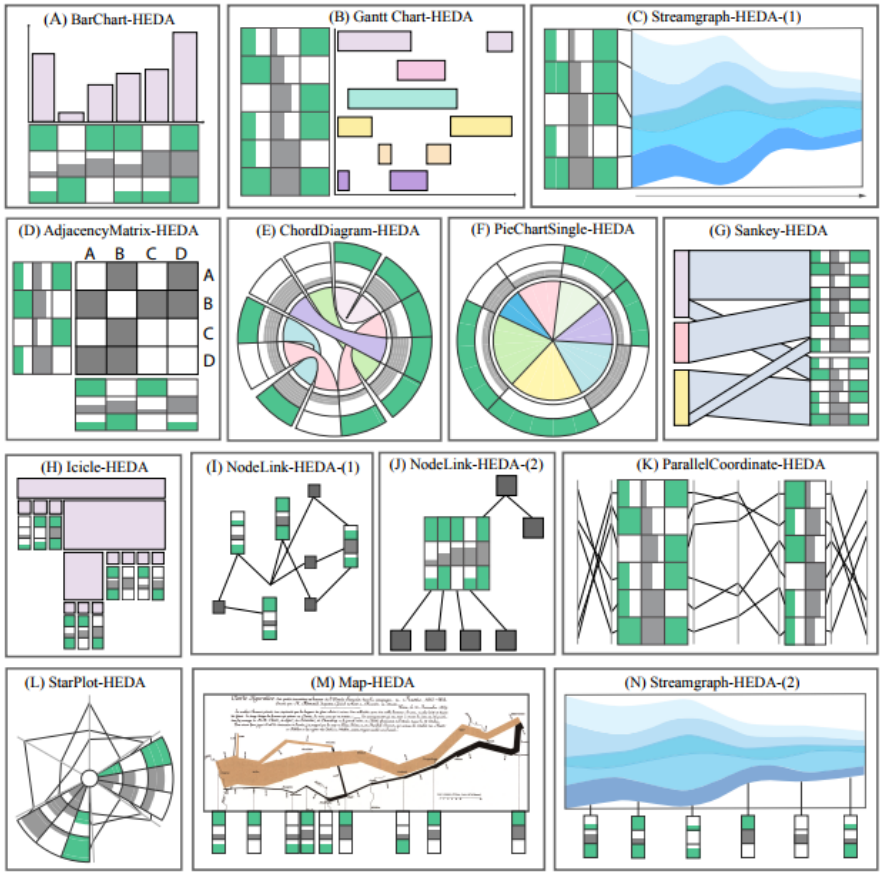


Figure 4: *Entity graph of the first presidential debate between Trump and Clinton, with a minimum entity-pair frequency of 2.*

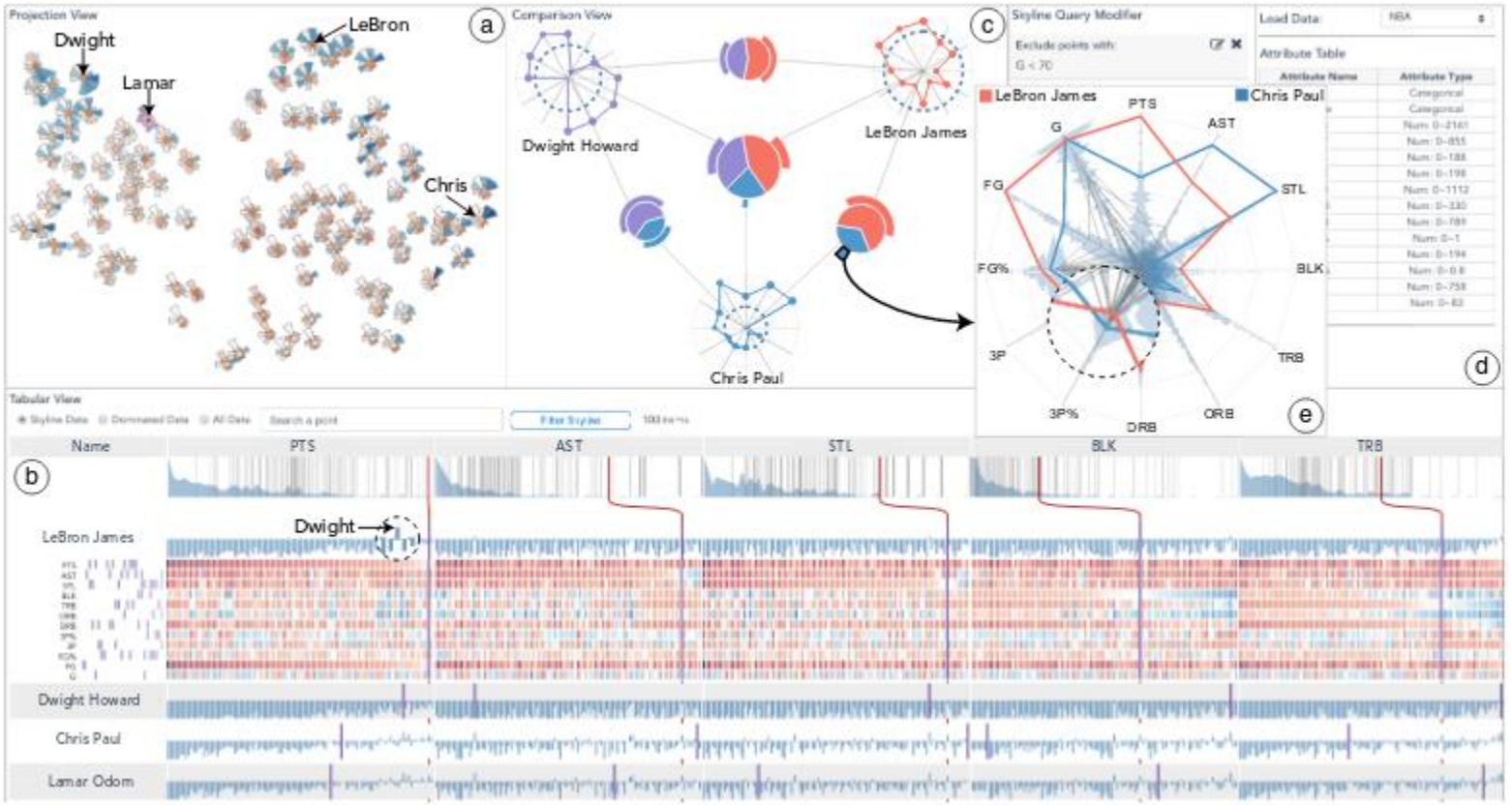
Source: https://scibib.dbvis.de/uploadedFiles/namedentityrelationship.pdf

Exploring the Possibilities of Embedding Heterogeneous Data Attributes in Familiar Visualizations (2017)



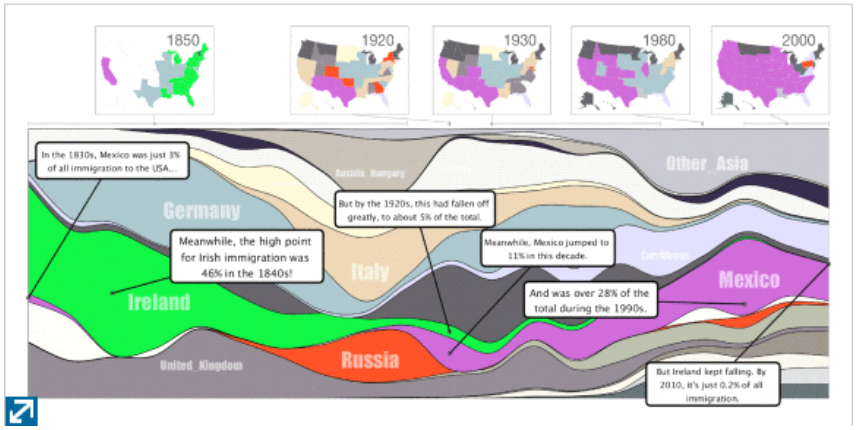
Source: http://innovis.cpsc.ucalgary.ca/innovis/uploads/Publications/Publications/HEDA-final.pdf

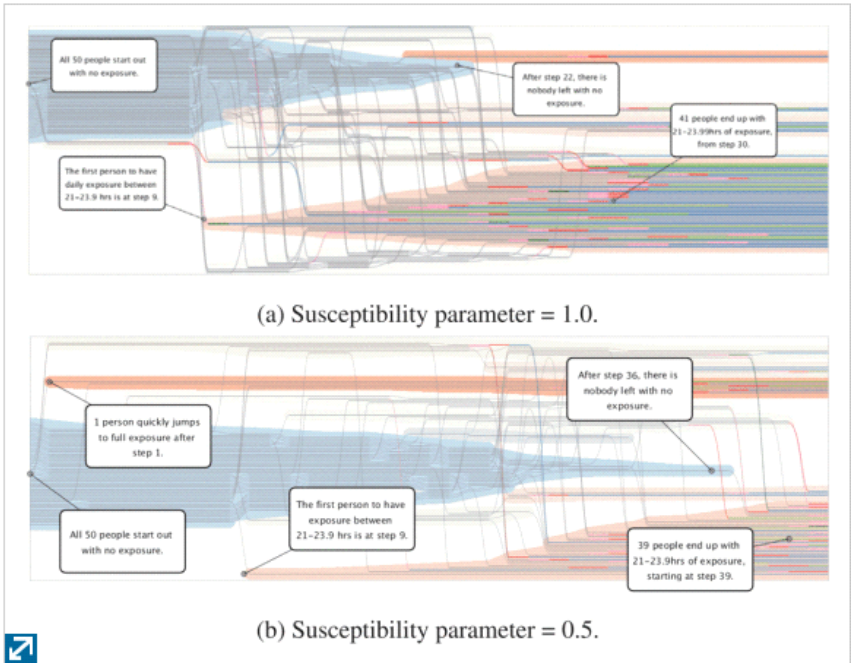
SkyLens (2017): Visual Analysis of Skyline on Multi-dimensional Data



Source: http://www.cse.ust.hk/~huamin/tvcg_xun_skyline_2017.pdf

Temporal Summary Images (2017)





Source: http://vis.cs.ucdavis.edu/~cjbryan/papers/tsi_paper.pdf

Assignments

Assignment 1 (week 7)

1. Submit any attempt by the deadline (no late submission)

2. Check the previous winning entries (graph B).

4. Submit your entry to GD 2018 contest (graph A).

5. You can draw by hand/use any tools/ do your own implementation.

6. Visualise the whole data set for an overview.
Then reduce the data set to subsets for the details.

7. You can submit >1 entries for each data
- should be significantly different layouts/tools
- add comparison analysis

Assignment 2

1. Group: 6 people
- choose a group leader and record weekly meeting minute
(Everyone should contribute)

2. Choose a suitable data set
- data processing
- design: analysis/visualisation
- implementation: tools
- evaluation: tasks

3. Implementation
- you can use any existing tools (for analysis, visualisation)
- you can build a system using open source tools (D3, Gephi, GEOMI)

4. Reports/Presentation
- initial report (5 mark): (week 8)
- presentation (10 mark): (submit by week 10) presentation: week 10-12
- final report (15 mark): (week 13)

Midterm Survey